RatSWD
German Data Forum (ed.)

Building on Progress

Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences

vol | ume | 1

Budrich UniPress

Building on Progress
Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences

Building on Progress

Expanding the Research Infrastructure
for the Social, Economic, and Behavioral
Sciences

edited by the
German Data Forum (RatSWD)

Vol. 1

Budrich UniPress Ltd.
Opladen & Farmington Hills, MI 2010

You must not circulate this book in any other binding or cover and you must impose this same condition on any acquirer.

A CIP catalogue record for this book is available from
Die Deutsche Bibliothek (The German Library)

© 2010 by Budrich UniPress Ltd. Opladen & Farmington Hills
www.budrich-unipress.eu

ISBN 978-3-940755-58-2
2 Vols. in slipcase

© These volumes are published by Budrich UniPress and are published under the Creative Commons licence:
http://creativecommons.org/licenses/by-nc-nd/3.0/de/deed.en_GB
You are free to copy, distribute, display, and perform the work as long as you tribute it to the original author/editors. No commercial use, or changes unless approved by Budrich UniPress.

SOME RIGHTS RESERVED

Both volumes may be downloaded free of charge at the Budrich UniPress website www.budrich-unipress.com
The two volumes in slipcase may be ordered from your local bookseller or directly from the Budrich UniPress.

Budrich UniPress Ltd.
Stauffenbergstr. 7. D-51379 Leverkusen Opladen, Germany
28347 Ridgebrook. Farmington Hills, MI 48334. USA
www.budrich-unipress.eu

Jacket illustration by Walburga Fichtner, Cologne, Germany
Printed in Europe on acid-free paper by
paper & tinta, Warszaw, Poland

Preface by the Federal Ministry of Education and Research (BMBF)

Nobel Prize-winning economists Amartya Sen and Joe Stiglitz, in collaboration with a number of co-authors of the internationally acclaimed report "On the Measurement of Economic Performance and Social Progress," noted that:

> "Those attempting to guide the economy and our societies are like pilots trying to steering a course without a reliable compass. The decisions they (and we as individual citizens) make depend on what we measure, how good our measurements are and how well our measures are understood. We are almost blind when the metrics on which action is based are ill-designed or when they are not well understood. For many purposes, we need better metrics. Fortunately, research in recent years has enabled us to improve our metrics, and it is time to incorporate in our measurement systems some of these advances. There is also consensus among the Commission members that better measures may enable us to steer our economies better through and out of crises."

The German Data Forum (RatSWD) was founded to address these needs for more reliable statistics and better empirical research in Germany and beyond. The German Data Forum advises the German federal government and *Länder* governments on issues that impact the expansion and improvement of the research data infrastructure in the empirical social, behavioral, and economic sciences. Since it was established in 2004 by the German Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*), the German Data Forum has significantly advanced the agenda set forth by the Commission to Improve the Information Infrastructure (KVI, *Kommission zu Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik*) and has supported the work of research funding agencies by making recommendations on how the KVI agenda can be most effectively implemented. The German Data Forum has hereby helped make a wide range of high-quality, reliable microdata available to empirical researchers in the social, behavioral, and economic sciences at Research Data Centers and Data Service Centers throughout Germany.

These data are enabling researchers to expand the frontiers of scientific knowledge. Viewed in isolation, findings from discrete research disciplines appear unspectacular; only on rare occasions do they yield a fundamentally new picture of the world or of society. It is for precisely this reason that patience and a long-term perspective are so crucial for research funding and support. Of the many new conclusions that have been developed on the basis of empirical data from the Research Data Centers, two groundbreaking findings can be cited as evidence of this: First, data from German pension insurance carriers have been used by several researchers to identify significant differences between male and female life expectancy depending on the level of education and corresponding differences in workplace health risks. Second, data from the Federal Labor Office, in which firm statistics were merged painstakingly with data on employment structures, have been used to show that exporting firms pay higher wages than non-exporting firms. This would be impossible to see from the raw statistical data, since exporting firms have a different product portfolio and personnel structure than non-exporters.

The development and distribution of "Campus Files", a noteworthy contribution to university education, is also among the achievements of the Research Data Centers and Data Service Centers established by German Data Forum and the German Ministry of Education and Research. By working with original statistical data, students obtain more advanced methodological training with greater practical relevance. This will undoubtedly pay off substantially in the years (and decades) to come – particularly when the graduates begin putting their statistical expertise to work professionally in such fields as policy analysis and market research.

Despite the gains it has already made in expanding the research infrastructure, the German Data Forum is not content to rest on past achievements. To the contrary, in 2008 it launched the project, "Developing the Research Data Infrastructure for the Social and Behavioral Sciences in Germany and Beyond: Progress since 2001, Current Situation, and Future Demands." Building on its work from the last several years, the German Data Forum now aims to develop the research infrastructure even further, to ensure that it can meet future demands, and to identify emerging data needs in the German, European, and international contexts. The Federal Ministry of Education and Research will continue to lend its support in this important undertaking.

The support of the Federal Ministry of Education and Research has made it possible to bring together over 100 renowned experts from a wide range of disciplines in an ongoing dialog. The current publication delivers the results of this concentrated effort in two volumes. The nearly 70 advisory reports in the second volume offer a detailed look at the situation from the perspective of various branches of the social, behavioral, and economic sciences in order to identify specific data needs. It is a comprehensive and systematic compendium designed for use by research organizations, funding agencies, and statistical offices.

Government policy alone cannot create optimal conditions for improving the research infrastructure. Dialog with the research community and the federal statistical agencies is critical. Acting as a platform for this dialog is one of the key tasks of the German Data Forum. The Federal Ministry of Education and Research looks forward to being a participant in this discussion.

Berlin, November 2010

Cornelia Quennet-Thielen
State Secretary
Federal Ministry of Education and Research (BMBF)

Table of Contents:

Volume 1:

Volume 2:

EDITORS

Representatives of the empirical social and economic research community in the German Data Forum (RatSWD) between 2008 and 2011:

- Prof. Dr. Frank Kalter, University of Mannheim (Dec. 2008 – Jan. 2011)
- Prof. Dr. Eckhard Klieme, University of Frankfurt am Main (Nov. 2006 – Nov. 2008)
- Prof. Dr. Heiner Meulemann, University of Cologne (Nov. 2006 – Nov. 2008)
- Prof. Dr. Notburga Ott, Ruhr University Bochum – *Deputy Chairperson 2008 – 2011* (Dec. 2008 – Jan. 2011)
- Prof. Dr. Susanne Rässler, University of Bamberg (Dec. 2008 – Jan. 2011)
- Prof. Dr. Ulrich Rendtel, Free University of Berlin (Nov. 2006 – Nov. 2008; Dez. 2008 – Jan. 2011)
- Prof. Dr. Heike Solga, University of Göttingen – *Chairperson 2007 – 2008* (Nov. 2006 – Nov. 2008)
- Prof. Petra Stanat, Ph.D., Director of the Institute for Educational Progress (IQB, *Institut für Qualitätsentwicklung im Bildungswesen*) at Humboldt University of Berlin; and Free University of Berlin (Nov. 2006 – Nov. 2008; Dec. 2008 – Jan. 2011)
- Prof. Dr. Joachim Wagner, Leuphana University of Lueneburg (Dec. 2008 – Jan. 2011)
- Prof. Dr. Bettina Westle, University of Marburg (Nov. 2006 – Nov. 2008)

Representatives of the data production community in the German Data Forum (RatSWD) between 2008 and 2011:

- Roderich Egeler, President of the Federal Statistical Office (Dec. 2008 – Jan. 2011)
- Eckart Hohmann, President of the State Statistical Office of Hesse – *Deputy Chairperson 2007 – 2008* (Nov. 2006 – Nov. 2008; Dec. 2008 – Jan. 2011)

- Prof. Dr. Joachim Möller, Director of the Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung*) of the Federal Employment Agency (BA, *Bundesagentur für Arbeit*) (Oct. 2007 – Nov. 2008; Dec. 2008 – Jan. 2011)
- Sibylle von Oppeln-Bronikowski, Federal Statistical Office (Aug. 2008 – Nov. 2008)
- Walter Radermacher, President of the Federal Statistical Office (Nov. 2006 – July 2008)
- Prof. Dr. Hans Rattinger, President of GESIS – Leibniz Institute for the Social Sciences (*Leibniz-Institut für Sozialwissenschaften*) (Dec. 2008 – Nov. 2009)
- Uwe G. Rehfeld, German Federal Pension Insurance (RV, *Deutsche Rentenversicherung*) (Nov. 2006 – Nov. 2008; Dec. 2008 – Jan. 2011)
- PD Dr. Hilmar Schneider, Director of Labor Market Policy at the Institute for the Study of Labor (IZA, *Institut zur Zukunft der Arbeit*) (Nov. 2006 – Nov 2008)
- Prof. Dr. York Sure, President of GESIS – Leibniz Institute for the Social Sciences (*Leibniz-Institut für Sozialwissenschaften*) (Dec. 2009 – Jan. 2011)
- Prof. Dr. Gert G. Wagner, Director of the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) at the German Institute for Economic Research (DIW Berlin, *Deutsches Institut für Wirtschaftsforschung*) – *Chairperson 2008 – 2011* (Nov. 2006 – Nov. 2008; Dec. 2008 – Jan. 2011)

Business office of the German Data Forum (RatSWD) – scientific staff between 2008 and 2011:

- Denis Huschka (*Managing Director*)
- Claudia Oellers
- Dr. Gabriele Rolf-Engel

PREFACE BY THE GERMAN DATA FORUM (RATSWD)

"Valid and reliable data are the indispensable foundation for research in the social sciences and economics: they ensure that research is in line with contemporary realities and provide convincing arguments for actions by citizens, policy-makers, and business leaders."

This is the opening sentence of the 2001 evaluation report by the German Commission on Improving the Information Infrastructure between Science and Statistics (KVI, *Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik*), prepared on behalf of the Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*).[1] Ten years later, this statement still holds: the provision of valid and reliable data through a sophisticated and sustainable research infrastructure is an important task for both academic research and official statistical institutions, and will remain so in the years to come.

The German Data Forum (RatSWD) was founded by the BMBF in 2004. Its origins, however, date back to 1999, when the BMBF appointed the KVI to submit a comprehensive report with recommendations to improve the German research infrastructure for the social and economic sciences. This report, published in 2001, still constitutes the basis for a large part of the work performed by the German Data Forum. Although the Forum's tasks have gradually expanded, collaboration with the Research Data Centers and Data Service Centers, both of which have come into existence since the founding of the Forum, continues to form the backbone of its activities. However, since KVI report's publication, much has changed – and improved – in terms of data collection, preservation, access, and analysis. Thus, the time is ripe to systematically assess the progress made so far in Germany's information infrastructure and to discuss current challenges and future needs in the German, European, and international contexts.

One of the key tasks of the German Data Forum is to offer informed advice to the policy-makers, official data providers (especially state and federal statistical offices), and research funding bodies involved in building and

1 Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Ed.) (2001): Wege zu einer besseren informationellen Infrastruktur. Baden-Baden, 37 [own translation]. See also the documentation of the recommendations: "Towards an Improved Statistical Infrastructure. Summary Report of the Commission set up by the Federal Ministry of Education and Research (Germany) to improve the statistical infrastructure in cooperation with the scientific community and official statistics", in: Schmollers Jahrbuch 121 (3), 443-468.

running national and international statistical and research infrastructures for the social, economic, and behavioral sciences. To this end, the German Data Forum promotes dialog between, as well as within, academic research infrastructures and official statistical services, facilitating the extensive communication and coordination processes required to identify and prioritize needs and to develop sustainable concepts for nationwide and international data provision.

The German Data Forum has made a major step towards achieving these objectives by commissioning advisory reports from internationally recognized scholars in the social, economic, and behavioral sciences. The 68 advisory reports contained in this final volume, "Building on Progress – Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences," cover a broad range of topics. Their preparation began in the summer and autumn of 2008 with two international workshops at which authors exchanged ideas with members of the German Data Forum. The intensive discussions that took place there regarding current challenges and future demands facing Germany's research infrastructure revealed the need to include many more fields than initially planned. By 2010, the original number of about 60 advisory reports had increased to almost 70. Together, these advisory reports form a compendium of recent developments and data infrastructure needs in numerous fields – not only in the economic and social sciences, but to some extent also in the behavioral sciences. They touch on an array of methodological, ethical, and privacy issues related to data collection, preservation, and access, and take recent European and international developments into consideration. Although the German Data Forum (RatSWD) has attempted to make this a comprehensive overview, one cannot claim to have covered every issue of relevance to the German research infrastructure in the behavioral, economic, and social sciences; the infrastructure for public health research, for example, is not discussed here. Furthermore, since the majority of advisory reports in this publication were written in 2009, it should be noted that the information presented reflects the state of affairs at that point in time. In order to guarantee the timely publication and broad international scope of this work, all advisory reports were released as *RatSWD Working Papers* and placed online prior to their final publication here.

This compendium is published in two volumes divided into three main parts. The first part presents the German Data Forum's recommendations on the further development of the research infrastructure for the social, economic, and behavioral sciences. One of the overarching goals of these recommendations – and of the German Data Forum itself – is to create optimal infrastructural conditions in Germany for innovative research both at universities and independent research institutes and within the system of official statistics and government research institutes. This requires that researchers in all these institutions be equipped with the capabilities and tools they need to create and access databases in Germany and abroad. A second and equally important goal is to create and cultivate a research environment that allows young scholars, official researchers, and official statisticians with innovative ideas to achieve their full potential.

A vibrant, structurally sound, and highly productive research environment cannot be created using a top-down approach: the impetus must come from the research community itself. Scholars as well as official statisticians and researchers need formal procedures that promote competition and allow research entrepreneurship to flourish. The recommendations contained in Part I of this publication seek to facilitate these processes by communicating the needs of scientific researchers and statisticians to policy-makers and by promoting dialog among the various institutions involved

The second part of this publication, also contained in the first volume, provides "executive summaries" of all of the advisory reports, including detailed recommendations on how to meet current and future data needs. The summaries serve to provide the reader with a compact overview of current issues and needs in each research field.

The third part is comprised of the 68 advisory reports commissioned by the German Data Forum and makes up by far the largest section of this final volume. The advisory reports cover a wide range of fields in the social, economic, and behavioral sciences: economics, sociology, psychology, educational science, political science, geoscience, and communications and media research. Some reports focus mainly on substantive issues, some on survey methodology and issues of data linkage, some on ethical and legal issues, and others on the assurance of quality standards.

The third part begins with the assessment reports that address future demands likely to be placed on Germany's research infrastructure as well as the progress made since the first KVI report of 2001. One of the main topics dealt with here is the harmonization of European research infrastructures and possibilities for the permanent institutionalization of certain elements thereof. These are followed by reports on specific research fields, and on new data types and their potential applications in scientific research – for example, geodata, biodata, and transaction data. Many of these reports highlight recent advances in research methodology, such as the use of paradata ("data about data") and, for example, "qualitative methods" that can enrich quantitative data. Others are concerned with questions of data security and research ethics.

Further reports deal with specific fields: migration and demography; vocational competencies, education, and research; labor markets and the economy; the state, the family, and health; political and cultural participation; and the role of the media. Since these have been identified as crucial research fields for research infrastructure, key aspects of each are discussed in several advisory reports.

Most of the authors of advisory reports work in academic or governmental organizations in Germany, but important reports also came from private-sector experts and from European and US scholars. Because of the wide scope of expertise spanning many different fields and issues, this compendium is of value not only for policy-makers, research funding bodies, and institutional data providers, but indeed for anyone interested in gaining an overview of Germany's research infrastructure within its international contexts in the social, economic, and behavioral sciences.

The entire process of preparing this compendium for publication was driven by a sense of enthusiasm, which became particularly evident at the workshops and in numerous discussions among contributors and German Data Forum (RatSWD) members. We are grateful to everyone involved in bringing this publication to fruition.

First of all, we would like to thank the Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*) for their generous support through project funding (grant number 01 UW 0805). This provided the basis for intensive and systematic critical engagement with the topic of research infrastructure for the social, economic, and behavioral sciences, the results of which are presented in this publication.

Our profound gratitude goes to the authors of the advisory reports, who, through their comments and suggestions at the two workshops, greatly assisted in developing a differentiated overview of the current data landscape. Without this crucial input and their advisory reports, this publication would not have been possible.

Further thanks go to all the members of the German Data Forum (RatSWD) for their help in summarizing the findings of the advisory reports and in formulating recommendations based on these results. Special thanks go to Bruce Headey of Melbourne University, who provided numerous valuable suggestions and was responsible for writing the executive summaries.

This publication would never have been possible without the support of the German Data Forum (RatSWD) business office, specifically Patricia Axt, Lena Gond, Toby Carrodus, and Simon Wolff, who provided organizational, proofreading, and indexing assistance. Christoph Beck monitored the advisory reports and did the final proofreading and layout, all with exceptional commitment and careful attention to detail.

Further special thanks go to Deborah Anne Bowen and Jennifer Dillon for the editing of numerous English-language manuscripts and for translating several contributions into English. It was a large and sometimes difficult project, and they completed it with perseverance, commitment, and analytical expertise.

We are especially grateful to Claudia Oellers for her tireless dedication, immense effort, and the overall coordination of "Building on Progress – Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences."

The German Data Forum (RatSWD) adopted these recommendations at its 25th meeting on June 25, 2010, in Berlin.

Berlin, October 2010

Heike Solga
Chairperson of the German Data Forum (RatSWD) 2007 – 2008

Gert G. Wagner
Chairperson of the German Data Forum (RatSWD) 2009 – 2011

Denis Huschka
Managing Director of the German Data Forum (RatSWD)

RECOMMENDATIONS

For Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences

The big picture: Measuring the progress of societies

The importance of better data for the social, economic, and behavioral sciences is underscored by recent international developments. For decades, social progress was judged mainly by measures of economic performance; above all, by increases in gross domestic product (GDP). In 2009, the Commission on the Measurement of Economic Performance and Social Progress ("Stiglitz Commission")[1] published its report, which opens with the statement that "what we measure affects what we do." It sought to bring about a change in social and political priorities by advocating that greater emphasis be placed on measures of well-being and of environmental and economic sustainability.

The Stiglitz Commission's recommendations form a backdrop to this report.[2] Recommendation 6 in particular can serve as a unifying theme for our recommendations; we quote it below in full.

Both objective and subjective dimensions of well-being are important

"Quality of life depends on people's objective conditions and capabilities. Steps should be taken to improve measures of people's health, education, personal activities and environmental conditions. In particular, substantial effort should be devoted to developing and implementing robust, reliable measures of social connections, political voice, and insecurity that can be shown to predict life satisfaction."

In Germany, the Statistical Advisory Committee (*Statistischer Beirat)*, which advises the Federal Statistical Office, made the Stiglitz Commission's report the backbone of its recommendations for the next few years. The Committee writes:

1 Report by the Commission on the Measurement of Economic Performance and Social Progress, chaired by Joseph E. Stiglitz, Amartya Sen and Jean-Paul Fitoussi, http://www.stiglitz-sen-fitoussi.fr, and Stiglitz, J./Sen, A. and Fitoussi, J.-P. (2010): Mismeasuring Our Lives: Why GDP Doesn't Add Up. New York.

2 International organizations like the Organisation for Economic Co-operation and Development (OECD) are dealing with similar issues. For example OECD established the "Global Initiative on Data and Research Infrastructure for the Social Sciences (Global Data Initiative)" as part of its "Global Science Forum."

"Initiatives for the further development of national statistical programs – above all demands for new data – often come from supra- and international institutions: the EU Commission, the European Central Bank, the UN, OECD and the IMF. The Statistical Advisory Committee (*Statistischer Beirat*) believes that valuable key initiatives will come from the Stiglitz Commission and the theme *Beyond GDP* advanced by the European Commission. Official statistics, in cooperation with the scientific community, must react to these initiatives and their system of reporting must develop accordingly."[3]

We want to stress this point in particular: *Beyond GDP* will be a fruitful concept only if it is discussed and shaped collaboratively by government statistical agencies and academic scholars. As the Statistical Advisory Committee wrote:

"The Federal Statistical Office should take stock of the non-official data which may be available with a view to measuring the multi-dimensional phenomenon of *quality of life*. The development of statistical indicators should be undertaken in cooperation with the scientific community."[4]

Further, at the 12th German-French Council of Ministers in February 2010, President Sarkozy and Chancellor Merkel agreed on the *Agenda 2020*, which included joint work on new measures of social progress. This was yet another message that policy-makers are interested now more than ever in sound empirical evidence about a wide range of social and economic trends indicative of human progress or regress.

The following principles and themes are not intended to contribute directly to discussion of the Stiglitz Commission report or the initiative of the German-French Council of Ministers. But they do lay the groundwork for improved measurement of economic performance and social progress.

We strongly believe that recent improvements in survey methods and methods of data analysis hold promise of contributing substantially to improved measurement of social progress.

3 Statistischer Beirat (2010): Eckpunkte zur Weiterentwicklung der amtlichen Statistik in der 17. Legislaturperiode, p. 8 [own translation].

4 Ibid.

Background

This report is based on contributions by approximately one hundred social scientists[5] who were invited by the German Data Forum (RatSWD) to write advisory reports on key research issues and future infrastructure needs within their areas of expertise; their reports are published as part of this publication.[6] The number of experts who have contributed is even larger than it was when the predecessor of this report was published in 2001.[7]

The advisory reports cover a wide range of fields of the behavioral, economic, and social sciences: sub-fields of economics, sociology, psychology, educational science, political science, geoscience, communications, and media research. Some reports focus mainly on substantive issues, some on survey methodology and issues of data linkage, some on ethical and legal issues, some on quality standards. Most contributors work for German academic or governmental organizations, but important reports were also received from individuals in the private sector and from European and American academics. All had a focus on German infrastructural needs, but German as well as international contributors emphasized the importance of international collaborative and comparative research. All reports have been repeatedly peer reviewed; they have been discussed and amended at successive meetings and in working groups organized by the German Data Forum (RatSWD).

We first set out some *guiding principles* underlying the recommendations. The core of the recommendations is structured around a set of *principles* and *specific recommendations* regarding infrastructure for the social sciences.

Research in the fields of public health and social medicine is not reviewed. These are clearly such important and distinct fields that they require their own major reviews.

5 To avoid long-winded expressions, the term social sciences will be used in the remainder of this report to refer to all the behavioral, economic, educational, and social sciences, as well as related disciplines.

6 Some working papers that were not commissioned by the German Data Forum but that are of interest too are available on the homepage of the German Data Forum. See http://www.ratswd.de/eng/publ/workingpapers.html, especially Working Papers 50, 52, 79, 113, 131, 135, 137, 139, 141, 151, and 153.

7 Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Ed.) (2001): Wege zu einer besseren informationellen Infrastruktur. Baden-Baden. For an English translation of the recommendations, see: "Towards an Improved Statistical Infrastructure – Summary Report of the Commission set up by the Federal Ministry of Education and Research (Germany) to Improve the Statistical Infrastructure in Cooperation with the Scientific Community and Official Statistics." Schmollers Jahrbuch, 121 (3), 443-468.

Principles guiding the recommendations

Evidence-based research to address the major issues confronting humankind

The social sciences can and should provide *evidence-based research* to address many of the major issues confronting humankind: for example, turbulent financial markets, climate change, population growth, water shortages, AIDS, and poverty. In addressing some of these issues, social scientists in Germany need to cooperate with physical and biological scientists, with scholars in the humanities, and also with the *international community* of scientists and social scientists.

Competition and research entrepreneurs

In making recommendations about the future of research funding and research infrastructure, we recognize the importance of competition and research entrepreneurs. This may seem an unusual perspective. In many countries, including Germany, there is a tradition of centralizing research funding and infrastructure decisions. In our view, this is suboptimal. Science and the social sciences thrive on competition – competition of theory and ideas, and competition of methods, and competition of infrastructures.

Public funding of research infrastructure is certainly needed because research findings and research infrastructure are public goods and would be undersupplied in a free market.[8] But decisions should not be made in a centralized, top-down fashion – an approach that has the effect of stifling rather than promoting innovation. The experience of the last few years has demonstrated – notably in the field of empirical educational research – that many fruitful new ideas and initiatives can emerge from a decentralized structure that would almost certainly never have resulted from a "master plan." First of all, in Germany the National Educational Panel Study (NEPS) and the Panel Analysis of Intimate Relationships and Family Dynamics (pairfam) are worthy of mention. Both are new panel studies with a long time horizon.

The history of Germany's Research Data Centers and Data Service Centers illustrates the same point. All the Research Data Centers and Data Service Centers established in the last six years were the result of independent initiatives intended to meet distinctive research needs. The KVI laid the groundwork by initiating the establishment of the first four Research Data Centers and two Data Service Centers through central funding. All the later centers were bottom-up developments. The Federal Ministry of Education

8 See also UK Data Forum (2009): UK Strategy for Data Resources for Social and Economic Research. RatSWD Working Paper No. 131.

and Research (BMBF, *Bundesministerium für Bildung und Forschung*) and other initiatives provided some project funding for a few centers. What was crucial was the basic concept for the Research Data Centers, and that was developed by the KVI in its 2001 report.

It is true that the German Data Forum (RatSWD) later institutionalized this framework by establishing a Standing Committee of the Research Data Centers and Data Service Centers (*Ständiger Ausschuss Forschungsdaten-Infrastruktur des RatSWD)*. This committee helps the centers to work together and put forward common interests, but it does not initiate new centers. Indeed, the German Data Forum (RatSWD) is of the firm opinion that it should not do so. What is necessary is a common framework for new initiatives that aim to raise Germany's social science infrastructure to a higher level.

In this report we take some further steps towards developing a common framework for research infrastructure in the social sciences. In doing so, we bear in mind the increasing opportunities open to German researchers to contribute to European and international databases and projects, as well as to projects in Germany itself. We formulate some principles and highlight a range of concepts and ideas drawn from the advisory reports.

We do not make detailed recommendations about specific research fields or particular infrastructural facilities. This would run counter to our view that innovative research directions and new ideas develop mainly at the grass-roots of scientific and statistical communities. The advisory reports did include a large number of recommendations for promoting research in specific fields and on specific issues. A few of these recommendations are included in this report as examples, but in general our approach is to make recommendations about institutions and processes in which competition and research entrepreneurship can flourish. Nevertheless, by providing the advisory reports in this publication, we hope to give research funding bodies some idea about the budgets that may be needed if particular ideas are put forward by "scientific entrepreneurs."

The important role of younger researchers

Closely connected to the need for competition and innovation in science is the need to develop and foster excellent young researchers and ensure that they have sufficient influence in the research community for their ideas and research skills to flourish. It is, in general, true that a centralized research environment favors older, well-established researchers. Almost unavoidably, it is they who are appointed to the main decision-making positions. However eminent they are, their decisions may tend to favor well-established research topics and well-established methods. Innovation, on the other hand, is more likely to come from younger and mid-career researchers.

An important aim and principle underlying this report is to enhance the roles, influence, and opportunities of younger and mid-career researchers. They should be encouraged and given incentives to act as research entrepreneurs, competing amongst themselves as well as with older, established researchers to develop infrastructure. They may, however, have occasion to form research networks among themselves, and this should be supported.[9]

The need to encourage younger researchers is particularly clear in the official statistical offices. They need more freedom to improve official statistics by doing research. Further, with more research opportunities available, employment in official statistical offices will become more attractive to innovative post-doctoral researchers. Recommendations along these lines are developed under Theme 2 below, where we also suggest that it would be valuable to form new kinds of partnerships with private-sector data collection agencies for the performance of specific infrastructure tasks.

Social science requires improved theory and methods, not just more data

The main focus of this report is necessarily on research infrastructure and databases, but we want to highlight explicitly the importance of further improvements in social science theory and also in statistical and survey methods.

Social scientists in almost all fields complain about data deficiencies. The usually unstated assumption is that if only they had the right data, they could do the rest. This is self-serving, misleading and often used to defend a lack of pertinent results. Theory and method are also crucial, and new developments in these domains often go hand in hand with availability of new data sources. The advisory reports published in Part III of this compendium describe exciting new data sources available to social scientists, including data arising from "digitization," geo-referencing, and bio-medical tests. We make some recommendations about linkages between new and increasingly available data sources and potential improvements to social science theory and method.

Research ethics and data protection are of growing importance

Most data in the social sciences are of course data on human subjects. This means that principles of research ethics and privacy need to be observed. In Germany the right to privacy and data protection is enshrined in the Federal Data Protection Act (BDSG, *Bundesdatenschutzgesetz*), which protects individuals against the release of any information about their personal or material

9 See the editorial in *Science*, April 2, 2010, Vol. 328, 17, and letters in *Science*, August 6, 2010, Vol. 329, 626-627.

circumstances that could be used to identify them. Principles of research ethics, on the other hand, are not embodied in law but are dealt with by the scientific community through codes of ethics promulgated by their professsional associations.

Due to new technological developments, data protection and research ethics are of growing importance. Two of the themes outlined below reflect this importance.

Specific recommendations

In this section, we summarize insights arising from the advisory reports and subsequent discussions within the German Data Forum (RatSWD). We do this by presenting ten themes. Most of them represent general ideas and fairly abstract recommendations. We aim to encourage debate in the scientific and policy-making communities.

Theme 1: Building on success: Cooperation between official statistics and academic researchers

The German Data Forum's (RatSWD) current activities, as well as the present compendium, build on substantial achievements flowing from the 2001 KVI report. A major theme of that report was the need for improved cooperation between academics and the official statistical agencies, particularly in regard to making official datasets available for academic research. Initially, four Research Data Centers and two Data Service Centers were set up to provide academics and other users with access to official data files and with training and advice on how to use them. The original Research Data Centers are associated with the Federal Statistical Office, the Statistical Offices of the German *Länder*, the Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung*) of the Federal Employment Agency (BA, *Bundesagentur für Arbeit*), and the German Pension Insurance (RV, *Deutsche Rentenversicherung*). Since then, nine more Research Data Centers have been founded (June 2010) and, after being reviewed by the German Data Forum (RatSWD), they joined the group of certified Research Data Centers. It is also worth noting that, after their first three years, all the original Research Data Centers and Data Service Centers were formally reviewed and received positive evaluations.

One of the advisory reports provided for this review offered the observation that, as a result of the Research Data Centers, Germany went from the bottom to the top of the European league as an innovator in enabling scientific use of official data. It has also been suggested that the Research Data Centers have had benefits that were not entirely foreseen, in that civil servants and policy advisors are increasingly using research-based data from Research Data Centers to evaluate existing policy programs and plan future programs. Civil servants have more confidence in academic research findings knowing that they are based on high-quality official data sources and that the researchers have received advice on how to use and interpret the data.

Official data files have also become more readily available for teaching in the higher education sector as a result of the recommendations of the 2001 KVI report. CAMPUS-Files, based on the Research Data Center files, have been created for teaching purposes and are widely used around the country.

It is important to note that the Research Data Centers have made good progress in dealing with a range of privacy and data linkage concerns that loomed large ten years ago. Particular progress has been made in linking employer and employee data. Research Data Centers have also, in some cases, been able to develop procedures for enabling researchers to have remote access to data once they have worked with officials in the relevant agencies and gained experience in using the data.

Partly due to the progress already made, but mainly due to technological and inter-disciplinary advances, new and more complicated issues relating to data protection, privacy, and research ethics keep arising. Some of these issues emerge because of the increasing availability of types of data that most social scientists are not accustomed to handling, including biodata and geodata. Other issues emerge due to the rapidly increasing sophistication of methods of record-linkage and statistical matching. These issues are discussed in more detail under Theme 8 ("Privacy") and Theme 9 ("Ethical Issues").

Based on these considerations, it is recommended that work continues towards providing a permanent institutional guarantee for the existing Research Data Centers. In the best-case scenario, Research Data Centers that belong to the statistical offices and similar institutions should be regulated by law. At present, the costs of Research Data Centers are borne by the agencies that host them, and users are usually not required to pay more than a nominal fee. In fact, we believe that this is the best way to run the centers because it ensures maximum use of official data. In the event that funding issues pertaining to the Research Data Centers arise in public and policy discussions, it is recommended that cost-sharing and user-pays models be investigated.

It is recommended that methods of obtaining access to a number of important databases that are still de facto inaccessible to researchers be investigated. Examples include criminal statistics and data on young men collected through the military draft system.

In particular, it is recommended that methods of permitting remote data access to Research Data Center files continue to be investigated.

It is recommended that the microdata of the 2011 Census – the first Census in almost 30 years – should be accessible and analyzed in-depth by means of concerted efforts on the part of the scientific community and funding agencies for academic research.

It is recommended that peer review processes be established and sufficient resources allocated to provide "total quality management" also of the data produced by government research institutes (*Ressortforschungseinrichtungen*).

We are in favor of a coordinated and streamlined process. We take a critical view, however, of the current trend towards increasing numbers of evaluations: this is neither efficient nor beneficial to the scientific content.

It is recommended that data providers in Germany collaborate more closely with the European Union's statistical agency, Eurostat.

Theme 2: Inter-sector cooperation: cooperation between academic research, the government sector, and the private sector

A major theme of the 2001 KVI report was the need for greater cooperation and collaboration among academic social scientists, official statistical agencies, and government research institutes (*Ressortforschungseinrichtungen*). Since then, it has become clear that in many areas of data collection and analysis, official institutes and academic organizations can form effective partnerships. Such partnerships would be strengthened if younger researchers in all areas were permitted more independent roles.

Much remains to be done. Academic research teams and official statistical agencies and research institutes probably still do not always realize how much they have to gain from collaboration. But each side must pay a price.

Academics need to understand and respect the social, political, and accountability environments in which official agencies operate. The official agencies (including the ministries and parliaments behind them), for their part, need to be willing to give up monopoly roles in deciding what specific data to collect and disseminate.

A strong case can be made that the improved level of cooperation that has been seen in recent years between academic social scientists and official statistical agencies and authorities should now be extended to include the private sector as well. Many large social and economic datasets, especially surveys, are collected by private-sector agencies. Since these agencies operate in a competitive market, they need a reasonably steady and secure flow of work in order to be able to make the investments required to maintain high-quality standards in data collection and documentation. Public-private partnerships may be desirable for initiating, attracting funding for, and continuing long-term survey-based projects. The UK's Survey Resources Network has experience in these ventures and may be able to offer useful guidance. Last but not least, a permanent flow of sufficient amounts of work is necessary to ensure competition between private fieldwork firms.

There are many opportunities for methodological investigations carried out in cooperation among academics and government and private-sector survey agencies. One clear example is investigation of the advantages, disadvantages, and possible biases of mixed-mode surveys. Mixed-mode surveys, which are more and more widely used, involve collecting data using a variety of methods, for example, personal interviews, telephone, mail, and Internet. In practice, respondents are commonly offered a choice of method, and the choice they make may affect the evidence they report.

Leaving aside cooperative ventures with public sector and academic clients, it is clear that private sector fieldwork agencies already collect a vast amount of market research data of great potential value to academic researchers.

The potential of market research data for secondary analysis lies mostly in the fields of consumption patterns and media usage. The German market research industry is huge – it has an annual turnover of more than two billion euros – and over 90 percent of its research is quantitative. However, samples are often highly specialized; telephone interviewing is the most common mode of data collection; and data documentation standards are not as high as academic social scientists would wish. However, secondary data analyses seem to be worthwhile – last but not least as a kind of quality control for these data. Clearly, too, the commercial clients for whom data are collected would have to give permission for secondary analysis. The data would have to be anonymized not only to protect individuals, but also to protect commercially sensitive information about products.

In addition, transaction data (e.g., about purchasing behavior) that is generated by commercial firms can be of interest for scientific research. In this case, anonymization is extremely important. The German Data Forum (RatSWD) makes no specific recommendation about this issue beyond the view that recognition of market research data and transaction data merits consideration in the scientific and statistical communities.

Theme 3: The international dimension

The main focus of the detailed advisory reports contained in this publication is of course on German social science infrastructure and research needs, but the international dimension is critical too. Plainly, many of the problems with which social scientists as well as policy-makers deal transcend national borders; for example, turbulence in financial markets, climate change, and movements of immigrants and refugees. Furthermore, international comparative research is an important *method of learning*. Similar countries face similar issues, but have developed diverse and more or less satisfactory policy responses. To do valuable international comparative research, researchers usually need to work with skilled foreign colleagues.

International data collected by the EU and other supra-national organizations have important strengths but also important limitations. The data are at least partly "harmonized" and cross-nationally comparable. Generally, however, data coverage is restricted to policy fields for which international organizations have substantial responsibility. Data are much sparser in areas that are still mainly a national-level responsibility. Furthermore, the needs of policy-makers, for whom the data are collected, do not exactly match the needs of scientists.

For example, policy-makers require up-to-date information, whereas scientists give higher priority to accuracy. Policy-makers are often satisfied with use of administrative and aggregate data and accept "output harmonization," whereas scientists favor the collection of micro-level survey data and prefer "input harmonization," that is, data collection instruments that are the same in each country.

With regard to international cooperation, which still raises some difficult problems for German research – in part because of legal restrictions on data sharing – we recommend that a working group be set up by the German Data Forum (RatSWD) to find ways of making German official statistics available *in a simple manner* as anonymized microdata to reliable foreign research institutes.

There are several cooperative European ventures that will be discussed in an open and constructive manner. These include a new European household panel survey under academic direction, Europe-wide studies of birth and other age cohorts, and a Europe-wide longitudinal study of firms. It would also be of great benefit to comparative European research if access to micro-level datasets held by Eurostat could be improved. Ideally, these data would be made available by virtual remote access, with appropriate safeguards to ensure data security.

It is noted that, following a British initiative, an International Data Forum (IDF) has been proposed. Along the lines of the UK Data Forum and the German Data Forum (RatSWD), this body would aim to bring together academic researchers and official statistical institutes, including international organizations like the OECD. The plan is currently being developed via an Expert Group set up under the auspices of the OECD. It is recommended that Germany participate in this and related initiatives through the German Data Forum (RatSWD) and possibly other bodies.

Finally, it is clear that the academic data providers are not very well organized at the international and supra-national level. Most surveys are conducted only within the bounds of a country at each wave. Notable exceptions are international survey programs like the European Social Survey (ESS) and the Survey of Health, Ageing and Retirement in Europe (SHARE), and networks of archives like the Council of European Social Science Data Archives (CESSDA), "Data Without Boundaries," and the "Committee on Data for Science and Technology (CODATA)." We recommend that the academic sector consider setting up an independent organization to represent its interests at the European and worldwide levels. This academic organization would be one of the partners in the international bodies that are likely to be established following the OECD initiative.

Theme 4: Data on organizations and "contexts"

It is clear that, since the 2001 KVI report, a great deal of progress has been made in Germany to improve academic researchers' access to firm-level data – that is, to data on employers and employees. These are high-quality data mainly collected in official surveys; firms are required to respond and to provide accurate information about the firm and its employee structure. Most statistical data of this kind are now available from Research Data Centers. Progress has been made on issues of data linkage, while protecting confidentiality, with the result that it is now often possible for researchers to link data from successive official surveys of the same firm. It is not, however, at present legally possible to link surveys of German firms to international datasets. This would be a desirable development, given that many firms now have global reach.

Progress made in improving access to data on business organizations points the way towards what needs to be achieved in relation to the many other organizations and contexts in which people live and work. Individual citizens are typically linked to multiple organizations: firms, schools, universities, hospitals, and of course their households. Linking data on these organizations and contexts with survey data on individuals would be desirable. Yet technical problems concerning algorithms for linking data are certainly easier to solve than the important questions regarding research ethics and data confidentiality that are in need of discussion.

At present, then, there are no German datasets that have adequate statistical information on all the organizations in which individuals operate. Data thus need to be collected in surveys on persons and activities in multiple organizations, and where possible, linked to data about the organizations themselves. This could potentially be achieved by (1) adding additional questions about organizational roles to existing large-scale surveys, perhaps even including the large sample of the German Microcensus, as well as by (2) linking existing survey datasets on these organizations with Microcensus data and other surveys on individuals and households.

A very special kind of new data type is information about historical contexts, which can be linked to time series data or microdata with a longitudinal dimension. The European Social Survey (ESS), for instance, provides such a databank. It contains information on minor and major historical events, and is updated on a daily basis. It is worthwhile to think about offering such a centralized historical database to the research community at large.

Government and research-based statistical data on political and civil society organizations are in short supply in Germany. In many Western countries, evidence about political parties – the most important type of political organization – is regularly obtained from national election surveys. Election surveys are also the main source of evidence on mass political participation. We want to note that in Germany, there is no guaranteed funding for election surveys, although a major election project (GLES, *German Longitudinal Election Study*) is currently being undertaken. This project could develop into a national election study.

Several of the advisory reports prepared for the German Data Forum (RatSWD) discussed detailed practical ways of realizing these possibilities. The German Data Forum (RatSWD) recommends that funding agencies consult these advisory reports when assessing specific applications to conduct organizational research.

Theme 5: Making fuller use of existing large-scale datasets by adding special innovation modules and "related studies"

Many of the advisory reports recommended that fuller use could be made of existing large-scale German datasets (such as ALLBUS) by adding special innovation modules, thereby creating greater value for money. Suggestions were made both for *special samples* and for *special types of data* to be collected. In all cases, it was suggested that the particular benefit of adding modules was that the underlying survey could serve as a national benchmark or *reference dataset* against which the new, more specialized data could be assessed.

The availability of a reference dataset enables researchers to obtain a more contextualized understanding of the attitudes and behaviors of specific groups. Conversely, the availability of detailed and in-depth evidence about subsets of the population can strengthen the causal inferences that analysts of the main reference dataset are able to make.

The advisory reports covering international and internal migration document substantial data deficits, which, it is suggested, could be largely overcome by adding special modules to existing longitudinal surveys (such as the SOEP). It has been pointed out that existing datasets do not allow researchers to track the life-cycles of migrants over long periods. This is particularly a problem in relation to highly skilled migrants, a group of special interest to policy-makers. Migrant booster samples, added to existing large-scale surveys, would largely overcome the problem.

Reports written by experts in other fields made similar recommendations. For example, it was suggested that data deficits relating to pre-school education and vocational education and competencies could be partly overcome by adding short questionnaire modules to ongoing surveys.

It is more or less conventional in the social sciences to collect exploratory qualitative data – for example, open-ended interviews – to develop hypotheses and lay the basis for quantitative measures prior to embarking on a large-scale quantitative project. It is suggested that this sequence can also sensibly be reversed. Once a quantitative study has been analyzed, individuals or groups that are "typical" of certain subsets can be approached with a view to conducting qualitative case studies. The researcher then knows precisely what he/she has a "case of." Extended or in-depth interviews can then be undertaken to understand the decisions and actions that subjects have taken at particular junctures in their lives, and the values and attitudes underlying their decisions.[10]

In an advisory report it is proposed that innovation modules using "experience sampling methods" be added to existing large-scale surveys. Again, the procedure would be to approach purposively selected respondents, representing sub-sets of the main sample, and ask them to record their answers to a brief set of questions (e.g., about their current activities and moods) when a beep alerts them to do so.

Theme 6: Openness to new data sources and methods

Advisory reports prepared for the German Data Forum (RatSWD) highlighted the potential of several exciting new sources and methods of collecting data. We want to mention some of these new technical possibilities, but without making specific funding recommendations. We do, however, want to stress that Germany needs to develop funding schemes involving use of these new data sources and data collection methods.

Digitization

Survey data and publications in the social sciences have generally been available in digital form for some time. Thanks to the grid technology promoted by the Federal Ministry of Education and Research (BMBF) as part

10 It is important to address the privacy and ethical implications of approaching survey respondents for additional interview data. Clearly, the respondents must be asked for explicit consent to link the data sets.

of the D-Grid Initiative, it is now possible to work with these digital data on a much larger scale and – more crucially – in new research contexts, thus enabling completely new approaches in empirical research. Yet the possibilities offered by grid technology have not been exploited in the social sciences to any notable extent.

Large quantities of data that would be of interest in social sciences research are generated by the Internet (particularly online social networks) and by the use of mobile phone, GPS, and RFID technologies. To date, researchers have drawn little benefit from such data, as numerous questions concerning access and data confidentiality remain unclarified. A few initiatives have been undertaken. For example, the networking site Facebook reports that social scientists in all English-speaking countries are analyzing messages posted on the site each day to assess changes in moods and perhaps happiness levels.

However, it will not be possible to make substantial progress until access and privacy issues are resolved. The German Data Forum (RatSWD) notes that the UK's Economic and Social Research Council (ESRC) has set up an Administrative Data Liaison Service to deal with similar issues by linking academics to producers of administrative data.

Geodata – A multifaceted challenge

Most of the data used in the social sciences have a precise location in both space and time. While geodata are used widely in geography and spatial planning, this is generally not the case in the social sciences. Spatial data from various sources (e.g., concerning urban development or the weather) can readily be combined via the georeferences of the units under investigation. This makes georeferenced data a valuable resource both for research and for policy advice and evaluation. While administrative spatial base data have been widely available for Germany for a long time, there has been an enormous increase in recent years in the supply of spatial data collected by user communities (e.g., OpenStreetMap) and private data providers (e.g., Google Street View). Furthermore, remote sensing data (aerial photos or satellite data) have become more important. These data are provided by different sources, which makes it important to launch geodata infrastructure projects that bring together different geodata sets. It must be emphasized that data security is of high importance for this type of data; issues of personal rights are particularly sensitive.

Closely related to geodata are data for regions, which can be defined as areas as large as a German *Land* or as small as a municipality. Regional data have been available for many years and have been used for cross-regional investigations and as context variables in studies investigating the behavior of persons or firms. Access to many datasets at various levels of regional aggregation is straightforward in Germany through the use of cheap CDs/DVDs and the Web.[11] The main challenge is to offer access to geodata in ways that allow easy combination with other data. Both current and older data need to be made available to allow for longitudinal studies. Furthermore, data on households, and buildings should be entered with a direct spatial reference; this is especially important for the forthcoming 2011 Census.

An important recommendation for the future is to intensify collaboration between social science researchers and researchers in institutions in the currently rather segregated areas of geoinformation and information infrastructure. Thus, the German Data Forum (RatSWD) will set up a *working group* on geodata and regional data with a view to bringing the different data providers and users together.

Biodata – Research incorporating the effects of biological and genetic factors on social outcomes

In recent times, greater attention has been paid in the social sciences to biomedical variables, including genetic variables that influence social and economic behaviors. Many opportunities, and some serious risks, exist in this growing research field. Historically, social scientists have received no training in biomedical research and are unlikely to be aware of the possibilities. Certainly, they have little knowledge of appropriate methods of data collection and analysis. It is under discussion whether the German Data Forum (RatSWD) will set up a *working group* with a view to positioning German social scientists to be at the forefront of developments. The group would need to include biologists and medical scientists, as well as social scientists and – equally important – not only data protection specialists but also ethics specialists. In addition, one issue that such a working group would have to address is the difficulty that researchers who are working at the interface of the social and biomedical sciences currently have in attracting funding.

A role model for this kind of multidisciplinary data collection may be found in the SHARE study, which has already conducted several pilot studies, collecting biomedical data from sub-sets of its European-wide sample. It has been shown that, with adequate briefing, medically untrained interviewers can do a good job of getting high-quality data in biomedical surveys, without a significant increase in non-participation or drop-out rates.

Virtual worlds for macro-social experiments

Advocates of the use of computer-generated “virtual worlds” (such as “Second Life”) for social science research believe that they offer the best

11 http://www.geoportal.bund.de, http://www.raumbeobachtung.de, http://www.regionalstatistik.de. [Accessed on: August 7, 2010].

vehicle for developing and testing theories at a “macro-societal” level. Many of the problems facing humanity are international or threaten whole societies: climate change, nuclear weapons, water shortages, and unstable financial markets, to name just a few. By setting up virtual worlds with humans represented by avatars, it is possible to conduct controlled experiments dealing with problems on this scale. The experiments can be run for long periods, like panel studies, and they can allow for the involvement of unlimited numbers of players. They pose no serious risk to players and avoid the ethical issues that limit many experiments.

Advocates of macro-social experiments recognize that initial costs are high, but claim that the worlds they create hold the prospect of eventually being self-funding, paid for by the players themselves.

Theme 7: Data quality and quality management

An increasingly important role is being played by questions related to the quality of (1) available measurement instruments, and (2) documentation required to facilitate secondary analysis of existing datasets.

Experts in several areas in their advisory reports made the point that a fairly wide range of measurement instruments were available to them, but that researchers would benefit from guidance in assessing their comparative reliability, validity, and practicality in fieldwork situations. In the advisory reports, it was suggested that something like a *central clearing house* was needed with a mandate to assess and improve standards of measurement. It was noted that the recent founding of the Institute for Educational Progress (IQB, *Institut zur Qualitätsentwicklung im Bildungswesen*) could serve as a model for additional subfields.

The IQB was launched at a time when the poor performance of German students in standardized international tests led to increased concern with measuring learning outcomes. The institute is measuring the performance of representative samples of students in the 16 German *Länder*, and will also be available to serve as a source of advice on measurement issues

A related but somewhat separate concern mentioned in several advisory reports is the poor quality of documentation provided for many surveys and other datasets that, in principle, are available for secondary analysis. It appears that academic data collection has much to learn in this respect from official statistical agencies, which generally adhere to high standards in data collection and documentation.

In thinking about data storage and documentation, a distinction should probably be drawn between two types of academic projects: those that are of interest only to a small group of researchers and those that are of wider interest. A mode of self-archiving (self-documentation) should suffice for the former type, although even here minimum satisfactory uniform standards need to be established. The latter type should be required to meet high professional standards of documentation and archiving (see Theme 10).

To a large extent, improvement of survey data documentation is a matter of adopting high *metadata standards*. These are standards relating to the accurate description of surveys and other large-scale datasets that need to be met when data are archived. Historically, researchers paid little attention to the quality of metadata surrounding their work; archiving was left to archivists. This mind-set is changing. There have been rapid advances in the development and implementation of high-quality metadata standards, standards which apply to datasets throughout their life cycle from initial collection through to secondary use.

An important source of survey metadata is the information collected in the recruitment of survey participants and in the actual survey itself concerning survey methods, the administration of the survey, and, when applicable, geographic location. These data, sometimes termed *paradata*, are typically recorded by interviewers and stored at the surveying institute. The data are valuable for analyzing problems of survey non-response and for assessing the advantages and disadvantages of different data collection modes. Paradata can be used for "continuous quality improvement" in survey research. It is recommended that efforts be made to standardize and improve the quality of paradata collected by public and private-sector survey agencies. The European Statistical System has published a handbook on enhancing data quality through effective use of paradata.

In Germany, the Research Data Centers have taken the lead in trying to improve current standards of documentation. Based on their experience, it appears that there are two internationally acceptable sets of metadata standards – the Data Documentation Initiative (DDI) and the Statistical Data and Metadata Exchange (SDMX) Standard – which could be more widely used in Germany. Adoption of these standards requires the establishment of a IT infrastructure compatible with the industry standard for Web services. This infrastructure can then facilitate the management, exchange, harmonization, and re-use of data and metadata.

We would like to highlight in particular one potential means of improving documentation: the use of a unique identifier for datasets (e.g., a digital object identifier or DOI). Unique identifiers for particular measurement scales (e.g., the different versions of the "Big Five" inventory) could possibly also be helpful (see also Theme 10 below).

The need for high-quality metadata appears even more pressing when recalling that many Internet users who are not themselves scholars are making increased use of these data for their own analyses. Results generated by lay users are especially likely to be skewed or misleading if the strengths and limitations of the data are described inadequately or in jargon a layperson could not be expected to understand.

Theme 8: Privacy issues

This section deals with privacy issues, particularly those that arise due to increasingly sophisticated methods of data linkage. *Record linkage* refers to the possibility of linking up different datasets containing information about the same units (e.g., individuals or firms). Linkages may be made, for example, between different surveys or between survey data and administrative data. Normally, datasets can only be linked if a common identifier is available. However, linkage can be achieved by means of "statistical matching" when datasets either do not contain the same identifiers for particular individuals or datasets of similar yet not identical units.

When an individual or firm consents to take part in a specific research project, the commitment and its limits are usually reasonably clear. But what is the situation if researchers acquire the permission of respondents to link a file obtained for this specific project to other files about the respondent, which, for example, contain information about her employer, tax files, health, or precise geographical location? Clearly, such linked data are of immense value to researchers, both in conducting basic scientific research and in providing policy advice. While it is clear that such linking may only take place with the explicit consent of the concerned individuals, how "explicit" must this consent be? Do the individuals whose data are being linked need to provide specific consent prior to each new linkage?

The advisory reports written for the German Data Forum (RatSWD) express a wide variety of views on this matter. While some legal experts have described such data linking as a breach of law, we believe that these problems could be best resolved by passing legislation that would require researchers to observe a principle of "research confidentiality" (*Forschungsdatengeheimnis*). This legislation, which was recommended by the KVI in 2001, would require that if authorized researchers obtained knowledge of the identity of their research subjects – even by accident – they would be obliged not to reveal the identities under any circumstances. Most important, the act would prevent both police and any other authorities from seizing the data. When pushing forward the issue of "research confidentiality," it will be important to refer to the European legislation.

A further proposal discussed in one of the advisory reports, is for data stewards (*Treuhänder*) to be appointed for the purpose of protecting the privacy of research subjects. Data stewards would be responsible for keeping records of the identity of subjects and would only pass data on to researchers for analysis with the identifying information removed.

A more general recommendation given in the reports is that a "National Record Linkage Center" be set up with high security precautions to cover all fields in which record linkage is an issue. This has been proposed to avoid the duplication that would occur if each branch of social science made its own separate efforts. The German Data Forum (RatSWD) expressly abstains from making any specific recommendations, but believes that the mentioned problems and possibilities are worth detailed consideration.

Theme 9: Research Ethics

This theme deals with two separate sets of ethical issues: the ethics of research using human subjects, and the ethics of scientists in publicizing their results.

Research using human subjects

The need to define and enforce ethical standards in research using human subjects has always been urgent and has become more so in view of the increasing availability of new types of data highlighted in this report: administrative and commercial data, data from the Internet, geodata, and biodata.

In practical terms, Germany does not yet have a detailed set of ethical requirements specifically designed to protect individuals who take part in research projects in the social sciences – a field typically concerned, of course, with the administration of surveys, and not human experiments. However, all researchers have to abide by the requirements of the Federal Data Protection Act. Additionally, the main professional associations in sociology and psychology have issued ethical guidelines, but these mainly affect behavior towards peers, rather than towards research subjects.

A review of ethics procedures in the UK and the US was undertaken by an advisory report to see if they offered useful examples for Germany. British procedures appear worth consideration; US procedures are perhaps too heavily geared towards the natural sciences.

In the UK, beginning in 2006, the *Economic and Social Research Council* (ESRC), which is the main funding body for academic research, forced universities whose researchers were seeking funding from ESRC to set up ethics committees. In practice, committees have been put in place in all universities, usually operating at the departmental or faculty level and not always on a university-wide basis. The committees are required to implement six key principles, four of which protect human subjects. Subjects have to be fully informed about the purposes and use of the research in which they are participating; they have the right to be anonymous; the data they provide must remain confidential; participation must be voluntary, and the research must avoid harm to the subjects.

The principle of "avoiding harm" is particularly important in view of the increasing availability of Web data, geodata, and biodata. "Avoiding harm" appears to be a principle of more practical relevance than the principle of "beneficence," which German social scientists, borrowing from the biological sciences, have sometimes incorporated into ethical guidelines.

Above all, given that research is conducted increasingly on the basis of international exchange, and research data are exchanged between different countries and national research institutions, it is of growing importance that respondents be able to rely on researchers to handle their data responsibly. Due to differences in national data security regulations as well as in research ethics standards, this is a difficult task, which, at worst, can hinder research. However, universal data protection rules are desirable, but extremely unlikely. Thus, it is important that, at a minimum, the scientific and statistical expert communities seek to foster the development of ethical standards which are then voluntarily adopted by those engaged in research and statistical work.

Scientific responsibility in publicizing results

A key set of ethical issues surrounds the responsibility of scientists in publishing and publicizing their results. In a recent editorial in *Science*,[12] it is noted that "bridging science and society" is possible only if scientists behave properly – that is, in accordance with scientific standards. The editorial mentions not just the need to avoid obvious scientific misconduct relating to data fraud or undisclosed conflicts of interest, but also the importance of avoiding "over-interpretation" of scientific results.

It is worth noting that many economists appear to believe that over-interpretation (by simplifying results) is necessary if a scientist wants to reach the general public. The former Federal President of Germany, Mr. Köhler, an economist by training, appeared to endorse this approach by calling for social scientists to announce "significant" findings without burying important results under too many details.[13]

We believe that it would *not* be wise for social scientists to take this advice, precisely because scientific results often become the subject of contentious public policy debates. Empirical results *can* have the effect of making policy debates more rational, but only if the assumptions underlying research and shortcomings that mar obtained results are communicated honestly. It is a duty of the scientific community to promote this type of honesty.

---

12 *Science*, February 19, 2010, Vol. 327, 921.

13 Köhler, H. (2009): Ein Kompass für die Gesellschaft. Grußwort von Bundespräsident Horst Köhler beim Festakt zum 40jährigen Bestehen des Wissenschaftszentrums Berlin für Sozialforschung am 17. Februar 2009 in Berlin. In: http://www.bundespraesident.de/Anlage/original_652450/Grusswort-beim-Festakt-zum-40jaehrigen-Bestehen-des-Wissenschaftszentrums-Berlin-fuer-Sozialforschung.pdf. [Accessed on November 17, 2010].

Theme 10: Giving credit where credit is due

A key principle of these recommendations is "*to give credit where credit is due.*" This principle[14] should apply to efforts at developing the research infrastructure just as much as to academic authorship in publications. In general, valuable new infrastructural initiatives will only be launched if the staff of infrastructures under academic direction, of official statistical agencies – and perhaps of private-sector organizations that collect and provide data as well – feel recognized and rewarded for undertaking this important work. Junior and senior staff of all types of organizations need to be clearly recognized for their important contributions.

Existing academic conventions about "authorship" are not entirely satisfactory, nor are "science metrics" that evaluate the output of researchers, universities, and research institutes. In a recent article in *Nature*[15] it is suggested:

> "Let's make science metrics more scientific. To capture the essence of good science, stakeholders must combine forces to create an open, sound and consistent system for measuring all the activities that make up academic productivity. … The issue of a unique researcher identification system is one that needs urgent attention."

Effective partnerships and joint investments by academic research institutes, official statistical agencies, and private fieldwork organizations occur despite seriously inadequate incentives and recognition for the creation and maintenance of research infrastructure. However, in order to make such collaborations more than rare events, the "rules of the game" must be changed. The establishment and running of infrastructure resources like biobanks, large social surveys, and the Scientific Use Files of official data must be rewarded more adequately than at present. This applies to official statistics, public administrations, private organizations, and the entire scientific system.

The German Data Forum (RatSWD) sees itself as one of the key players in promoting discussion and proposing effective steps on this issue. Here we want to mention two instruments that might help to ensure that credit is given where it is due.

First, the establishment of a system for the persistent identification of datasets (e.g., the DOI system) would not only allow easier access to data, but also make datasets more visible and easily citable, thereby enabling the authors/compilers of the data to be clearly recognized. Even particular measurement "devices" (e.g., specific scales for the "Big Five" inventory) might be identified and citable by unique identifiers. And a digital object identifier makes it easier to see the links between a scholarly article, the relevant datasets, and the authors/compilers of the datasets. There are already

14 *Nature*, December 17, 2009, Vol 462, 825.
15 *Nature*, March 25, 2010, Vol. 464, 488-89.

some organizations that have assigned DOIs to datasets (e.g., CrossRef and DataCite).

Second, the issue of a unique researcher identification system is equally important and needs urgent attention. The recent launch of Open Researcher Contributor ID (ORCID) looks particularly promising. The use of a unique researcher ID makes the scientific contributions of each individual researcher who works on a dataset clearly visible.

Concluding remarks

In Germany, there are several organizations for funding scientific research. Due to this "fragmented" funding environment, some policy-makers, government officials, and senior researchers believe that a more centralized organization would perform better. However, we at the German Data Forum (RatSWD) disagree. We are convinced that competition opens up more space for new ideas than would be available under a centralized system.

Even though we do not support centralized organization of research, we nevertheless recognize an increasing need to provide long-term funding to establish and run large-scale social science infrastructure. Fortunately, the academic community, official statistical agencies, and government research institutes are thinking more than ever before about how to reorganize and finance infrastructure in research and statistics. So, for example, the German Council of Sciences and Humanities (WR, *Wissenschaftsrat*), and Germany's Joint Science Conference (GWK, *Gemeinsame WissenschaftsKommission*) have working groups underway that are considering matters of research infrastructure.[16]

The discussions in these working groups have already made obvious that not only Research Data Centers and data archives but also more and more libraries – university and research institute libraries as well as centralized specialist libraries (*Fachbibliotheken*) – are an important part of the research infrastructure, providing crucial data documentation and access services. The Federal Archive (*Bundesarchiv*) could also play a specific role. Nothing is

16 These are (in 2010) the "Research Infrastructure Coordination Group (*Koordinierungsgruppe Forschungsinfrastruktur*)" and the "Working Group on a Research Infrastructure for the Social Sciences and Humanities (*Arbeitsgruppe Infrastruktur für sozial- und geisteswissenschaftliche Forschung*)" of the German Council of Science and Humanities (WR, *Wissenschaftsrat*) as well as the "Commission on the Future of Information Infrastructure (KII, *Kommission Zukunft der Informationsinfrastruktur*)" of the Joint Science Conference by the Federal and Länder Governments (GWK, *Gemeinsame Wissenschaftskonferenz des Bundes und der Länder*).

settled yet. However, it is time to find a new and appropriate division of labor among these institutions.

Many approaches will no doubt be considered, but in our view it is preferable to develop *principles* for funding and managing research infrastructure, rather than to attempt the almost impossible task of formulating a detailed *master plan*.

The German Data Forum (RatSWD) is itself neither a research organization nor a funding organization. It exists to offer advice on research and data issues. This places it in an ideal position to moderate discussions and help find the most appropriate funding arrangements.[17]

17 See also the "Science-Policy Statement on the Status and Future Development of the German Data Forum (RatSWD)" by the German Council of Science and Humanities (WR, *Wissenschaftsrat*). Schmollers Jahrbuch, 130 (2), 269-277.

EXECUTIVE SUMMARIES*

* The executive summaries have been compiled and edited by Bruce Headey. These summaries are not necessarily identical with those in the expert reports.

TOWARDS AN IMPROVED RESEARCH INFRASTRUCTURE FOR THE SOCIAL SCIENCES: FUTURE DEMANDS AND NEEDS FOR ACTION

1. Providing a Permanent Institutional Guarantee for the German Information Infrastructure *(Johann Hahlen)*

Background and current issues

Proposals relating to an institutional guarantee for social science infrastructure should be soundly based on an understanding of existing constitutional, legal, and other requirements in Germany. In particular, the Federal Constitution enshrines strict rights to "informational self-determination."

Following the last 2001 KVI report, and taking account of legal constraints, four Research Data Centers and two Data Service Centers were set up. These centers take responsibility for "anonymizing" data and are themselves organized on a subject-matter basis. Formal evaluations of these centers have been positive.

Start-up funds for the centers came mainly from the Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*) with the intention that other relevant institutions would provide subsequent funding. A permanent funding solution is now required, and while users can cover some of the costs it is important that the prices they are charged do not deter research.

Recommendations

(1) Each of the four institutions that houses a Research Data Center should provide financial support for its "own" Research Data Center.

(2) Special research projects, including methodological research, should continue to receive project funding on a temporary basis.

(3) Users should pay some costs, but subsidies should be available for financially "weak" users, like PhD candidates. Better-off users (e.g., economic research institutes) should pay full costs, especially if they have the capacity to pass costs on to clients.

(4) It is suggested that new Research Data Centers are needed to cover additional subject areas (e.g., health, education, crime, migration).

(5) An additional Research Data Center may be desirable to hold data that covers more than one subject area. This Research Data Center could function as a “data trust,” archiving data for future scientific use.

2. The European Dimension *(Klaus Reeh)*

Background and current issues:

A great deal of social and economic data is now collected at the European level. The data are at least partly “harmonized” and thus cross-nationally comparable. Generally, however, data are restricted to those policy areas for which European institutions have substantial responsibility. Much less data is available in areas that are still mainly a national-level responsibility. Furthermore, the needs of policy-makers for whom the data are collected do not entirely square with the needs of scientists. For example, policy-makers want up-to-date information, whereas scientists are more interested in accuracy. Policy-makers are often satisfied with use of administrative and aggregate data and accept “output harmonization,” whereas scientists favor the collection of micro-level survey data and prefer “input harmonization,” that is, data collection instruments that are the same in each country.

Recommendations

(1) The German Data Forum (RatSWD) needs to recognize that the provision of high-quality data for science is a higher priority in Germany than at the European level. It is therefore recommended that the German Data Forum take the lead in pressing for improved European level data and statistics and working with Eurostat and sympathetic national statistical agencies.

(2) The German Data Forum could also take the lead in developing agreements among scientists about how best to compromise between their own needs and the differing needs of policy-makers for statistical data.

3. The Role of the German Research Foundation (DFG) *(Eckard Kämper, Manfred Nießen)*

Background and current issues

The future strategy of the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*) should be based on past achievements and lessons learned from those achievements. First, the focus of efforts should be mainly on generating valuable new data, not sharing existing data. Second, the leaders of projects whose primary purpose is to provide a collective good for the research community (e.g., large-scale surveys) should be required to build an effective infrastructure to assist the community of users. At present, compliance with requirements to deposit data in archives for use by other researchers is far below 100 percent. The reasons for non-compliance need to be investigated. They certainly include the considerable costs of compliance in both time and money, costs that active researchers are unwilling to bear.

The DFG has ample means to support its aims and is willing to play an active role under its elected leadership bodies.

Recommendation

It is up to the research community to adapt itself in cooperative ways to make effective use of available funding instruments. Cooperation is required (a) to identify research themes that merit priority, (b) to identify funding options to support these priorities, and (c) to help define a division of labor in research funding between different national (including ministries) and international funding bodies.

4. Providing Data on the European Level *(Peter Elias)*

Background and current issues

This advisory report reviews the potential demand for and provision of European data for social science research. The concept of data provision is defined broadly, covering the ease with which specific types of data can be found, interpreted, understood, and accessed by researchers. The advisory report first addresses the issue of why researchers need European (as opposed to national) data sources. This leads to a discussion of the potential demand for data at the European level. The main section focuses on the characteristics of data currently available or under development. The concluding section provides an assessment of the need for new and/or improved data infrastructures and suggests where efforts could be focused to meet such needs.

Recommendations

The major needs are:

(1) A new European Household Panel.

(2) Facilities to encourage analysis of birth and other age cohort studies.

(3) A European-based longitudinal study of business organizations.

(4) Improved access to micro datasets held by Eurostat. This should be feasible, ensuring appropriate data security via virtual remote access.

5. Infrastructure for High-Quality and Large-Scale Surveys. Cooperation between Academic Research and Private-Sector Agencies (*Peter Ph. Mohler, Bernhard von Rosenbladt*)

Background and current issues

Germany already has a fairly well established set of large-scale measurement instruments (LMIs) – mainly surveys – in the social sciences. The LMIs provide high-quality measurement of social and economic trends and should be viewed as a core element of the country's research infrastructure. The private sector is the main sector involved in designing surveys and collecting data, although the government and academic sectors are also important. The system works well at present but the degree of cooperation between the private sector and other sectors may not be adequate for the future.

Recommendation

Closer cooperation among government, the private sector, and academia would be beneficial for the development of LMIs. The private sector as a whole needs the assurance of a planned flow of work in order to undertake the large-scale investments in survey infrastructure that are required. The German Data Forum could take the lead in initiating closer cooperation and could look to the UK's Survey Resources Network as a useful example. Public-private partnerships are desirable for initiating, attracting funding, and continuing long-term survey-based projects. Such partnerships could promote methodological innovations, as well as collecting large datasets.

6. The Availability of Market Research Data and its Potential for Use in Empirical Social and Economic Research *(Erich Wiegand)*

Background and current issues

The potential of market research for secondary analysis lies mostly in the fields of consumer behavior, consumption patterns, and media usage. The German market research industry is huge – it has an annual turnover of more than two billion euros – and over 90 percent of its research is quantitative. However, samples are often highly specialized (rather than being representative and heterogeneous), telephone interviewing is the most common mode of data collection, and data documentation standards are not as high as academic social scientists would wish.

The chances of getting market research data released for secondary analysis would be improved if a win-win situation could be created by which, as a quid pro quo, the industry gained access to microdata from official statistical agencies. At present this is forbidden by law; individual data from official statistics are only available for scientific and not for commercial purposes.

Recommendation

It is recommended that the academic social and economic research community should consider supporting market research companies in their efforts to gain access to official statistics at the individual level. This would increase the readiness of companies and their clients to make data available for secondary analysis by social scientists. The appropriate body for the academic community (e.g., the German Data Forum) to negotiate with is the Working Group of German Market and Social Research Institutes (ADM, *Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute*).

Progress since 2001 and current state

1. The Recommendations of the 2001 KVI Report and Their Realizations up to 2008 *(Gabriele Rolf-Engel)*

Background and current issues

This advisory report describes the structure of the 2001 KVI report and lists the eight themes into which its 36 recommendations were categorized. It then reviews the extent to which each recommendation has or has not been implemented between 2001 and 2008. Each recommendation is assigned a green light (full implementation), a yellow light, or a red light. The advisory report makes no recommendations, but leaves it open to the German Data Forum to press for the implementation (or improved implementation) of 2001 recommendations that were either not implemented or only partly implemented. Key successes include:

(1) The foundation of four Research Data Centers, making confidentialized data files accessible for scientific purposes as well as CAMPUS-Files for teaching purposes.

(2) Establishment of two Data Service Centers.

(3) Access to business data via projects supported by the Federal Ministry of Education and Research (BMBF, *Bundeministerium für Bildung und Forschung*).

(4) Long-term funding for the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) and the National Educational Panel Study (NEPS).

2. Access to Microdata from Official Statistics *(Stefan Bender, Ralf Himmelreicher, Sylvia Zühlke, Markus Zwick)*

Background and current issues

A major outcome of the 2001 KVI report was the founding of four publicly funded Research Data Centers. These centers have greatly improved the access of researchers to official microdata. The centers have developed in constructive ways that were not entirely foreseen. Their services are widely used and many policy decisions are now planned and/or evaluated on the basis of data originating from them. Germany has gone from the bottom of the European league with regard to use of official statistical data for research purposes to the position of being an innovator. Innovations have been made, in particular, in providing access to data for teaching purposes via CAMPUS-Files, and in producing linked employer-employee datasets.

The Research Data Centers have developed strict criteria that provide for equal treatment of all data users, regardless of the subject/content of their research. Strict privacy and data protection conditions are in force. Researchers are required to spend substantial time working on the premises of a Research Data Center in order to learn about content and methodological issues relating to the data they are using. Access via controlled remote data sites may then be available.

Recommendations

(1) The four Research Data Centers should continue to increase their cooperation.

(2) One area of cooperation is the development of improved procedures for remote data access.

(3) Cooperative work is also underway to match survey data to administrative data.

It is noted that discussions are underway relating to the possible permanent establishment of the Research Data Centers of the Federal Statistical Office and the Statistical Offices of the German *Länder*.

3. Publicly Financed Survey Data: Access and Documentation *(Wolfgang Jagodzinski, Christof Wolf)*

Background and current issues

Four types of publicly financed surveys are considered: (1) academic surveys, (2) surveys using data from projects funded by the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*), (3) surveys utilizing data collected for research funded by the Federal State and the *Länder* (*Ressortforschung*), and (4) surveys employing data collected by national and international statistical agencies.

Recommendations

(1) Minimum standards of data accessibility should be required for all publicly funded scientific projects. All data should be stored in a digital repository provided by the social science infrastructure.

(2) A distinction should be drawn between two types of projects: those that are only of interest to a small group of researchers and those that are of wider interest. A mode of self-archiving (self-documentation) should be established for the former type, prior to data being lodged in a central depository. The latter type should be required to meet high professional standards of documentation and archiving. A pilot project should be established to define these standards.

(3) Access to data from the government research institutes (*Ressortforschung*) is at present quite limited and should be the standard expectation. Confidentiality requirements are often cited as the reason for restrictions but should rarely prevent access to an entire dataset.

(4) Access to data funded by national and international agencies is at present quite satisfactory, but it would be desirable for all documentation to reach the standard set by the European Social Survey (ESS).

4. Teaching and Statistical Training *(Ulrich Rendtel)*

Background and current issues

Well-educated researchers are needed for fruitful analysis of large social and economic datasets. Further, the creation of research data centers has generated increased demand for such analysts at the Diplom/Master's and PhD levels. But within the field of economics there is intense competition between sub-disciplines to attract students, and survey statistics has not fared well. The situation is better in sociology faculties.

Recommendations

(1) Some CAMPUS-Files (i.e., files freely available to teachers and students) are already available, including some from the Federal Statistical Office. However, the creation of new CAMPUS-Files, covering a wider range of subjects, is recommended as a way of attracting more high-quality students.

(2) It is recommended that new Master's programs be created in survey statistics, in part to compensate for the fact that, following the Bologna reforms, Bachelor-level students are not likely to have sufficient statistical training to undertake analysis of large datasets.

(3) Finally, it is recommended that it should become possible for students to receive academic credit for completing training courses in data analysis, which are currently offered by private sector data producers, the

Leibniz-Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*), and the Research Data Centers. This would be a sensible extension of the increasing collaboration between these organizations and universities.

5. e-Infrastructure for the Social Sciences *(Ekkehard Mochmann)*

Background and current issues

Social scientists have been slow to take advantage of collaborative research opportunities made possible by e-Science infrastructure. In principle, grids of fiber optic cable can link widely dispersed networks of researchers who can share data and undertake analyses using virtually unlimited computing capacity. For example, The EU research network Géant links 10,000 scientists at 300 sites in 50 countries and provides access to 80,000 CPU cores 24 hours a day.

The German Grid Initiative was launched in 2005, but so far social scientists have not contributed. Most social scientists appear to believe that Web 2.0 technology is adequate for their needs. A good example of technology at this level is the Council of European Social Science Data Archives (CESSDA) Portal, which holds important international datasets, including the European Social Survey (ESS) and the European Values Study. Documentation of studies is based on Data Documentation Initiative specifications (DDI), with Web software tools enabling users to browse and analyze data.

Recommendations

(1) The German social science community needs to decide whether it wants to take a concerted initiative to make use of data grid technology. If it does, then an institutional basis may be needed similar to the National Center for eSocial Science (NCeSS) in the UK. The Open Access Initiatives (e.g., the Berlin Declaration 2003) and the OECD (2004) declaration on open access to publicly financed data provide a basis for taking steps in this direction.

(2) If the social science community decides that it may wish to proceed, one way forward would be to set up a working group to make a needs assessment in relation to grid technology and to deal with a range of methodological, technical, and legal issues.

CURRENT STATE OF AND FUTURE DEMANDS IN DIFFERENT FIELDS

I. *(New) Data Types and their Potentials*

1. Macroeconomic Data *(Ulrich Heilemann)*

Background and current issues

No really lamentable or material gaps exist, and the cost of obtaining data is low. No major changes are likely before scheduled reforms to the National Accounts in 2011 and then 2014.

The provision of macro-data improved enormously 50 years ago when the National Accounts were introduced. In the last decade, we have seen huge improvements in research infrastructure for microeconomics, which perhaps has now "caught up" with macroeconomics. In many areas of social science, it is now no longer reasonable to regard data as a limiting factor.

2. Interdisciplinary Longitudinal Surveys *(Jürgen Schupp, Joachim Frick)*

Background and current issues

Household panel studies under academic direction are conducted in several countries. The Panel Study of Income Dynamics (PSID) in the US was the first to be launched and has been followed by similar major panel studies in Germany, Britain, and Australia. Experience shows that academic direction of these panel studies is beneficial for both the capacity to innovate and the capacity to extend their scope to include topics of interest to new disciplines. In practice, most of the current national panel studies were initially directed by economists and/or sociologists. But they now include questions and measures relating to psychological concepts, cognitive capabilities, and physical and mental health. They have also been extended to include age-specific modules of interest to developmental psychologists and biologists (e.g., mother and child and retirement modules). The German Socio-Economic Panel (SOEP, Sozio-oekonomisches Panel) will specifically add an innovation sample to try out new questionnaire modules and data collection methods, and will also conduct behavioral experiments with sample members as subjects. None of these changes were envisaged when the household panels started, but under academic direction innovation has been embraced.

The fact that data from these household panels are in continuously high demand from both the academic and policy communities is testimony to their capacity for innovation.

Recommendations

(1) It is recommended that increased attention be given in household panels to the earliest and last phases of human life – early childhood and late adulthood. Event-triggered modules (e.g., a module triggered by the birth of a child) should be designed to cover specific phases of the life course in more detail.

(2) Sample sizes need increasing to improve statistical power in analyzing data for population sub-sets (e.g., immigrants) and regions. It is recommended that sub-sample sizes of 500 per birth and age cohort should be considered an acceptable standard.

(3) It is recommended that national household panels be used as "reference" datasets for more specialized surveys. That is, they can effectively be used to provide comparisons (or baselines) for results from the more specialized studies.

3. Geodata *(Tobia Lakes)*

Background and current issues

In principle, all socio-economic data relate to a specific location in time and space. In practice, it is estimated that some geoinformation is provided for about 80 percent of all such data. The quantity, quality, and multidimensionality of geodata are improving rapidly in Germany and elsewhere, but are seriously under-exploited by social scientists. Large databases have been built up in both the public and private sectors. In Germany, the Federal Agency for Cartography and Geodesy (BKG, *Bundesamt für Kartographie und Geodäsie*) is a major source for users, but private-sector sources are also important. In both sectors it is common to charge fees for user access, especially when complex database searches are required, involving use of advanced mapping and spatio-temporal algorithms. However, some sources (e.g., GoogleEarth) provide free data and access to free software.

Recommendations

(1) There is a need for more cooperation between what, at present, are rather segregated public and private sector sources of geodata.

(2) The upcoming 2011 Census could be used as a focus for such cooperation, partly by efforts to link Census data to other sources.

(3) The German Data Forum could assist the process for social scientists whose work could benefit greatly from more sophisticated use of geoinformation.

(4) The German Data Forum could also facilitate international cooperation in the use of geodata.

4. Regional Data *(Gerd Grözinger, Wenzel Matiaske)*

Background and current issues

Space/location is an increasingly important dimension of social science analysis. It is clear that intra-national (or regional) comparisons can prove just as valuable as the more fashionable international comparisons. A great deal of high-quality regional data is available, provided by official and semi-official statistical agencies and generally in the form of user-friendly DVDs. Academic researchers and, especially, commercial firms also collect spatial data, often at a very detailed local level. Regional analysis has also been facilitated by methodological advances; in particular, the development of multilevel statistical analysis.

Recommendations

(1) It would be valuable for researchers if some existing datasets that are not yet available for spatial analysis could be released (e.g., the PISA E dataset).

(2) On many topics (e.g., criminal behavior) insufficient data are available at the regional and local level.

(3) An agreed classification of localities should be used in research. The European Nomenclature of Territorial Units for Statistics (NUTS, *Nomenclature des unites territoriales statistiques*) classification is the clear choice.

5. Genetically Sensitive Sample Designs *(Frank M. Spinath)*

Background and current issues

Many social and economic outcomes, including earnings, life satisfaction, and physical and mental health, result from the interplay of genetic and environmental factors. There is an opportunity to modify existing household panel studies in order to allow for state-of-the-art multi-group analyses of genetic and environmental effects. The panels already include respondents who are genetically related to each other in a wide variety of ways. What is needed is an additional over-sampling of twins. In Germany there is no central twin register, but previous studies have nevertheless had considerable success in recruiting twins. There is, however, usually some bias towards oversampling women and monozygotic twins.

Recommendation

It is recommended that twin cohorts be added to and integrated into panel studies, including the German Socio-Economic Panel (SOEP, *Socio-oekonomisches Panel*).

6. Biological Variables in Social Surveys *(Rainer Schnell)*

Background and current issues

Social scientists have almost completely ignored biological variables in their research. Yet it is clear that these variables are important in accounting for many aspects of social and economic behavior. There are, in fact, many biological measures that can be taken by medically untrained observers (e.g., survey interviewers) in standard surveys. These include body-mass index, grip strength, and simple pulmonary function tests. However, as a matter of law, blood samples can only be taken by a medical doctor in Germany. Small sized sensors and “intelligent clothing” may become increasingly important for use in surveys. Generally, respondents react favorably to the use of new instruments, but cooperation may later decline as the techniques become more common.

Recommendations

(1) Biological variables (biomarkers) should be collected in a wide range of surveys. With this in mind, biologists and behavioral scientists, for example, should become members on the advisory board of the Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*). Graduate programs in the social sciences should alert students

to the value of biomarkers, and students should be trained in appropriate methods of analysis.

(2) Research is needed on the willingness of respondents to cooperate in providing biological measures and on possible biases resulting from differential cooperation. Ideally, controlled experiments should be conducted.

(3) Funding opportunities for cross-disciplinary work are limited in Germany. An interdisciplinary special research program of the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*) would be a valuable first step. An EU project may also be a promising way forward.

7. Administrative Transaction Data *(Julia Lane)*

Background and current issues

This advisory report describes the potential for the social sciences of data from a wide range of sources, including Internet clickstreams (e.g., in the use of social networking sites), e-mails, cell phones, GPS systems and radio frequency identification devices, credit card purchases, telephone calls, retail store scanning records, health records including biomarkers, and employment records.

In sheer quantity, administrative data dwarf all other datasets, but at present social scientists make little use of them. Note, however, that the UK's Economic and Social Research Council (ESRC) has set up an Administrative Data Liaison Service to link academics to producers of administrative data, and the US National Science Foundation (NSF) has recently awarded large grants for the study of social and information networking. Peter Elias, on behalf of several international agencies, is working to establish the International Data Forum.

Recommendations

(1) Invest in new methods of data collection to harvest administrative data. It is necessary to solve issues of privacy and confidentiality, but considerable progress has already been made in this regard. Funding agencies are at present keen to fund such efforts. This opportunity should be taken advantage of.

(2) Devise new ways of analyzing transaction data. The data are often characterized by a high noise-to-signal ratio and by non-linearity. Standard tables and regression analyses tend to be of limited value. Visual representations are often preferable. Social scientists could learn much about such techniques from computer and behavioral scientists.

(3) The study of administrative data should be conducted by “virtual communities” of scientists using an open and transparent infrastructure for data sharing and dissemination. In this respect, social science communities should become more like “hard science” communities.

(4) Improved methods of communicating administrative data to policymakers and broader communities need to be devised.

8. Transaction Data: Commercial Transaction Surveys and Test Market Data *(Bernhard Engel)*

Background and current issues

Commercial transaction surveys and test market data are virtually unused by the scientific community. Yet they are important sources for understanding consumer behavior. Their advantage is that they provide “hard” data on sales and marketing, not just “soft” data on consumer perceptions. There are three main problems facing scientists who may wish to use the data. First, the commercial owners need to give permission. Second, the data would need to be made anonymous with respect to both individuals and products, without losing information vital to research. Third, the quality of the data would need to be checked to determine their value for scientific research.

Recommendations

(1) In cooperation with the official statistical community and market researchers, the German Data Forum could facilitate *scientific use of commercial transaction and test market data* by initiating a project to investigate issues of data “anonymization,” with respect to both the identity of products and the identity of consumers.

(2) The German Data Forum could also take the lead in proposing standards of data quality.

9. Time Use and Time Budgets *(Joachim Merz)*

Background and current issues

Time use studies are uniquely valuable for studying, *inter alia*, the division of labor within households, household production, and leisure activities. The Federal Statistical Office conducts a time use survey approximately every ten years, and this is now integrated with the EU's Harmonized European Time Use Surveys (HETUS).

Recommendations

(1) The next German national time use survey, due in 2011-12, should be conducted by the Federal Statistical Office and needs secure funding. It should again be embedded in the EU's HETUS. It is recommended that the German Data Forum support this view.

(2) Supplementary questions about infrequent activities should be appended to the main diary collection instrument.

(3) Mobile devices should be used to provide additional experience sampling data. This requires a pilot study before incorporation in the main survey.

(4) Expenditure data and subjective satisfaction data should be collected alongside time use data.

(5) A new longitudinal study on time use is recommended to answer questions about changes in individual time use profiles in response to major life events and changing environmental conditions.

II. *Methods*

1. Survey Methodology: International Developments *(Frauke Kreuter)*

Background and current issues

Survey methodology has been heavily influenced by two factors in recent years: falling response rates and technological advances in data collection. Falling response rates have led researchers to emphasize that these rates were never a valid guide to *response bias*. Two alternative measures of response bias are now receiving more attention: single indicators for an entire survey (e.g., the variance of non-response weights) and item-specific estimates (e.g., comparisons between survey results for a particular variable and interviewer observations or administrative data).

In order to counteract falling response rates, efforts are being made to reduce *response burden.* One approach is multiple matrix sampling, which involves putting different sub-sets of questions to sub-sets of respondents drawn from an initial sample.

Changing technology has allowed development of mixed-mode surveys (e.g., different sub-sets of respondents record their data using CAPI, or CATI, or by mail). Substantial research efforts are now being directed at methods of estimating response bias for each mode and, overall, for multi-mode surveys.

Recommendations

(1) Many recent developments, including deployment of mixed-mode surveys and requirements for interviewer observations, place increased demands on survey interviewers. This suggests a need for further research on interviewer performance and its consequences for response bias.

(2) German researchers are particularly well placed to investigate response bias in mixed-mode surveys, due to major efforts already undertaken in data linkage (e.g., between surveys and official sources).

(3) Randomized experiments that test alternative survey modes and methods could usefully be conducted within survey organizations that are already responsible for carrying out many surveys at the same time.

(4) All of the above recommendations require increased cooperation between researchers and survey organizations.

2. Improving Research Governance through Use of the Total Survey Error Framework *(Marek Fuchs)*

Background and current issues

The infrastructure for survey research has greatly improved in the last 20 years. It would now be valuable to conceive of survey methodology as a framework or "science for conducting and evaluating surveys." The specific framework proposed here revolves around the concept and measurement of *total survey error*. Total survey error includes both sampling or "representation" error and also measurement/response error (Groves et al, 2004). In practice, it is usually too expensive to calculate mean square errors for particular sample estimates because multiple repetitions of one's survey design are required. However, researchers can benefit greatly from using the total survey error framework because it alerts them to all possible components of error and serves as a guide in designing cost-effective surveys.

Recommendations

(1) It is recommended that the research community adopt the total survey error framework as a guide to survey design and evaluation.

(2) Survey methodology should be regarded as a valuable “cross-disciplinary discipline.” The emergence of Master’s degrees in this field should be encouraged.

3. Metadata *(Arofan Gregory, Pascal Heus, Jostein Ryssevik)*

Background and current issues

In the social sciences, *metadata* can be defined as available documentation of primary datasets. Historically, researchers paid little attention to the quality of metadata surrounding their work; archiving was left to archivists. This mindset is changing. There have been rapid advances in the development and actual implementation of high-quality metadata standards; standards which apply to datasets throughout their life cycle from initial collection through to secondary use, perhaps in conjunction with quite different datasets. The German Research Data Centers, which were set up following the 2001 KVI report, together with the Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*), have taken a leading role in these developments.

This advisory report describes two sets of metadata standards in some detail: the Data Documentation Initiative (DDI) and the Statistical Data and Metadata Exchange (SDMX) Standard. These are seen as central to a potential global metadata management framework for social science data and official statistics.

Recommendations

(1) It is suggested that the German Data Forum endorse the importance of high-quality data documentation and the implementation of the metadata quality standards described above.

(2) Adoption of these standards requires the establishment of an industry standard, Web-service-oriented, and registry-based IT infrastructure. This infrastructure can then facilitate the management, exchange, harmonization, and re-use of data and metadata.

4. **Paradata** *(Frauke Kreuter, Carolina Casas-Cordero)*

Background and current issues

The use of computers in survey data collection generates a great deal of "paradata," a term coined by Mick Couper (1998). Paradata are data "surrounding" a survey and consist mostly of records of efforts to contact respondents, together with interviewer observations. Audio recordings made in the course of computer-assisted data collection also constitute paradata. The data are valuable for analyzing problems of survey non-response and for assessing the pros and cons of different data collection modes. Paradata can be used to achieve "continuous quality improvement" in survey research. In this context, the European Statistical System has developed a handbook on improving data quality through effective use of paradata.

In Germany, data collection agencies generate and disseminate fewer paradata than in some other Western countries.

Recommendations

(1) It is desirable that the research community as a whole demand high-quality paradata. This would encourage data collection agencies to make the necessary investments.

(2) Experimental survey designs – for example, to assess alternative data collection modes, or alternative respondent contact procedures – particularly benefit from the collection of paradata.

(3) Panel surveys provide special opportunities for the collection of valuable paradata because the same respondents are repeatedly interviewed under (potentially) varying conditions.

5. **Record Linkage from a Technical Point of View** *(Rainer Schnell)*

Background and current issues

Record linkage involves linking the same objects (e.g., survey respondents) in two or more databases using a set of common identifiers. These identifiers may include unique individual ID numbers, but other unique characteristics or combinations of characteristics may be used as well.

This advisory report reviews problems in record linkage and comments on available software.

Recommendations

(1) Research needs to be undertaken on the practical performance of record linkage programs and algorithms. Large real social science datasets – not simulated data – need to be used for this purpose.

(2) A European research program is needed on pre-processing keys for privacy-preserving record linkage.

(3) A National Record Linkage Center is needed in Germany. At present, different disciplines duplicate efforts in this area.

6. Experiments, Surveys, and the Use of Representative Samples as Reference Data *(Thomas Siedler, Bettina Sonnenberg)*

Background and current issues

In the last two decades social scientists have made increasing use of laboratory experiments to research social preferences and behavioral outcomes. A problem with most experiments is that the subjects are students and self-selected (volunteers). There is some evidence that this biases results, and that students who self-select into experiments are not even representative of the student body from which they are drawn. It is therefore valuable to compare the results of experiments with results of representative sample surveys that have investigated the same topic. Ideally, a sub-set of survey respondents should be found to take part in laboratory experiments. At the bare minimum, using the survey data as *reference data* allows the experimenter to estimate biases in his/her results. At best, the comparison may help to validate both sets of results. In this regard, the advisory report cites research on risk attitudes in which data from the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) and experimental results were compared. It transpired that survey respondents who rated high on willingness to take risks then actually took high-risk decisions in an experimental setting.

Recommendation

It is recommended that surveys be used as *reference data* for social and economic experiments.

7. Experimental Economics *(Simon Gächter)*

Background and current issues

Experimental economics should be regarded as a *method* applicable within many sub-fields economics, rather than as a sub-field itself. The aim is to use controlled laboratory-type conditions to answer *if-then* questions about the choices that economic agents face. Hypotheses can be more rigorously tested in the lab than by using observational data, but issues then arise about the generalizability of results to the "real world." One such issue arises because most studies use students as their laboratory subjects, with the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) and the British Household Panel Study (BHPS) being exceptions. This special form of sample selectivity/bias needs addressing.

Recommendations

(1) It would be valuable if experimental economists could develop an agreed questionnaire for administration to all subjects, which would document their specific socio-economic characteristics. Sample selectivity could then be assessed and taken into consideration in assessing empirical results.

(2) For recommendation (1) to be beneficial, it would be necessary to set up data depositories for experimental economics. At present, there appear to be only two journals that require authors to make data available, and there is only one major depository (in the US) at which data is available to other researchers. It would be desirable to set up depositories in Germany and elsewhere in Europe. The depositories would then be the right place to lodge the results of questionnaires completed by experimental subjects (Recommendation 1).

(3) The German Data Forum might wish to advocate these developments and facilitate their implementation. However, the difficulties are considerable. Many researchers feel they have strong property rights over their data. Further, the task of making subject samples and data more comparable across studies would be time-consuming for researchers.

8. Experience Sampling *(Michaela Riediger)*

Background and current issues

Experience sampling refers to the repeated capturing of experiences – such as feelings, thoughts, behaviors, and events – at or close to the moment at which they actually occur in an individual's life and in his or her normal/natural environment. Data are typically recorded by the subject in response to a cue (e.g., a beeper going off) using a hand-held computer. Compared with standard survey reports, which are based on recall, data produced by experience sampling have a high level of validity and are particularly valuable for assessing within-person changes across time. On the other hand, experience sampling studies are resource-intensive and place a heavy burden on subjects, who usually have to be paid. This means that small samples are the norm, with sample attrition still a problem. Further, subjects' responses may be affected by participation in a study (reactivity effects).

Experience sampling is going through a boom period, but few datasets are available for secondary analysis. Most studies are small-scale, conducted by psychologists. The use of experience sampling in large household panels is in its infancy. However, the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) has successfully piloted an experience sampling procedure with a small but representative sample.

Recommendations

(1) Experience sampling should be used in large-scale surveys as part of a multi-method approach. Similarly, it can be used to conduct "studies within a study."

(2) Mobile technology should be used to reduce the burden on respondents.

(3) Careful sample selection criteria should be used to minimize self-selection and other forms of sample bias. Control group designs are needed to assess reactivity effects.

(4) It should be a requirement of funded research that datasets be deposited for secondary use.

(5) Experience sampling could be included as a research topic in the Priority Program on Survey Methodology of the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*).

9. Virtual Worlds as Petri Dishes for the Social and Behavioral Sciences *(Edward Castronova, Matthew Falk)*

Background and current issues

Social scientists need to develop theories and test hypotheses at the macro-societal level. Computer generated *virtual worlds*, with humans represented as avatars, should have priority as a tool for generating and testing these theories. Virtual worlds have many advantages. They can be used to conduct realistic controlled experiments, varying one or more parameters as the experimenter sees fit. They can be constructed to have a good fit with empirical reality and they allow for the interaction of numerous (even millions of) players. They can be run for long periods, like panel studies. Generally, they pose no serious risks to players, avoiding the ethical issues that limit many other types of studies. They are, however, initially expensive and time-consuming to set up, although, like the virtual worlds run by the gaming industry, they may eventually be self-funding.

Recommendations

(1) Virtual worlds should be recognized as a research tool for future research at the macro-societal level.

(2) Initial research funding is needed.

(3) Virtual worlds have good prospects of becoming self-funding or profitable by means of charging users both initial and ongoing fees, as happens with Internet worlds marketed by the gaming industry.

10. Qualitative Interviewing of Respondents in Large Representative Surveys *(Olaf Groh-Samberg, Ingrid Tucci)*

Background and current issues

Large representative surveys are using mixed methods to an ever-increasing degree. For example, biomarkers, register data, and experiments provide different types of evidence linked with survey data. However, the practice of conducting qualitative interviews with sub-sets of respondents from large scale surveys, including longitudinal surveys, is still quite rare. The key advantage of this approach, in contrast to many reported case studies, is that the researchers know precisely what they have "cases of."

Qualitative methods have proven just as valuable as quantitative methods in providing insights into social reality that reflect the multidimensionality of individual life courses and lived realities. Furthermore, in-depth interviews can provide an improved understanding of individual decision-making processes and behavior resulting from more or less unconscious strategies. They also provide insights into decisive turning points in people's lives. Finally, use of multiple methods to investigate the same issues enables researchers to "triangulate" their results and so assess their validity.

Recommendation

It is recommended that consideration be given to conducting qualitative interviews with purposively selected respondents from large-scale surveys, including longitudinal surveys.

III. Data Protection and Research Ethics

1. Data Protection and Statistics – A Dynamic and Tension-Filled Relationship *(Peter Schaar)*

Background and current issues

A balance has to be struck between the requirements of individual privacy and the research needs of the scientific community. Despite the development of ingenious methods of protecting privacy, including use of aliases, it is clear that recent decisions of the Federal Constitutional Court, interpreting the Basic Law, are likely to make it more difficult for researchers to collect comprehensive datasets, retain them, and make them available to others. The Court takes the view that individuals have a right to privacy, which can only be abrogated by informed consent for specific purposes. Further, the Court holds that informed consent given for one study does not allow datasets to be combined and regularly updated. "Profiling" of individuals via combining datasets is also clearly illegal. Posting data on the Internet runs such serious privacy risks that it can only be allowed if absolute anonymity is guaranteed.

Recommendation

The Research Data Centers of the Federal Statistical Office and other public agencies provide Scientific Use Files for research, teaching, and other specific uses. The files are created in such a way as to ensure virtual or full anonymity of subjects. This is one way forward.

2. Record Linkage from the Perspective of Data Protection *(Rainer Metschke)*

Background and current issues

A cherished dream of social science researchers is to be able to link diverse survey datasets relating to the same individuals or households. The realization of this dream is beset with many obstacles, not least constitutional and legal provisions relating to data protection and privacy. This advisory report discusses current and potential future methods of making data available to the research community within the law. These methods include pseudonymization of respondents and data encoding, as well as the related use of "data stewards" (see below).

Recommendations

(1) Researchers and official statistics need to determine which datasets it is appropriate to link for research purposes, and then list the legal, technical, and methodological problems likely to be encountered.

(2) One method that is used for linking datasets, while still protecting privacy, is use of a legal entity known as a data steward (*Treuhänder*). The precise legal status of data stewards needs to be clarified.

(3) Additionally, the appropriate relationship between data stewards and Research Data Centers needs to be defined.

(4) Recommendations for the modernization of the law relating to statistics and data integration should eventually be drawn up. It would be sensible to conduct a pilot study first.

3. New Methods in Human Subjects Research: Do We Need a New Ethics? *(Karsten Weber)*

Background and current issues

New methods of data collection in the social sciences, including online research (e.g., data mining of websites) and use of biomarkers, pose *ethical issues* related to autonomy, beneficence, justice, privacy, and informed consent. These general ethical principles need to be more tightly defined or redefined by ethics committees dealing with modern data collection methods. For example, the principle of beneficence – promoting the good of others – appears to have few clear applications in social science research and needs to be redefined as preventing harm. Special consideration needs to be given to the development of ethical principles and practices relating to research on children and other vulnerable groups.

Recommendations

(1) Ethics committees should be established to approve and monitor human subjects research.

(2) The committees should employ risk assessment procedures to assess risks to research subjects.

(3) Particular attention should be paid to risk assessment and the application of ethical principles to research on children and other vulnerable groups.

4. Does Germany Need a (New) Research Ethics for the Social Sciences? *(Claudia Oellers, Eva Wegner)*

Background and current issues

In practical terms, Germany does not have a detailed set of ethical requirements that protects research subjects and is designed for the social sciences. However, all researchers have to abide by the requirements of the Federal Data Protection Act, and professional bodies in sociology and psychology have issued ethics guidelines affecting behavior towards peers rather than towards research subjects.

A review of ethics procedures in Britain and the US was undertaken to see if they offered useful examples for Germany. British procedures appear worth consideration; US procedures are perhaps too heavily geared towards the natural sciences.

In Britain, beginning in 2006, the Economic and Social Research Council (ESRC), which is the main funding body, forced universities seeking funding to set up ethics committees required to implement six key principles, four of which protect human subjects. Subjects have to be fully informed about the purposes and use of the research and have a right to remain anonymous, the data provided must remain confidential, participation must be voluntary, and the research must avoid harm to subjects. In practice, most universities have ethics committees in place at a faculty and/or departmental level and not just at the broader university-wide level.

Recommendation

Germany should consider the introduction of ethics principles and procedures similar to those in Britain.

IV. Fields

1. Migration and Demography

1.1 Migration and Globalization *(Martin Kahanec, Klaus F. Zimmermann)*

Background and current issues

Existing international migration datasets do not effectively capture the complexity of migration trajectories. Little is known about the prior experiences of immigrants in their home countries, about migrants who make more than one move, about the moves of additional family members, or about out-migration. The experiences of skilled migrants – the migrants that host countries are most keen to attract – are especially poorly documented because many make multiple (including circular) moves.

Lack of quality data about immigrants reduces the effectiveness of public policy, especially in education and job training.

International organizations, including the EU and the World Bank, have begun to make some datasets available to researchers, as have non-governmental organizations, for example, the Institute for the Study of Labor (IZA, *Forschungsinstitut zur Zukunft der Arbeit*). These organizations use advanced data management technologies to store data and provide access to users, but this does not remedy underlying data deficits.

Recommendations

(1) International coordination of data collection methods and standardization of immigrant identifiers.

(2) Guidelines for collecting adequate information about immigrants, including retrospective data on experiences in their home countries.

(3) Longitudinal data collection.

(4) Boosting immigrant samples in large social surveys.

(5) Appropriate anonymity standards relating to immigrant respondents.

(6) Data Service Centers using modern technologies to facilitate user access.

(7) Making arrangements for future data access a priority in planning data collection.

1.2 Migration and Statistics *(Sonja Haug)*

Background and current issues

Empirical research on migration is faced with problems relating to the fact that (1) most surveys under-sample some migrant groups, and (2) different sets of official statistics contain differing estimates of migrant numbers. However, several significant improvements have been made in migration statistics or are currently projected. The concept of “migration background,” replacing the concept of foreign-born, has been adopted in the German Microcensus and is recommended for the main 2011 Census. If a projected central population register is set up, future migration researchers will have an ideal sampling frame from which to draw adequate and/or special migrant samples. At present, the most accessible large dataset is the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*), which does over-sample migrants.

Recommendations

(1) Improvements are needed in sampling methods applied to migrant populations, especially small groups.

(2) New longitudinal studies focused on migrants should be a priority in Germany and internationally.

1.3 Internal Migration *(Andreas Farwick)*

Background and current issues

Research on internal (within-country) migration covers a wide range of issues relating to the reasons, distance, and direction of moves, as well as processes of decision-making. Both official aggregated data and cross-sectional data are useful for descriptive purposes but have limited value for explaining why households change residences. This advisory report describes longitudinal datasets that are valuable for understanding causal relations, but also notes their limitations. Retrospective longitudinal studies have the advantage of providing long histories of recalled events, including migration events. Their limitation is that they do not provide valid data on reasons for and attitudes to changes of residence. Prospective longitudinal studies are generally preferable in this respect, but could be improved by providing standardized data on aspects of migration (see below).

Recommendations

(1) Longitudinal (and other) studies should collect standardized data on place of residence and changes of residence at the smallest available spatial level, using the Nomenclature of Territorial Units for Statistics (NUTS, *Nomenclature des unites territoriales statistiques*) coding scheme.

(2) Data should be coded according to standard typologies of the characteristics of places/locations of residence, changes of residence, reasons for moving, intentions to move in future, the dwelling itself, the neighborhood, and commuting.

1.4 Fertility and Mortality Data for Germany *(Michaela Kreyenfeld, Rembrandt Scholz)*

Background and current issues

The data infrastructure for research on fertility and mortality in Germany has improved in recent years. In particular, several large datasets have been made available through Research Data Centers. Fertility data, in particular, have been improved through the Microcensus, which now collects information about the total number of children born to each woman during her life. There are still some "weak spots." Accurate counts and information about the exact composition of reconstituted families are lacking. Also, it is known that immigrants are healthier than average, but their mortality risks are inadequately understood.

Recommendation

Collecting information in the Microcensus via a household relationship matrix would considerably improve the quality of data on households/ families and should, perhaps, be considered. Clearly, however, adoption of a matrix approach would represent a major change in the design of the Census instrument.

2. Measuring Competencies

2.1 Measuring Cognitive Ability *(Roland H. Grabner, Elsbeth Stern)*

Background and current issues

Many survey researchers want, in principle, to include cognitive tests in questionnaires. One quite common motive is to obtain unbiased estimates of the effects of socio-economic variables on some outcome (e.g., wages). The practical problem is that most cognitive tests that are regarded as valid by psychologists take too long to administer to be included in socio-economic surveys. This advisory report reviews a promising new type of cognitive test – the WM or working memory test. WM tests measure a person's ability to store and process information in working memory. There is considerable evidence showing that this ability is highly related to domain-specific abilities and to general cognitive abilities (the *g* factor). The advisory report reviews several promising WM tests; the shorter ones would take five to ten minutes to administer in a computer-assisted survey setting.

Recommendation

Working memory tests require further development and testing. There are several promising candidate tests, but few evaluations with large and diverse samples have been undertaken.

2.2 Measuring Cognitive Competencies *(Ulrich Trautwein)*

Background and current issues

In order to make well informed decisions in the educational arena, politicians and other decision-makers need high-quality data on the development of student competencies. This advisory report argues that there is often no substitute for well constructed standardized tests, and that it is important to measure a range of competencies and not just rely on measures of general cognitive ability.

Recommendations

(1) Policy-makers need better quality longitudinal data about the development of student competencies to inform their decisions.

(2) Data from multiple sources (e.g., school achievement studies and data from national agencies) need to be linked.

2.3 Measuring Vocational Competencies *(Martin Baethge, Lena Arends)*

Background and current issues

The EU has responded to the internationalization of labor markets in part by seeking to improve the quality and transparency of Vocational and Educational Training (VET). A European Qualifications Framework has been proposed, together with a European Credit System for VET. A logical extension of these policy initiatives is development of agreed measures of vocational competencies. It is clear that current international measures of adult literacy, numeracy, etc. are too broad to be termed measures of "vocational competencies." Beyond that, there is no consensus even about what types of measures are required. One school of thought favors measurement of *internal conditions* (dispositions and skill sets), which are taken to indicate capacity to perform vocational tasks. A second school of thought favors measurement of *external performance* of specific vocational tasks. The authors favor the first approach, which views individuals as carriers of skills that could be adapted to a variety of vocational tasks and may form a basis for lifelong learning.

At present, EU Member States are attempting to achieve convergence on these issues through the Copenhagen process. A Feasibility Study is currently underway, with participation by experts from interested countries. This study will provide a clear picture of national VET programs that might be included in international comparisons, but there is no immediate prospect of agreement on measurement issues.

2.4 Measuring Social Competencies *(Ingrid Schoon)*

Background and current issues

There are differences in the way social competencies are conceptualized and measured in psychology, education, sociology, and economics. In general, social competency requires adapting individual characteristics to social demands and specific situations. Limited data are available on the development of social competencies during individual lifetimes, or about their possible biological basis. Several archives hold national and/or international datasets that include some data on social competencies. These include the UK's Economic and Social Data Service (ESDS), the Council of European Social Science Data Archives (CESSDA), and the Inter-University Consortium for Political and Social Research (ICPSR).

Recommendations

(1) Existing data on social competencies need to be cataloged and documented in a consistent way in order to promote secondary analysis.

(2) Longitudinal data are needed to assess the acquisition of social competencies and their expression in specific contexts. Longitudinal data are also required to understand intergenerational transmission of competencies.

2.5 Subjective Indicators *(Beatrice Rammstedt)*

Background and current issues

Subjective indicators – the best-known relating to life satisfaction – are widely used in survey research and have been shown to be associated with a large array of social and economic outcomes. The psychometric properties of subjective indicators have not been adequately investigated. The main difficulty is that the validity of self-reports (e.g., reports of satisfaction, or worries, or trust in others) is hard to assess, even by peer and/or expert reports.

Recommendations

(1) Most large surveys, like the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) and the European Values Study (EVS), use single indicators. The reliability and validity of single indicators need to be assessed, and if they prove deficient, short multi-item measures should be developed and assessed.

(2) The cognitive processes used by respondents in making their subjective judgments require investigation.

3. Education and Research

3.1 Education Across the Life Course *(Hans-Peter Blossfeld)*

Background and current issues

There is enormous demand in Germany for high-quality longitudinal data on education through the life course and on returns to education. Until the National Educational Panel Study (NEPS) was set up in 2008, there was no long-term German panel study providing nationwide data on educational experiences, competences, and outcomes. Previous studies dealt primarily either with particular transitions (e.g., from secondary school to university) or were focused on particular areas of the country.

In planning the NEPS, it was considered that birth cohort studies take too long to bear fruit; it takes nearly 20 years before the first “subjects” enter the labor force. So, following the lead of the National Center for Education Statistics in the US, NEPS will have a “multi-cohort sequence design.” This involves collecting data on “subjects” during key transitions: kindergarten to elementary school, elementary school to secondary school, and so on. At each transition, decisions are made about participation in different educational institutions and processes, and this participation leads to development or non-development of various competences. Varying outcomes and returns to education are recorded. Particular attention will be given in NEPS to immigrant educational experiences and outcomes.

NEPS will give high priority to preparation of a Scientific Use File for researchers and will offer training courses on how to make effective use of the data.

3.2 Preschool Education *(C. Katharina Spieß)*

Background and current issues

There is widespread international recognition of the importance of preschool education as a key determinant of later educational outcomes. However, there is a dearth of datasets in Germany that enable researchers to assess linkages between preschool educational experience and later outcomes. Existing datasets focus mainly on preschool attendance and are particularly deficient for children below the age of three and from migrant families. Two recent developments – the “Educational Processes, Competence Development, and Selection Decisions in Pre- and Primary School Age” (BiKS, *Bildungsprozesse, Kompetenzentwicklungen und Selektionsentscheidungen im Vor-*

und Grundschulalter) study at the University of Bamberg and the National Educational Panel Study (NEPS) – will improve matters somewhat, but the former study stops at the second grade of elementary school and the latter only includes children from age four onwards.

Recommendations

(1) Improved data are needed to measure the *quality* of preschool education, including the education of children under three.

(2) These data should be linked to cost data, so that cost-benefit studies can be undertaken.

(3) It is important to have adequate sub-sample sizes for disadvantaged children and children from migrant families.

(4) It may be beneficial to improve the preschool data infrastructure jointly with research infrastructure on families as well as on abilities and competencies in these other areas. Possibilities of data linkage need to be investigated.

3.3 Data in the Domain of Secondary School Education *(Petra Stanat, Hans Döbert)*

Background and current issues

Compared to most Western countries, Germany knows little about its school system. Data are lacking on how student competencies develop over time and on the factors which affect development. Official school statistics are at an aggregate level only. The Microcensus is valuable for some purposes but provides only cross-sectional data and has no information on preschool attendance or learning outcomes. The German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) measures some aspects of cognitive abilities but not subject-specific competencies.

Partly as a result of the poor performance of German students in standardized international tests, there has been increased interest in measuring learning outcomes. Major recent innovations are the founding of the Institute for Educational Progress (IQB, *Institut zur Qualitätsentwicklung im Bildungswesen*) and the National Educational Panel Study (NEPS). The IQB will administer competency tests to representative samples of students in the 16 *Länder*. The NEPS is a multi-cohort study, starting in 2009, that will cover eight key educational and career transitions. Data from the IQB and the Panel Study will be available for secondary analysis through research data centers.

Recommendations

(1) School statistics should be reported at the individual-level in all *Länder*.

(2) To allow for longitudinal analysis, school statistics should ideally include unchanging student identifiers. The legal and practical feasibility of using such identifiers needs to be assessed.

3.4 Knowing More about Vocational Training *(Steffen Hillmert)*

Background and current issues

Vocational training is a key aspect of the lifelong learning required in modern economies.

To understanding the costs and benefits of this training, it is essential to have *longitudinal data*, which can capture multiple periods of training undertaken by the same individual. However, at present, longitudinal evidence is limited. The German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) and the German Life History Study (GLHS) are valuable sources, but evidence is still needed from administrative sources. The data records generated within the "dual system" of vocational training are quite comprehensive, but do not allow individuals to be traced from one period of training to another. In other sectors of vocational training, even more serious data deficiencies exist.

Recommendation

Each individual should have a common ID number within the vocational education system so that his/her education and training career can be traced over a lifetime.

3.5 Higher Education *(Andrä Wolter)*

Background and current issues

In the last five years, there has been a major increase in research on higher education in Germany. This has been partly due to the boom in education research generally, and partly due to the Bologna Reforms, which have led to increased demand for internationally comparable data.

Data come from two sources: official statistics and surveys. In principle, all data are available for secondary analysis, although some practical problems arise (see below).

Recommendations

(1) Access for secondary analysis of education surveys could be made more convenient by setting up a Research Data Center associated with the Higher Education Information System (HIS, *Hochschul-Informations-System*).

(2) Some specific sets of questions need to be integrated into all education studies. These include questions about migration status, learning competencies, and evidence of lifelong learning.

(3) Panel studies are a particularly important deficit, although this will be partly remedied by the establishment of the National Educational Panel Study (NEPS).

3.6 Adult Education and Lifelong Learning *(Corinna Kleinert, Britta Matthes)*

Background and current issues

Adult education and lifelong learning are regarded as increasingly important due to the emergence of a "knowledge society" and the increased economic competition resulting from globalization. Germany has many different sources of cross-sectional data on adult education, plus several longitudinal studies, including the new National Educational Panel Study (NEPS), the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*), and the German Life History Study (GHLS). A major problem is that the sources provide contradictory evidence; for example, the Microcensus reported that only 13 percent were involved in adult education in 2003, while the Adult Education Survey (BWS, *Berichtssystem Weiterbildung*) reported 41 percent. Such large divergences highlight the need to develop standardized questions.

Recommendations

(1) It is *not* recommended that new sources of data be provided.

(2) The main requirement is to develop standardized questions that capture all aspects of lifelong learning: formal learning, on-the-job learning, informal learning, and development of measured competences. Also, the *household context,* in which decisions about continued learning are made, needs to be recorded.

3.7 Research, Science, Development *(Stefan Hornbostel)*

Background and current issues

Research institutions are under increasing pressure to measure and even predict ("foresight studies") their research performance. They need to do so in order to avoid being disadvantaged in national and international competition for funding.

Outcome measures are generally preferred (e.g., citations in high-impact journals). However, input measures, including attraction of third-party funding, are also often used. Germany's federal research report reflects the demand for evidence to assess research performance, but data are often delayed and are not appropriate for outcomes analysis. By contrast, the German Council of Science and Humanities (WR, *Wissenschaftsrat*) provides up-to-date and transparent ratings that are available for scientific use. Internationally, Google Scholar and other open access repositories are increasingly valuable.

Germany may be falling behind in its capacity to conduct "bibliometry analyses" of research performance. The Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*) is currently promoting a consortium to try and close this gap.

Recommendation

Germany has a decentralized education and research system. Recognizing this, it is desirable to develop a decentralized data collection system (CRIS, *Current Research Information System*), which could then develop national standard definitions of research performance.

The Norwegian research information system (Frida) and Open Research Archives (NORA) provide a good example of what can be done. Institutions have to provide data to Frida and NORA in order to receive government funding.

4. Economy and Labor Markets

4.1 Data from the Federal Employment Agency *(Stefan Bender, Joachim Möller)*

Background and current issues

Access to labor market data was greatly improved as a result of the 2001 KVI report. The establishment of Research Data Centers and Data Service Centers has been particularly valuable. Anonymization techniques have developed rapidly and have facilitated access to data. Policy developments have provided researchers with new opportunities. Important examples are (1) the availability of data on active labor market programs required for evaluations of the Hartz reforms and (2) job search data generated as a result of the Social Code 11 reforms (2005). The research network of the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*) "Flexibility of Heterogeneous Labor Markets" has used and generated a great deal of valuable data.

Recommendations

(1) Increased use of datasets that link different types of data: economic and environmental data (AFiD) and company data from official statistics, the Bundesbank, and the Federal Employment Agency (BA, *Bundesagentur für Arbeit*) / Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung*) (KombiFID).

(2) Improvements in international datasets are also necessary, in part because of transnational movements of labor.

4.2 More and Better Data for Labor Market Research. Proposals for Efficient Access to the Currently Unused Potential of Official Statistical Data *(Hilmar Schneider)*

Background and current issues

The official labor market statistics are inadequate, primarily because they are based on the outdated idea of compiling aggregate statistics for specific purposes. The key need is for panel data at the individual and household levels. It is also important to have the possibility of making linkages between surveys. Existing deficits can be most clearly illustrated in regard to *hourly wage rates.* Accurate measurement of hourly wage rates is crucial for labor market research, but even the Income and Consumption Survey (EVS, *Einkommens- und Verbrauchsstichprobe*) does not permit accurate measurement because it only asks about wages earned during contractually stipulated, not actual work hours.

The situation has improved somewhat since Falk and Steiner made their recommendations for the 2001 KVI report, but even so, neither the Microcensus nor the Income and Consumption Survey have been developed into adequate surveys for labor market research.

Recommendations

(1) The Microcensus and the official Income and Consumption Survey (EVS, *Einkommens- und Verbrauchsstichprobe*) should contain the variables needed to calculate actual hourly wage rates. The laws governing these two surveys will be reviewed in 2012 and 2013 and should be amended to allow for this improvement.

(2) The design of official surveys should be coordinated to create possibilities for data linkage.

(3) Data linkage between surveys – including linkage between firm data and employee data – should be permitted for purely statistical purposes without the express agreement of individual respondents.

(4) Remote data access and processing should be made feasible for users of the Research Data Centers of the Federal Statistical Office and the Statistical Offices of the German *Länder*.

4.3 Interdisciplinary Longitudinal Surveys. Linking Individual Data to Organizational Data in Life-Course Analysis *(Stefan Liebig)*

Background and current issues

This advisory report is based on three fundamental insights from social science and economics: (1) The causes and consequences of individual behavior can only be satisfactorily studied with longitudinal data; (2) individual behavior is embedded in and strongly affected by social contexts and aggregates; and (3) formal organizations (e.g., firms and universities) are becoming more and more important for individual life courses. It follows from these premises that social and economic research needs a data infrastructure that provides information about individuals over time in the context of the organizations in which they live and work.

In the last eight years in Germany, there have been major efforts to provide the scientific community with linked individual and firm-level data. However, the main datasets which are currently available provide only limited information about individuals and firms and tell us nothing about the households in which individuals live. For many research purposes, including the study of social inequality, it is important to add household data to existing files.

A project currently underway at the University of Bielefeld is testing the feasibility of the proposed approach. It will assess both methods of maximizing firm/organizational participation and issues relating to confidentiality and data protection.

Recommendation

It is recommended that the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) shallcollect information about the firms and other organizations in which sample members live and work. An attempt would then need to be made to contact the organizations to collect data from them. If this project succeeded, it would yield a dataset unique in international terms.

4.4 Organizational Data *(Stefan Liebig)*

Background and current issues

Organizational data describe the central characteristics of organizations, their internal structures and processes, and their behavior as corporate actors. Data on business organizations – firms – are already widely used by researchers, and there is now increasing interest in studying other organizations, including schools, universities, and hospitals. In recent years, the official statistical organizations have made substantial improvements in providing data on firms for social science research. However, data from non-official sources are rarely available for secondary analysis. In fact, there are no adequate records of the datasets that exist, and documentation of methodological standards and quality is inadequate. These are serious deficits in view of increased demand for high-quality international comparative and longitudinal studies.

A current project at the University of Bielefeld is testing the feasibility of the approach. It will assess both methods of maximizing firm/organizational participation and issues relating to confidentiality and data protection.

Recommendations

(1) Documentation on existing non-official organizational datasets needs to be compiled and made readily accessible to researchers.

(2) Universities and publicly funded research institutes should be required to make their data available for secondary analysis.

(3) Methods and data quality indicators should be properly documented.

(4) Methods of conducting organizational surveys need to be taught in universities.

(5) A network of organizational research projects should be established, in part to deal with the implications of data protection laws and related issues of data linkage.

(6) It is proposed that a new Research Data Center be established for firm and organizational data. This center would take the lead in documenting existing surveys and archiving them. It would provide expertise in secondary analysis of organizational surveys and seek to improve methodological standards. A Research Data Center is essential for German research to come up to best international practice.

4.5 Firm-Level Data *(Joachim Wagner)*

Background and current issues

Researchers use firm-level data to document the stylized facts and assumptions used in formal models, and then to test hypotheses derived from the models. The most comprehensive data come from official sources: the Federal Statistical Office, the Federal Employment Agency (BA, *Bundesagentur für Arbeit*), and the Deutsche Bundesbank. Data from official surveys have the advantage that they cover all target firms. The firms are required to respond and respond accurately. Academic surveys have been valuable for specific purposes, but rely on small samples and limited response. Following the 2001 KVI report, the availability of data for research purposes improved markedly. Most of the important collectors of firm-level data established Research Data Centers and some offer Scientific Use Files. Furthermore, researchers can combine data from repeated surveys to produce longitudinal data on firms.

Recommendations

(1) From a research standpoint, it would be desirable to match data about firms collected by different agencies. This would require a change of law.

(2) It would also be desirable to create international datasets, in part because many firms now have global reach.

5. State, Family, and Health

5.1 Public Finance *(Thiess Büttner)*

Background and current issues

Public finance is concerned with the decisions of firms and households, not just governments. Budgetary statistics provide high-quality data relating to some government decisions and public services, but data on the *quality* of public services are generally lacking. Furthermore, tax arrangements are so complex that it is usually necessary to resort to simulation, rather than obtaining exact empirical data. The greatest future need is for datasets that combine governmental, firm, and household level data. The recent Combined Firm Data for Germany (KombiFiD, *Kombinierte Firmendaten für Deutschland*) initiative by the Federal Statistical Office, the Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung*), and the Deutsche Bundesbank, is a major development along these lines.

Recommendation

Major research advances could be made by combining governmental, household, and firm-level data. The resulting datasets would be particularly valuable for studying the impact of taxes and assessing possible tax reforms.

5.2 Household Income, Poverty, and Wealth *(Richard Hauser)*

Background and current issues

This advisory report focuses on *official statistics* relating to household income, poverty, and wealth. It characterizes the main research questions in this field, and presents an overview of available statistics and Scientific Use Files produced by the four Research Data Centers.

The author underscores the importance of a European peer review group, applying standards based on the European Statistics Code of Practice, which has already detected some problems with the statistics produced by the Federal Statistical Office and the Statistical Offices of the German *Länder*.

Recommendations

(1) Peer review groups should be set up to assess the work of all data-producing agencies, including ministries.

(2) The recommendation of the 2001 KVI report is repeated to find ways in which Scientific Use Files could be made available to reliable foreign research institutes.

(3) Specific improvements are recommended in survey methods and questionnaire design in the Income and Consumption Survey (EVS, *Einkommens- und Verbauchsstichprobe*) and the German contribution to the European Union Statistics on Income and Living Conditions (EU-SILC).

(4) It is recommended that the statistics of the various minimum benefit programs be harmonized.

(5) A single Scientific Use File is recommended relating to all minimum benefit recipients.

5.3 Family Research *(Johannes Huinink)*

Background and current issues

A great deal of progress has been made in the availability of data for family research since the 2001 KVI report. The German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) has received long-term funding, the German Life History Study (GLHS) continues to provide valuable data, and the new Panel Analysis of Intimate Relationships and Family Dynamics (pairfam) is underway. Access to the Microcensus has been much improved, which is of great benefit to family researchers.

The greatest remaining need is for improved longitudinal data. Improvements are needed at the regional, national, and international level. Data collection by official statistics could also be improved.

Recommendation

It would be valuable for the German Data Forum to provide the auspices for family researchers to work on developing an improved framework for family data collection.

This framework could be used to improve official statistics as well as surveys under academic direction.

5.4 Intergenerational Relationships *(Bernhard Nauck, Anja Steinbach)*

Background and current issues

Intergenerational relationships within families and kinship groups are a major topic of research in the social sciences. The impetus for research has come partly from changes in the family, including reduced fertility and longer life expectancy, and the implications of these changes for public policy and the welfare state. Six dimensions of social exchange are widely used in intergenerational analysis: structural, associative, affective, consensual, normative, and functional. However, despite general agreement on appropriate dimensions, there is no accepted overall theory of intergenerational relations.

Numerous large-scale German and international datasets are available for analysis. These include the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*), the German Ageing Survey (DEAS, *Deutscher Alterssurvey*), the Survey of Health, Ageing and Retirement in Europe (SHARE), and the Panel Analysis of Intimate Relationships and Family Dynamics (pairfam).

Recommendation

The aim should be to develop an overarching theory of intergenerational relationships. This requires panel data with questions which enable researchers to take a lifespan perspective.

5.5 Administrative Data from Germany's Statutory Health Insurance Providers for Social, Economic, and Medical Research *(Günther Heller)*

Background and current issues

For the last 125 years, medical care in Germany has been financed under a statutory insurance system. Eighty-six percent of the population have statutory insurance, while 14 percent have private insurance. The data reviewed were mostly collected by statutory health insurers with the aim of making correct reimbursements to health providers. Hence, they are secondary data from the point of view of a social science user.

At present, the data are only available for research within the health insurance system, or to researchers working closely with a statutory insurer. Some insurers have established databases that are anonymized at the individual level and that link different health-related contacts and treatments. But it is not clear that all insurers have such a database.

There has been no comprehensive validity study conducted on the data, but its validity is checked for its primary purpose – information that bears directly on the accuracy of invoices for reimbursement. But other information that might interest social scientists (e.g., time of medical procedures or admission diagnoses) is not necessarily carefully checked because it does not substantially affect reimbursements.

Recommendation

A detailed handlingof legal privacy provisions is important in considering potential use of these datasets by social scientists.

5.6 Provision for Old Age: National and International Survey Data to Support Research and Policy on Aging *(Hendrik Jürges)*

Background and current issues

Population aging is a key trend in all developed countries. It poses major policy problems relating to the maintenance of economic growth and to provision of adequate living standards in old age. International comparative data are particularly valuable, because diverse "policy solutions" have already been (or are being) attempted, and their results can be assessed. A large number of German and international datasets are available, including the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) and the Cross-National Equivalent File (CNEF), which includes SOEP, the

German Ageing Survey (DEAS, *Deutscher Alterssurvey*), the Survey of Health, Ageing and Retirement in Europe (SHARE), and the Savings Behavior in Germany (SAVE) survey.

Recommendations

(1) There is a need to combine conventional survey data with two other types of data: administrative data and biomarkers.

(2) It is important to extend surveys to include institutionalized people, especially those in nursing homes.

5.7 Income Provisions and Retirement in Old Age *(Tatjana Mika, Uwe Rehfeld, Michael Stegmann)*

Background and current issues

The aim is to assess the incomes of current and future retirees. Historically, most retirees have mainly relied on federal social security pensions. This implies that estimation of retiree incomes depended on administrative data from the German Pension Insurance, which included information about employment and earnings histories, and also life events affecting pension entitlements.

Recent reforms have increased the importance of occupational pensions and private savings. Accordingly, additional sources of administrative data are now required, and these data need to be linked to the German Pension Insurance (RV, *Deutsche Rentenversicherung*) data. So far this has been done in a number of official datasets and surveys, including the Completed Insured Life Courses (VVL, *Vollendete Versichertenleben*), the Old-age Pension Schemes in Germany survey (ASID, *Alterssicherung in Deutschland*), and survey data from Retirement Pension Provision Schemes in Germany (AVID, *Altersvorsorge in Deutschland*). All but the last of these datasets are available for scientific research from the Research Data Centers or the Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*).

Recommendations

(1) It is desirable to link official records – preferably from several sources – to survey data. Survey data could be particularly valuable in providing information on self-assessed health and retirement intentions.

(2) There is at present no regular procedure in place for making administrative record-to-record linkages, let alone linking to survey data. Privacy

requirements are onerous and the separate governing boards of pension funds all have to give separate approval. Nevertheless, if particular research projects using improved data could demonstrate the value of record-to-record linking, then a regular procedure might become possible.

6. Political and Cultural Participation and the Role of the Media

6.1 Political Participation – National Election Study *(Rüdiger Schmitt-Beck)*

Background and current issues

This advisory report provides an overview of recent developments in research on elections and mass political participation. Similar to other Western countries, Germany does not provide guaranteed funding for national election studies. This is a key deficit.

Recommendations

(1) A National Election Study should be established by providing the current German Longitudinal Election Study (GLES) project with continuing logistic and methodological support under the auspices of the Leibniz-Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*). Permanent long-term public funding, with the study institutionalized at GESIS, is the desirable long-term outcome.

(2) A small number of political variables should be tagged for inclusion in all surveys conducted in German General Social Survey (ALLBUS, *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*) and the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*).

(3) The data services of the statistical offices should be modified to meet basic requirements of research on elections and political participation.

(4) Public agencies should be under a formal obligation to deposit survey data collected under their auspices into appropriate archives.

6.2 Civil Society *(Mareike Alscher, Eckart Priller)*

Background and current issues

Available data on civil society organizations (CSOs) remain seriously inadequate. To a large extent, researchers have to compile data from other sources which were not primarily designed to provide valid data on CSOs. However, considerable progress has been made through Germany's participation in the Johns Hopkins Comparative Nonprofit Sector Project. This project sets out the data requirements that would need to be met in order to provide a valid description of CSOs and their activities. More recently, the Civil Society Data Collection Project has set out to provide a reporting system for Germany, using the concepts of the Johns Hopkins project and data (mainly) from the Federal Statistical Office. The mid-term goal is to establish a National Accounts satellite system for CSOs.

Recommendations

(1) The long-term goal should be to set up a comprehensive, self-contained data provision system for CSOs.

(2) This goal can be reached by using existing surveys and data sources, including improved data from the CSOs themselves, and by adding questions about civic engagement to ongoing surveys, especially the annual Microcensus.

6.3 Culture *(Jörg Rössel, Gunnar Otte)*

Background and current issues

The expert report focuses on culture defined as the arts. Research on the arts falls into three categories: artistic production and its organization, the distribution and economic valuation of culture, and the consumption of culture. Sociology and economics are the two main social sciences in which the arts are studied.

Recommendations

(1) Two large baseline surveys are needed: (a) a survey providing life-course information on artists, as well as information about their current work, status, earnings, etc., and (b) a representative sample survey of cultural consumption.

(2) Development of a single national cultural statistic as set out in the study *Kultur in Deutschland* and compatible with efforts at the EU level.

(3) *Publicly* financed surveys on culture, including those conducted by statutory bodies, as well as citizen surveys and audience surveys, should be deposited at the Data Archive of the Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*) and be available for secondary research.

6.4 Mass Media Research *(Heiner Meulemann, Jörg Hagenah)*

Background and current issues

Mass media research focuses on both the production of media “programs” and their consumption. Content analysis is the main method used to analyze programs. Surveys, including time budget surveys, are used to analyze consumption. Several archives store media programs. The German National Library (*Deutsche Bibliothek*) in Frankfurt holds a copy of every newspaper published. Various public and private agencies archive and analyze the content of television and radio media. The largest of the private agencies is *Media Tenor*, which has conducted content analysis of programs from about 700 sources since 1993. Both public and private agencies also analyze media consumption; the private agencies being motivated partly by demand for advice on communications and advertising outlets.

The Federal Statistical Office includes time spent consuming media in its Time Budget Studies (1991 and 2001). International data on media consumption have been collected by the Eurobarometer and the European Social Survey (ESS).

Recommendations

(1) It is recommended that a central media content archive be set up for Germany. This should include data collected by public and private agencies, and by individual researchers.

(2) Common content analysis categories should be developed, in part to facilitate international comparisons.

(3) The professional societies of the social and communication sciences should attempt to secure access to important surveys funded by media stations, as well as privately funded surveys.

6.5 Judicature *(Wolfgang Heinz)*

Background and current issues

An adequate system of crime statistics would enable us to answer questions about: (1) trends in the incidence of different types of crime, (2) the decisions of the authorities relating to prosecution, (3) the numbers and types of criminal sentences/penalties imposed, (4) the extent to which penalties are enforced, and (5) rates of reconviction/recidivism.

Assessed by these standards, current official German crime statistics are seriously deficient. As a result, it is currently necessary to supplement official statistics with periodical crime and victimization surveys. Similarly, prison statistics need supplementing with statistics about suspects who face preliminary proceedings. Additional ways also have to be found to collect data on the enforcement of criminal sentences and on reconvictions.

Recommendations

(1) A comprehensive crime statistics database would need to contain all police data on crime and all relevant judicial decisions. Data on individuals would need to be "pseudonymized" and then linked.

(2) This database would need to be regularly updated, in particular with respect to enforcement of sentences and reconviction/recidivism. It would then be possible to assemble case flow statistics and to conduct cohort studies.

6.6 Environment *(Cornelia Ohl, Bernd Hansjürgens)*

Background and current issues

Environmental problems are large in scale. They are typically long-lasting and also have wide geographical impacts. Furthermore, their impact is often subject to "true uncertainty"; that is, there is insufficient knowledge of damages and costs, and the probability of damages and costs is unknown. The complexity of the problems means that innovative research methods and modeling approaches are needed to supplement traditional monitoring methods used for assessing environmental impacts.

Recommendations

(1) Geographical Information System technology should be used to enhance mapping of environmental impacts. These impacts need to be shown in

relation to a range of socio-economic indicators mapped at the appropriate scale. The biggest challenge lies in mapping global climate change.

(2) It is also necessary to evaluate policy responses to environmental challenges and assess the vulnerability of affected social units.

(3) A nested data structure is needed in order for researchers to be able to assess developments from a polluter's point of view, a victim's point of view, and also a regulator's point of view.

ADVISORY REPORTS

TOWARDS AN IMPROVED RESEARCH INFRASTRUCTURE
FOR THE SOCIAL SCIENCES:
FUTURE DEMANDS AND NEEDS FOR ACTION

1. Providing a Permanent Institutional Guarantee for the German Information Infrastructure

Johann Hahlen

Contact:

Johann Hahlen, Staatssekretär a.D.
Am hohen Rain 52
50389 Wesseling-Urfeld
Germany
e-mail: johann.hahlen[at]web.de

1. A permanent information infrastructure must be tailored to the specific German data situation. This means it must – to the extent possible – take account of the variety of data, the multitude of data producers and, especially, the potential domestic and foreign users and ways they intend to use the data. Furthermore, it must be open to topics of future interest and new questions.

When setting up a permanent information infrastructure, it must be kept in mind that there is a network of interactions among the various datasets and data users, which in Germany is determined by a number of legal and structural conditions. It is realistic to say that these conditions cannot be changed and, consequently, it is reasonable to treat them as given. Any ideas for scientific policy regarding a permanent information infrastructure should take account of the following conditions:

- For natural persons, the German Constitution (GG, *Grundgesetz*) grants the right to informational self-determination protecting individuals from unlimited collection, storage, use and transmission of their personal data and safeguarding the individual's right to decide on the disclosure and use of personal data. Although the GG does not require data collected for statistical purposes to be strictly and concretely linked to a specific purpose, it does set limits on the information system. Transmitting statistical data for scientific use is permitted by the Constitution if limited to what is necessary for the particular scientific purpose, if no direct reference is made to individuals (no names or addresses), and if the recipient of the data does not have any additional information that could allow re-identification of the individual and thus result in a violation of the individual's right to informational self-determination. This was laid down by the Federal Constitutional Court in its fundamental population census judgment of 15 December 1983 (BVerfGE[1] 65, 1 et seqq.). This requirement is met by the clause relating to the scientific community in § 16 Abs. 6 of the Federal Statistics Law (BStatG, *Bundesstatistikgesetz*).[2] Local units, enterprises, and legal entities engaged in economic activity cannot claim the right to informational self-determination. However, they are protected by the right to perform business activities, which is also granted by the Constitution.

- The scientific use of personal data and of data on economic entities must comply with these constitutional principles, the numerous legal provisions on the collection and use of statistical data, and regulations protecting local units and enterprises with regard to their

1 Decision of the Federal Constitutional Court (BVerfG, *Bundesverfassungsgericht*).

2 The citations to German legal sources have been left in German to guarantee accuracy.

economic activity (e.g., the protection of business secrets or fair and open competition on the market).

- Data producers and data holders – to the extent that they are part of the public administration, for example, government authorities or institutions – are bound by the principle of the rule of law according to Art. 20 Abs. 3 GG. No information infrastructure of any kind and no scientific demand can exempt such data producers and data holders from complying with the above regulations.

- This remains unaffected by the freedom of science, research, and teaching guaranteed by Art. 5 Abs. 3 GG. It is true that the Federal Constitutional Court has interpreted the basic right of the freedom of science (Art. 5 Abs. 3 GG) to entail a government obligation to provide efficient institutions to ensure the freedom of science and the relevant teaching. However, this does not mean that an individual scholar can claim the right to access specific data stocks. Furthermore, it does not mean that a scholar's research might take priority over the legal protections afforded to individuals or enterprises.

- Germany is a federation (Art. 20 Abs. 1 GG) in which the exercise of state powers is generally a matter of the *Länder* (federal states) (Art. 30 GG). The *Länder* are generally responsible for executing federal laws (Art. 83 GG). The Federation, which – according to Art. 73 Abs. 1 Satz 11 GG – has the sole legislative power for "statistics for federal purposes," was allowed by Art. 87 Abs. 3 GG to establish the Federal Statistical Office as an independent superior federal authority. However, the federal legal provisions on official statistics are implemented by the *Länder* through their own administration (Art. 84 Abs. 1 GG). At the same time, Germany has committed to the project of European integration, and has transferred sovereign powers to the European Union (Art. 23 Abs. 1 GG), so that EU Regulations and Directives are directly applicable in Germany or have to be transformed into German law. Therefore, EU Regulation No. 223/2009 on European Statistics (former EU Regulation No. 322/97 on Community Statistics) and EU Regulation No. 831/2002 concerning access to confidential data (of the EU) for scientific purposes are directly applicable in Germany.

- Finally, the principle of democracy, which is explicitly referred to by the German Constitution (Art. 20 Abs. 1 and 2, Art. 21 Abs. 1, Art. 28 Abs. 1 and Art 38 Abs. 1 GG), requires a free, open, transparent and discursive process of forming opinions, which needs both the knowledge of the facts relevant for the decision-making concerned, especially the data available, and the scientific examination and processing of those facts.

2. There are many indications of the need for a permanent information infrastructure in Germany. What is needed is not just a free, non-government press and radio landscape and a free, self-determined research system but also access to data from official statistical institutions and – if possible – any other data stocks collected for government purposes. At the same time, it is crucial to safeguard the legal rights of the entities to which the data refer (individuals, local entities, enterprises). However, these needs – which are easy to reach consensus on in abstract terms – are confronted with a number of very real weaknesses:

- On the one hand, as Chancellor Angela Merkel once stated, any policy starts with the facts. On the other, we could just as easily quote the former Prime Minister of Saxony Biedenkopf, who talked about the widespread "resistance to facts" among politicians. On that point, Keynes said the following: "There is nothing a government hates more than to be well informed; for it makes the process of arriving at decisions much more complicated and difficult." Probably, however, the impression of a "resistance to facts" is merely due to the fact that those in power think they are sufficiently informed already, while frustrated statisticians and social scientists overestimate the importance of their findings.
- In any case, it is obvious that empirical social and economic research in Germany has been severely underfinanced for a relatively long time compared with other branches of empirical science, such as medicine and other natural sciences.
- What is more, in some areas, there has been obvious reluctance on the part of German economists to engage in empirical work.
- And finally, by no means only in Germany, there is a certain reluctance among scholars to scrutinize their own work for reproducibility and falsification. This, combined with the tendency – though perhaps simply a human one – towards competition and isolation, may prevent these scholars and institutions from obtaining the high infrastructural investments they need from government agencies.

3. Based on these conditions and structural constraints, the permanent information infrastructure that is needed can be defined in both negative and positive terms.

3.1 This is what a permanent information infrastructure in Germany should *not* be:

- The public data producers and data holders belong to different levels of state administration that are, in many cases, structured by *Länder* or other regional units. For example, in addition to the Federal Statistical Office there are 14 *Länder* offices producing and storing statistical data. The Federal Employment Agency (BA, *Bundesagentur für Arbeit*) and the German Pension Insurance (RV, *Deutsche Rentenversicherung*) are part of the indirect federal administration; education data are stored by the competent *Länder* ministries; the Central Register of Foreigners belongs to the Federal Office for Migration and Refugees (BAMF, *Bundesamt für Migration und Flüchtlinge*), which is a superior federal authority; the population registers are available at the towns and municipalities or at central *Länder* population registers. Health data come from some 100 widely varied sources.

  This patchwork is anything but convenient for anyone interested in data for scientific use; it is confusing and labor-intensive at best. The idea of an institution that is comprehensive both regionally and in terms of subject matter therefore seems obvious, but it turns out to be an unachievable vision.

  As experience shows, the various data producers and data archives in Germany are not willing to transmit their data to third parties or even to grant third parties the right of use. At most, they are willing to be represented by a regional partner (one *Land* for several or all other *Länder*). A highly structured information infrastructure is certainly not convenient, but modern information technology offers the potential for cross-referencing that can make things clearer and provide orientation in the maze of data providers.

- A register comprising all data, such as a large central archive where all the data producers and data archives store duplicates of their data, would theoretically be a solution to the dilemma described above – but it is impossible due to the legal situation. This is because, in Germany, data are strictly linked to a specific purpose in order to protect the individuals, local entitites, or enterprises referred to by the data. This means that, already during data collection, respondents must be informed what specific purposes their data are being collected for and who will receive access. Transmission of originals or duplicates to a "central scientific register" has not been dealt with legally and will only be possible to regulate in future legislation. Consequently, no stock data can be stored in such a register unless all respondents give their consent, which makes the whole matter unrealistic. Exceptions are not permitted by the clause relating to the scientific community as stated, for example, in § 4a Abs. 2 of the Federal Data Protection Act.

However, statistical data which are processed and used only in an anonymized form do not need to be linked to a specific purpose, so that they can be used for scientific purposes if anonymity (even de facto anonymity) is safeguarded. This does not yet make it possible to set up a comprehensive central scientific register because what could be stored there would only be aggregated data and microdata in a de facto anonymized form. Although the latter is possible – with sometimes considerable efforts – for specific data stocks such as the Microcensus, this is not possible for all official statistical data. Therefore, a central register limited to statistical data would be highly incomplete. The health monitoring system operated by the Robert Koch Institute and the Federal Statistical Office is not an example to the contrary because it uses only aggregated data from the various sources.

- What should not be envisaged to guarantee the information infrastructure is the creation of a new federal authority or – either in addition or alternatively – of new *Länder* offices. First, for the reasons shown above, they could not represent a central register. Second, this would involve considerable bureaucratic efforts; they would have to be integrated into existing divisions of responsibility and hierarchies, they would have to acquire the required wide range of special knowledge on the various data stocks and would be limited to coordinating activities, while scientific data users would still have to deal with the relevant data producers and data holders.

- The same reasons apply to attempts to establish an information infrastructure on a permanent basis through a university institute of some kind or through one or several professors. The existence of the Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*) and its practical success at the same time show the limits of such institutions. A university institute or a team of scientists would not be able to cope with these requirements.

- Also, it is not promising to use externally funded private institutions to establish a permanent and secure foundation for the information infrastructure. As experience in Germany shows, the financial resources of potential users (from the scientific community) would not be sufficient to pay the considerable staff required for such institutions to offer services that meet the wide range of requirements. The empirical social and economic research community cannot be expected to obtain enough funding from the relevant organizations in the near future to permit such institutions to be established or maintained.

3.2 What should a permanent information infrastructure in Germany look like if the models rejected here are not an option and if the goal is to achieve maximum use for data users, especially from the scientific community?

- Considering the possibilities of modern information technology, the infrastructure must be available online seven hours a day, 365 days a year. Where online use is not possible because of data protection and statistical confidentiality, local workstations should be kept available for use at common universities hours.
- The infrastructure should ensure that it is equally open to anyone interested and that it is neutral, i.e., that it does not assess or censor user requests. It should be independent in its methodological work and be based on accepted scientific standards. The openness, neutrality and methodological independence should each be ensured through supervision by a committee comprising representatives of data producers and scientific data users as well as the responsible data protection commissioner.
- The infrastructure should be sufficiently equipped with staff and material to fulfill its tasks. At the same time, it should be lean and economical, so that it can be used without insurmountable financial obstacles. Its work should be rationalized through permanent evaluation of its processes and through wide-ranging use of the appropriate information technologies.
- Considering the complex subject matter and regional structures of data production and storage in Germany, and the fact that centralization is unachievable, the infrastructure should be structured in terms of subject matter, it should cover all of Germany and it should be broken down into regions only to the extent absolutely required (e.g., by *Länder*).
- Although the infrastructure should be set up permanently, it should also be able – for example, through revision clauses – to react flexibly to changes in data availability and in the demand from the scientific community.
- In all these areas, in practical work it is necessary for the infrastructure institutions to achieve an optimal reconciliation between, on the one hand, the legitimate interests of data producers and data

archives as well as the rights – protected by provisions on data protection and statistical confidentiality – of the individuals, local entities, and enterprises referred to by the data and, on the other hand, the interests of the scientific users. Constantly keeping this in mind will be one of the main tasks of the committee set up by the relevant infrastructure institution, in addition to the tasks mentioned above.

4. The institutions set up in Germany on the basis of the recommendations of the KVI report of March 13, 2001, and with considerable support from the Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*) have basically proved successful:

4.1 The German Data Forum (RatSWD), where data producers and data users work together, has developed into an institution creatively enhancing the information infrastructure in Germany. In this council, representatives of the Federation and the *Länder* co-operate with individuals elected in a "grassroots" manner from the scientific community. Therefore, its proposals are practical and are welcomed by the community. Apart from its internal work, such as exchanging ideas with the major institution funding research (BMBF) and carrying out evaluations for official statistical institutions, the German Data Forum (RatSWD) is engaged in many external activities that have become important elements in the information infrastructure in Germany and should be continued.

What should be mentioned first of all here is the Conference for Social and Economic Data (KSWD, *Konferenz für Sozial- und Wirtschaftsdaten*). At this important event, which is held at regular intervals, council members are elected from the scientific community and research results are presented that have been obtained using the available data stocks. This provides a basis for discussing gaps identified in the information infrastructure.

Important suggestions towards improving the information infrastructure are given by the expertise contests organized by the German Data Forum (RatSWD) and the working papers and newsletters it publishes.

4.2 The most important progress that has been made since the 2001 KVI report, has been the establishment of the four Research Data Centers and the two Data Service Centers.

- The Research Data Center of the Federal Statistical Office was founded in 2001 – as the first Research Data Center in Germany –

and was positively assessed in 2004. It allows empirical social and economic researchers to access official statistical microdata, while safeguarding statistical confidentiality. For that purpose, the Research Data Center makes Public Use Files, Scientific Use Files, and CAMPUS-Files available for off-site use by the research and teaching community. Guest researchers can use less strongly anonymized data on the premises of the Federal Statistical Office in Wiesbaden, Bonn, and Berlin. Also, scholars can use data stocks of the Federal Statistical Office by means of controlled teleprocessing (on-site use).

- The decentralized Research Data Center of the Statistical Offices of the German *Länder* was set up in April 2002 and positively assessed in late 2006. It offers scientists the same access to official statistical data as shown above for the Research Data Center of the Federal Statistical Office. Subsequent to an amendment of the BStatG, the Statistical Offices of the German *Länder* established a system of centralized data storage for the whole of Germany for this purpose, with a breakdown by subject matter.

- The Research Data Center of the Federal Employment Agency (BA, *Bundesagentur für Arbeit*) was established in April 2004 at the Agency's Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung*) in Nuremberg and has also been assessed positively. It makes the large data stocks of the Federal Employment Agency available for scientific analyses within the scope of Art. 75 of Volume X of the Social Code.

- The Research Data Center of the German Pension Insurance (RV, *Deutsche Rentenversicherung*) was also established in 2004 at two locations in Berlin and Würzburg. The Scientific Use Files produced there with regard to the statistics of new and existing pensions and the statistics of persons insured allow, for the first time, scientific evaluation of the vast data treasures of the German Pension Insurance.

- The two Data Service Centers – also based on the 2001 KVI report – were established in 2003 at GESIS in Mannheim and at the Institute for the Study of Labor (IZA, *Forschungsinstitut zur Zukunft der Arbeit*) in Bonn. The Data Service Center at GESIS works under the name of German Microdata Lab (GML) and offers a service and research infrastructure for official microdata. The International Data Service Center for Labor Market Relevant Data (IdZA, *Internationales Datenservicezentrum für arbeitsmarktrelevante Daten*) at the IZA provides labor market researchers with a metadata portal for existing data; it has developed a special web-based tool (JoSuA) for data access via controlled teleprocessing.

All Research Data Centers and Data Service Centers have been welcomed enthusiastically by the scientific community and are used extensively for research and teaching, with the two Research Data Centers of official statistics having observed a marked recent shift in the demand for means of access to their data stocks. While the – initially very high – demand for Scientific Use Files has been declining, demand is increasing for individual datasets, which guest researchers can access from protected scientific workstations at the Research Data Centers, and for controlled teleprocessing.

The encouraging practical efficiency of the Research Data Centers has been achieved in two ways:

- Thanks to start-up financing from the BMBF, the Research Data Centers have made a wealth of official statistical data stocks available to the research and teaching community by producing Public Use Files, Scientific Use Files, and CAMPUS-Files, by offering safe scientific workstations for guest scientists, and by offering controlled teleprocessing.
- The financial obstacles to use of data of official statistics that existed in the 1990s – which were insurmountable in some cases for empirical social scientists – have been removed, thanks in part to start-up financing provided to the Research Data Centers by the BMBF. For example, in the mid-1990s the Statistical Offices had to charge as much as DM 30,000 (about EUR 15,000) per Scientific Use File of the Microcensus to cover the considerable production costs. Since the emergence of Research Data Centers, a social scientist can obtain such a Scientific Use File for a "charge" of just EUR 90 including the CD and shipping.

4.3 The information infrastructure developed since the 2001 KVI report also includes many larger and smaller projects and initiatives of widely varied institutions. These include:

- Every year since 1999, the Federal Statistical Office has been granting the Gerhard Fürst Award for dissertations and diploma/master theses dealing with empirical questions and using official statistical data.
- The Statistical Offices of the German *Länder* have set up branches of its Research Data Center at the German Institute of Economic Research (DIW Berlin, *Deutsches Institut für Wirtschaftsforschung*) and at Dresden Technical University.

- At its conferences, the German Statistical Society (DStatG, *Deutsche Statistische Gesellschaft*) offers workshops for junior scholars to introduce them to empirical work with the various data stocks.

5. Despite all the progress made so far, there still is much to improve and numerous problems that remain to be solved. We have not yet succeeded in developing a firm institutional foundation for the information infrastructure in Germany. Financial and content-related problems need to be solved.

5.1 Financial problems appear to be the most urgent issue at present and, although they are not so excessive in volume (the Research Data Centers of the Statistical Offices of the German *Länder*, for instance, reckons with total costs of only about EUR 3.7 million for the 2 1/2 years from 1 July 2007 to 31 December 2009), they are difficult but can be solved.

- The structures created on the basis of the 2001 KVI report (especially the RatSWD with its business office in Berlin and the four Research Data Centers) owe their establishment to the support provided by the BMBF. This was temporary project support in the form of start-up financing that requires the relevant institution to contribute funds of its own, considering the benefit it draws from the project.
- The German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*), which meanwhile is 25 years "old," has been financed institutionally since 2004. Thus an important recommendation of the KVI has been implemented and its work can be regarded as permanently guaranteed. In contrast, such institutional support seems out of reach for the Research Data Centers but it is not strictly necessary.
- The financial situation of the Research Data Centers varies considerably at the present time.
- In the beginning, the Research Data Center of the Federal Statistical Office was financed mainly by the BMBF. Meanwhile its core business, answering and handling user requests from the scientific community, is funded completely out of its own budget. The Research Data Center receives BMBF funds only for research projects to extend the data supply it offers, for instance by anonymizing panel data from economic statistics.

- Most of the funds required for the Research Data Center of the Statistical Offices of the German *Länder* will be provided by the BMBF up to the end of 2009.
- The Research Data Center of the BA at the IAB was partly financed by the BMBF and since the beginning of 2007 has been funded entirely by the BA.
- The Research Data Center of the German Pension Insurance will be supported by the BMBF until the end of 2008.

Consolidation and a uniform financing line for the Research Data Centers are therefore urgently needed. On the one hand, they would have to guarantee the ongoing existence of the Research Data Centers and the further development of their data supply. On the other, the Research Data Centers should not charge prices that users cannot afford. It is thanks to the 2001 KVI report and the project support by the BMBF that this – harmful – situation no longer exists in Germany. After all, the scientific community should be able to use the respective data stocks for research and teaching purposes. At the same time, one will have to accept that the BMBF generally confines itself to temporary start-up financing and regards the respective data archives and scholars as responsible thereafter.

Therefore the organizations supporting the Research Data Centers, the empirical social and economic research institutions and the BMBF should agree on the following model, which should entail sustainable financing of the Research Data Centers at affordable prices for their users:

- The respective organizations supporting the Research Data Centers, for example of the Federal Statistical Office and the Statistical Offices of the German *Länder*, will take over the basic financing of their Research Data Centers.
- The further development of methodology and special research projects of the Research Data Centers will continue to receive project funding on a temporary basis, provided that these are important for an expansion of the information infrastructure.
- The Research Data Centers will charge users to cover the expenses incurred in each case, but there will be far-reaching possibilities to reduce prices for financially "weak" users such as PhD candidates or university institutes, while "well equipped" users, for instance economic research institutes, which can pass on their expenses to their clients, will have to pay prices fully covering the expenses.

For the Federal Statistical Office and the Statistical Offices of the German *Länder* to support this solution, it would be advisable to amend the Federal Statistics Law, making it clear that the mandate of official statistics also includes the provision of data (both aggregated data and microdata) to the scientific community. The inclusion of such a provision into one of the next bills on statistical issues should be supported at the Federal Ministry of the Interior (BMI, *Bundes-ministerium des Inneren*). Once the cooperation of the statistical offices in an Research Data Center with jointly held data is made possible by the Federal Statistics Law in 2005 (through its § 3a Abs. 2 and § 16 Abs. 2), the Research Data Centers of the official statistical agencies would thus be enshrined in law and their funding would be indirectly guaranteed.

5.2 As regards its contents, the information infrastructure that has emerged in Germany since 2001 provides numerous starting points for expansion and consolidation. Depending on the perspective, different institutions prioritize one point or another. Priorities and posteriorities should be discussed in the German Data Forum (RatSWD) and a medium-term consolidation and extension program should be set up, focusing not only on what would be desirable but also on what chances there are to implement it. The initiatives listed in the following are therefore not listed in order of preference.

- The existing four Research Data Centers are far from opening up all data stocks that are of interest to empirical social and economic research. This is why there should also be Research Data Centers, for instance, for health, education, and media data. Other major fields awaiting investigation are crime control and the administration of justice and penal administration, for example, using the criminal statistics of the police and judicial statistics. The situation is similar with the Central Register of Foreigners kept at the Federal Office for Migration and Refugees, the business register of the Federal Statistical Office, and the population registers of the municipalities and the *Länder*. Finally, provisions have to be made in time for the scientific use of the data that will be collected in the EU-wide population census scheduled for the year 2011. In each of these areas, it would have to be determined whether Research Data Centers should be set up and, if so, how this can be fostered.
- As there are different Research Data Centers, each of them restricted to specific data stocks, it is crucial that a “special” Research Data Center be set up that combines the data stocks of various data producers or makes it possible to work with the data from different producers. A similar goal is pursued by the proposal to create a kind of

"data trust" keeping data stocks from various subject fields and making them accessible to the scientific community through the channels known from the Research Data Centers. Advantageous as both ideas may be from the viewpoint of empirical social and economic research, the obstacles of data protection legislation appear insurmountable so that one should not "fight a losing battle" here.

- Such a solution might be considered, if at all, for statistical data whose collection does not have to be strictly linked to a specific purpose. But then the data kept there would have to be at least de facto anonymized. This, however, would probably not be worthwhile. Also, it should be kept in mind that combining de facto anonymized personal data from different statistics increases the chances of reidentification, which is precisely what must be prevented.

- Non-statistical data, however, have to be strictly linked to a specific purpose. This means the following. First the data – and also the microdata – of the various producers would have to be transferred to the "special" Research Data Center or the "data trust." So far this would generally not be covered by the respective data collection purpose and would therefore be illegal. The clauses relating to the scientific community as contained in the German Federal Data Protection Act (e.g., § 14 Abs. 5 Satz 2) do not permit such data transmission and storage because the research purposes can actually be achieved with reasonable efforts even without a "special" Research Data Center or without a "data trust." The proposal to appoint the data protection commissioner in charge as trustee does not solve the problem. Apart from the fact that the Federal Commissioner for Data Protection has already dismissed such ideas for his institution, the unsolvable problem of having to alter the purpose would persist. If – despite all practical obstacles – the consent of all concerned to such a purpose-altering transfer could be obtained, reservations would remain because contrary to the order of the Federal Constitutional Court, the data would not be de facto anonymized at the earliest possible time.

- In view of this legal situation, it would be advisable to invite an expert, for example, from the Federal Employment Agency or its Institute for Employment Research (IAB) to the Research Data Center of the Federal Statistical Office and to entrust him or her with "data processing by order" – that is, with the evaluation of statistical data in combination with data of the Federal Employment Agency in relation to a specific issue. The Research Data Center of the Federal Statistical Office plans to do this with regard to the data of the Federal Employment Agency.

- Given the legal consensus that the clause on the scientific community in § 16 Abs. 6 of the Federal Statistics Law does not cover foreign universities or foreign scientists, the information infrastructures created in Germany to date have not furthered scientific cooperation with foreign countries, and this also holds true for the EU. It is true that there is now a "Safe Centre" at Eurostat in Luxembourg, whose establishment was made possible by Regulation (EC) No 831 / 2002 (concerning access to confidential data for scientific purposes). However, German statistical microdata would be available there only if they had been submitted to Eurostat, as well, which is an exception. The establishment of such an "EU Safe Centre" in Wiesbaden, which is planned by Eurostat together with the Federal Statistical Office, will therefore not bring any improvements for foreign scientists. To enable cross-border scientific work, empirical social and economic researchers should call for an extension of § 16 Abs. 6 of the Federal Statistics Law to cover foreign scientists. Article 23 of the new EU Regulation No. 223/2009 on European Statistics grants researchers access to confidential data which only allow for indirect identification of the statistical unit for scientific purposes.
- There has been no progress in the last few years regarding the 2001 KVI recommendation to introduce a research or scientific code of confidentiality. The restraint shown in responding to this suggestion may be due to the fact that such a research data secret would have to entail the scientist's right to decline to answer questions, and the prohibition of seizure. However, this recommendation still deserves to be studied in detail. Because of the complexity of the matter, the German Data Forum (RatSWD) should set up a working party for the purpose. After the recent cases of data abuse at a large telecommunication provider and in call centers, serious proposals have been put forward calling for a codification of the right to informational self-determination and of a right to privacy of information technology records. If these attempts should materialize, the scientific community would have to promote its interests in an elaborate proposal to introduce a research or scientific secret. Progress in this difficult matter might be easier if a code of conduct existed for scientists interested in using the data stocks, paired with the possibility to impose sanctions, which was also recommended by the 2001 KVI report. The RatSWD should also take steps in that direction, together with the other scientific institutions.
- Finally, the 2001 KVI report deserves further attention, since it aims at an expansion of empirical social and economic research (includ-

ing university education on this type of research). Beyond the establishment of "empirical economic research" as a university subject, there is a sufficient number of current problems justifying, for instance, the creation of special research areas (e.g., on questions of health and education policies) or of professorships for empirical work (co-) financed by trusts.

- When the information infrastructure has been established on a permanent basis, it will be important to carry out continuous checks for "proliferation," overlaps, duplication of labor, and the like. This should take place in the course of, and apart from, the now common and rather strict periodical evaluation of the facilities created. Experience shows that these problems are likely to arise, especially with new developments, and that the readiness to make necessary changes may still be lacking. In particular, the informational structures resulting from the federal system should be analyzed in this respect.

2. The European Dimension

Klaus Reeh

Contact:

Klaus Reeh
Adviser to the Director-General, Eurostat
Bâtiment BECH
5, rue Alphonse Weicker
2721 Luxembourg
e-mail: klaus.reeh[at]ec.europa.eu

Abstract

The purpose of this report is to identify how to better meet the needs of scientists and take into account their concerns around the use of economic and social data at the European level without compromising or neglecting the legitimate needs and justified concerns of European policy-makers.

1. Background

The volume and type of data that can be made available to scientists depend increasingly on targeted initiatives and general developments at the European level. These initiatives and developments are primarily motivated by the need for official statistics to serve the purposes of (European) policy making. A number of scientific needs are met because they overlap with the needs of European policy-makers:

- Comparability across national borders is of central importance for both policy-makers and scientists.
- Policy-makers and scientists both benefit from coordinated program planning between the Member States since this is the only way to have corresponding statistics on hand for all Member States.

Other scientific needs and concerns, however, are either at least partially at odds with the (legitimate) needs and concerns of European policy-makers or have a significantly different priority level:

- In science, for example, accuracy is usually more important than how recent the information is; the opposite is true for policy making. While policy-makers are often under pressure to make snap decisions, the world of science faces such time pressure only in exceptional circumstances.
- Methodological stability over time is often more important in science than the ability to adequately address up-to-the-minute political and institutional situations; the opposite is true for policy making. While policy-makers normally have to base their arguments on what is at play in the current situation, scientific perspectives draw from longer periods of time.
- Complex statistical procedures do not pose a problem for science; scientists often even demand them. For policy making, however, there are limits to complexity because it complicates communication.

- Scientists are always looking for new concepts that must then also be described statistically, whereas policy-makers cannot but prefer to work with well-established concepts. Conceptual innovation is a necessity for science, but is subject to limitations in the field of policy making.

Scientists have needs that can be satisfied without obstructing the needs of European policy-makers. Nonetheless, these needs are often neglected. An important reason for this is that they have not been, and are still not sufficiently emphasized by scientists themselves.

- Access to (anonymized) microdata has become increasingly important for science. The behavior of individual actors or groups of actors has become increasingly interesting for economic and social science research, particularly with a view toward improving what are still too frequently the rather simplistic assumptions within economic science itself, and to overcome the divide between micro and macro analysis. In contrast, the use of microdata is of limited importance for European policy making (or for European administration), and is sometimes excluded entirely.

Although policy-makers and scientists often have overlapping interests in and needs for (European) statistics, it must nevertheless always be borne in mind that – to borrow from the language of sociology – the "science system" and the "political system" follow differing logics and principles. Science (empirical science) endeavors to adopt at least a denationalized or even global approach in order to avoid politicization. Policy making, on the other hand, must remain to a large extent national and, by definition, also political, even where there is an attempt at depoliticization, which is made not least by pointing to the inherent necessities that can be substantiated by statistical evidence. Furthermore, (empirical) science constantly strives for neutrality in its value system; in contrast, policy making cannot escape value judgments – indeed, value judgments are its business.

Official statistics, which are after all part of both systems, can easily risk being torn between the two different fields and end up satisfying neither of them. To make matters more difficult at the European level, official statisticians usually have a much more general mission at the national level and are much freer to decide how to accomplish their mission than would be legally possible at the European level. It is therefore desirable for scientific research policy in particular to look into this issue and to support a broader spectrum of responsibilities for European statistics, which would make it possible to provide European statistics also for domains without a specific *political* competence at EU level.

2. A few specific problem areas

A number of specific barriers stand in the way of both the extensive, appropriate supply and the sensible use of European economic and social data by scientists. Below is a non-exhaustive, brief outline of some of these barriers. They are not listed in order of importance. Reference is made first to more technical barriers and then to barriers which are more organizational in nature.

- The purpose of European statistical policy is to organize official statistics in such a way that the information needed to implement European policies (for the appropriate exercise of European competences) is available. It follows that European statistics can cover only those areas for which a European political competence exists. It is therefore not a comprehensive system, and has never been presented as such. This incompleteness is frequently regretted by scientists, but is difficult to remedy at the European level, since the European Union does not have full competence in the field of statistics and the European Commission does not have the corresponding right of initiative to create an all-encompassing European statistical system.
- The harmonization of official statistics is the main focus of the European statistical policy. However, each harmonization brings with it inevitable discontinuity, at least in some Member States. Temporal continuity is sacrificed in favor of improved geographical comparability. Yet continuity over time is particularly important for science (time series econometrics). Scientists (generally more than policy-makers) therefore press for retroactive calculations of harmonized statistics.[1] These are very costly and therefore cannot be carried out without a specific request.
- On the other hand, the harmonization of individual statistics repeatedly encounters various limitations which result not least from these statistics being anchored within the different national systems and their basic respective orientations. Even when policy-makers consider individual harmonization results to be acceptable, scientists often find fault with them: the process of "output harmonization" often suffices to achieve data convergence for analyzing problems of "practical policy making," whereas it is all too commonly believed that "rigorous science" requires "input harmonization" in order to obtain secure findings. However, the content

1 The treatment of changes to territorial boundaries is a similar issue. Here again, scientists push for retroactive calculations or for the old territorial boundary to continue to be used.

superiority of “input harmonization” has not been clearly established and requires expensive comparisons, while the lower costs of “output harmonization” are a definite advantage.

- One technical (but also policy) problem is posed by the incoherence of data related to cross-border issues, such as flows between countries or entitlements with cross-border validity (e.g., rights). Particularly in the case of sample surveys due to the sampling error, but also in exhaustive surveys, exactly identical results in the country of arrival and the country of departure cannot be expected for a number of reasons when statistically measuring exactly the same flow. The same applies to the allocation of entitlements. This problem is indeed inconvenient for policy making, but is not considered too serious for the decision-making process, whereas in science it is seen to undermine research possibilities and the accuracy of conclusions.

- The growing complexity of official statistics has been brought about by the methodological and definition-related cross-linking of specialized statistics. On the one hand this is necessary, for instance, in order to develop a system of national accounts, which is important for policy making, and particularly for European policy. On the other hand, it impedes the targeted pursuit of specific scientific questions because it leads to conceptual definitions that are determined by considerations unrelated to the field of reference. Furthermore, the establishment of an omnipresent statistical “*perspective unique*” (single perspective) encourages the adoption of a “*pensée unique*” (single line of thought). This may even be helpful in European policy since it often makes decision making easier. However, it appears to endanger the safeguarding of a variety of perspectives, which is important in the world of science.

- Another problem for science is the general lack of flexibility of official statistics caused by their increasing codification, which is not least of all a consequence of their Europeanization. In many cases, European legislation is required where national legislation would never have been necessary. Think, for example, of the detailed regulations on the calculation of the HCPI (Harmonized Consumer Price Index). Without its functional significance for European policy even the calculation of national accounts would never have been codified. This to a large extent determines the demands on European statistics and considerably limits the possibilities for rapid, pragmatic action in the field of official statistics, with the result that new phenomena of particular interest for science are insufficiently recorded in European statistics and with a certain delay.

- Recently, policy-makers have insisted more on reducing the response burden (which is, on the whole, relatively undemanding) and in this context are pressing for the increased use of administrative sources in order to lighten the “burden” on respondents. This can lead to significant changes (and often also restrictions) in the availability of comparable data, as administrative structures and thus sources often differ enormously within the EU. This in turn can restrict scientific research possibilities. The partial substitution of observation by estimation is particularly problematic for (empirical) science in this regard. However, it must be borne in mind that these estimation procedures are also developed by the (methodological) sciences. The problem is thus not just a conflict between policy making and science, but also a conflict of interests between empiricists and theorists, possibly worsened by policy-makers.

- Policy-makers of course generally support a reduction in the cost of official statistics, especially at the European level. Here too, (methodological) science, in conjunction with technology, offers valuable cost-cutting assistance. But here again there is a conflict of interest between empiricists and theorists. The solid, suitably controlled, accurately targeted, and regular sample survey is still the most popular source for (empirical) science, but these surveys are very costly and are therefore becoming increasingly controversial, a trend reinforced by concerns about data protection. Science must come to terms with the fact that, in official statistics, the importance of the classic sample survey will diminish while that of administrative sources will increase.

- The functional use of official statistics for policy-making purposes has expanded at the European level in recent years. This has raised increasing doubts among scientists and others regarding the credibility of European statistics. It seems to be a widely-held belief (and probably also a basic assumption of the New Political Economy) that official statisticians angle their results, when necessary in the national interest, according to desired political outcomes. In this context, however, science all too often overlooks the harmony of interests between European policy making and science, and the fact that the Europeanization of statistics on the basis of trusting cooperation between the national statistical offices and Eurostat has led to the depoliticization of the statistical processes, from conceptualization to data collection, statistical preparation, and dissemination.

- In general, science seems to have difficulty dealing with the role of policy making in official statistics. As regards statistical methods, the influence of science is of course substantial; scientists are even asked for advice. But as far as the statistical program is concerned, it would be difficult for science to accept the primacy of policy making over statistics. Knowledge in many fields is desirable, but not everything can be researched on account of limited resources (aside from the fact that some things should simply not be officially recorded). Expense and yield, cost and benefit must first be weighed by official statistics within the framework of their legal remit, but ultimately this must always also be the duty of policy-makers as legislators and as the budgetary authority. It is therefore not enough for scientists to voice their concerns and needs to official statisticians; they must also seek support from policy-makers. In the European context, such efforts are two-tiered and therefore doubly expensive, and the world of science does not appear to be particularly well-equipped for this, since it must work at convincing official statisticians and policy-makers at both the national and European level.

- Finally, reference must be made to one more barrier which is particularly problematic in the European context: centralized (European) access to microdata. European legislation generally requires Member States only to provide tables, but not individual data. Microdata at the European level are therefore available for only a very limited number of statistics. These data are of course available to scientists, in accordance with Commission Regulation (EC) No. 831/2002. Access arrangements have admittedly become more user-friendly in recent years, but further improvements in the near future will be difficult to achieve owing to the pending change in the legal basis for European statistics. Instead, we can even expect the process of gaining access to data to become even longer, as a parliamentary inspection has been built into the approval procedure.

3. Possible solutions

For some of the difficulties listed here, there are no simple solutions (e.g., limitations and consequences of harmonization, changes to territorial boundaries) – science will simply have to live with them. It will doubtlessly be possible to find solutions to other problems, but this will take time and above all budgetary resources, and possibly also an amendment to the legal framework. However, these solutions can be found only through dialogue between scientists and official statisticians as well as between scientists and policy-makers.

3.1 Recommendations relating solely to science policy

Scientists without question believe there is room for improvement in the general policy on scientific research at the European level with respect to official statistics. The provision of economic and social statistics is not a particularly important issue for European research policy, unlike German policy, an importance demonstrated at least in recent years by the very existence of the German Data Forum (RatSWD). At the European level, whatever support is allocated is largely directed toward methodological research in the field of statistics. There are certainly good reasons for this, but the result is that Eurostat – the central authority for the provision of European data and the focal point of European statistics, or more precisely for official statistics at European level – is not and cannot be very active in the provision of statistics for (European) policy and the public. There is no body (as yet) comparable to Germany's national and regional Research Data Centers, which specifically address the needs of science. Likewise, there is no infrastructure (as yet) to connect all the relevant data holders and thereby facilitate the use of European data through different channels and different sites. The following recommendations are therefore proposed:

- First recommendation: German research policy (BMBF, *Federal Ministry of Education and Research*) should more actively represent the needs and concerns of scientific users of economic and social data at the European level. If it is appropriate in a national context to give science better access to available data, which has been difficult or impossible to access or use until now, then the same applies to the European context. The German Data Forum (RatSWD) should be called upon to draft recommendations for the further development of a truly European data infrastructure (not only access to data but also data type and volume).
- Second recommendation: in light of the forthcoming amendment to the Commission Regulation (EC) No 831/2002, German research policy (BMBF) and German official statistics should push for simplified access and a greater variety of forms of access. The German Data Forum (RatSWD) could be asked to give an opinion on this in the context of the European amendment procedure.
- Third recommendation: in the summer of 2009, the European Statistical Advisory Committee (ESAC) will take over from the European Advisory Committee on Statistical Information in the Economic and Social Spheres (CEIES). German scientists must lobby the 24 members of this body, some of whom will be representatives from the sciences, for improvement to data access and data volume at the European level (for instance via the

RatSWD). Furthermore, German scientists could urge this body to provide incentives for improved cooperation between official statisticians and scientists (both empirical and methodological scientists).

- Fourth recommendation: scientists should in general make targeted use of the opportunities to voice their views offered under the new "governance structure" of European statistics that has taken shape in recent months. Their efforts will be even more effective if other Member States share these views. It would therefore be a good idea for the RatSWD to establish closer contacts with user bodies in other Member States.

- Fifth recommendation: lastly, it could be helpful for researchers to look into the social and political processes that generate the need for statistical information and tried to analyze these processes. This would certainly also make it easier for scientists to take part in these processes and influence them in such a way as to ensure that greater account is taken of their own concerns. Such processes have, after all, become considerably more complex in recent years and, with the new media, also more participatory, not least at the European level.

3.2 Practical steps

While policy initiatives to improve the legal framework conditions are important, significant improvements are nevertheless also possible under the current conditions.

- Sixth recommendation: German official statistics should engage in technical cooperation with those national statistical offices which also want to improve access for scientists to European data and, as sponsors (where appropriate through the European structures that have been created for that purpose), should take the initiative. Particular consideration should be given here to whether the data made available in the context of this cooperation would go beyond the already Europeanized microdata (on the basis of EU legislation). Data which has not been harmonized owing to a lack of Community competence and which Eurostat cannot take care of are also of interest to empirical science.

- Seventh recommendation: at the same time, German official statisticians should increase their efforts to lobby for improved access to, and an extended scope of, economic and social data at the European level. The European Commission (Eurostat) is of course restricted in the exercise of its right of initiative to those statistical fields that relate to policy areas where the Community is competent. However, when it is a matter of infrastructure that, once created, will be used both for Europeanized and non-Europeanized statistics, it should be possible for the European Commission (Eurostat) to at least assume the role of a catalyst.

Perhaps it will also be necessary to break new ground and separate content, access, and control possibilities from infrastructure. The infrastructure could then be used to provide access to European microdata through Eurostat and at the same time also provide Europe-wide access to national microdata under the joint control of the national statistical offices – in whatever form such a joint structure might take. German national and regional Research Data Centers are probably best placed and suited to submit proposals.

- Eighth recommendation: the use of European statistics presents a number of particular difficulties, some of which have already been mentioned (structural breaks caused by harmonization, contradictions in the double recording of intra-Community flows and entitlements, etc.). Science can make important contributions in how to deal with these difficulties by making them a research subject in their own right. Here again, the German Data Forum (RatSWD) could provide valuable stimulus.

- Ninth recommendation: the German Data Forum (RatSWD) could also be a driving force when it comes to the provision of data on statistical units without a clear national affiliation (e.g., multinational companies). The EuroGroups Register is currently being developed and one objective could be to improve the data on multinationals so that they can be subjected to systematic empirical analysis.

- Tenth recommendation: lastly, it must be pointed out that, not least for its own benefit, science should actively support the statistical policy of the European Commission (Eurostat). Successful harmonization, coordinated and forward-looking program planning, efficient collection and processing procedures, and widespread dissemination of the results generally also improve possibilities scientific research. However, this should apply not only to the core area of European responsibilities and those fields in which the open method of coordination is used, but also for purely national fields. The research avenues open to empirical science depend on the availability not only of temporal but also of spatial data. The European Commission (Eurostat) is of central importance for making the latter type of data available and should therefore be actively and enthusiastically supported by the scientific community.

To sum up, we wish to restate and thereby emphasize the following: in order to improve data for the economic and social sciences, the German Data Forum (RatSWD) should first begin to become a more Europeanized organization. Establishing contacts with partners in the European Union is necessary to allow for their common interests to be asserted jointly, based on the broadest possible coalitions. Secondly, German Research Data Centers at national and regional levels should cooperate with partners in other EU Member States, not least of all to maintain the drive generated by their creation. And, thirdly, representatives of German policy on scientific research (BMBF) should push for European policies to improve the supply and use of economic and social data across Europe.

3. The Role of the German Research Foundation

Eckard Kämper and Manfred Nießen

Contact:

Eckard Kämper
Kennedyallee 40
53175 Bonn
Germany
e-mail: Eckard.Kaemper[at]dfg.de

Abstract

The strategy adopted by the German Research Foundation (DFG) for future data research infrastructures should be based on what has been achieved thus far and the lessons that can be learned from that. First, the focus of effort should be on providing data rather than on sharing data. Second, projects whose primary purpose is to provide a common good should seek to build research infrastructure. The DFG has powerful means at its disposal to fund outstanding infrastructure projects. It is up to the scientific community to adapt and utilize these funding instruments. Different types of strategic cooperation are required among interested parties in the field. These include: cooperation on identifying thematic priorities within the research community; cooperation between the research community and funding institutions in determining funding options; cooperation around defining the division of labor between different funding institutions (including ministries) on the national and international level. The DFG is prepared to play an active role in this cooperative effort under the leadership of its elected bodies (the *Fachkollegien* and *Senat*).

Keywords: large scale studies; strategic cooperation

The research infrastructure of the social sciences, like that of other disciplines, has long had a place on the agenda of the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*), both in terms of its funding policies and its funding activities. The DFG provided the funding for both ZUMA[1] and SOEP[2], for example, and nurtured them through their formative years. The DFG has also funded activities at the ZA[3], IZ[4] and ZPID[5]. All these activities have been about data – about methods and methodologies for collecting and analyzing data, and about organizations and structures for preserving data and making them accessible.

Data-related research infrastructure has become a more prominent topic in research policy over recent months and years, nationally as well as internationally. In the general science policy debate, much emphasis has been placed on “sharing data,” often also referring to open access initiatives.

1 ZUMA: Center for Survey Design and Methodology & Social Monitoring and Social Change. See: http://www.gesis.org/en/institute/gesis-scientific-sections/center-for-survey-design-and-methodology/ and http://www.gesis.org/en/institute/gesis-scientific-sections/social-monitoring-social-change/.

2 SOEP: German Socio-Economic Panel. See: www.diw.de/gsoep.

3 ZA: Data Archive and Data Analysis. See: http://www.gesis.org/en/institute/gesis-scientific-sections/data-archive-data-analysis/.

4 IZ: Specialized Information for the Social Sciences & Information Processes in the Social Sciences. See: http://www.gesis.org/en/institute/gesis-scientific-sections/specialized-information-for-the-social-sciences/ and http://www.gesis.org/en/institute/gesis-scientific-sections/information-processes-in-the-social-sciences/.

5 ZPID is the psychology information center for the German-speaking countries. See: http://www.zpid.de/index.php?lang=EN.

1. Sharing data: a realistic approach

The idea of sharing data focuses on data produced in research projects that pursue specific hypotheses and generate the data accordingly. That is, the data are generated or collected to answer the specific questions of the project at hand; thus, the data are project-specific.

It is taken for granted that sharing data will increase efficiency and reduce research costs by necessitating replication studies and reducing duplication in data production. However, data sharing is by no means a new idea. It has a long history that is well worth examining more closely.

The DFG has long required that all funded projects transfer their data to public data archives, for example, to the ZA or ZPID. But relatively few datasets have actually been transferred. As a result, some of the DFG's national programs (SPP, *Schwerpunktprogramme*) have imposed strict time limits on the transfer of data to public archives for every project funded. While the success rate – the number of projects complying with this provision – has increased, it still is far below 100 percent.

We may lament the discrepancy between official policy and the actual behavior of the research community, exert more pressure, and impose tighter controls. But we should also ask: what are the reasons for this discrepancy? Why do relatively few projects "share" their data by transferring them to a data archive?

Project-specific data, generated to answer specific research questions, do not necessarily lend themselves to use by others. Both contextualization and specification are a necessary provision for sharing these kinds of data. After completion of the research project, scarce resources – researchers' time in particular – must be further invested to produce a dataset that is potentially valuable to others and that can be transferred to an archive for their use. The question is: can the reluctance of the research community to invest in this type of data sharing be understood as an indicator of the low value ascribed to shared data?

And what about the datasets that *have* been transferred to archives – data from projects whose primary aim was not to produce data "for others" but to pursue specific research questions? To what degree are these data being used by the research community? In other words: is there sufficient demand?

Both of these questions – why the research community is reluctant to invest in sharing data and how high is the actual demand for shared data – need to be analyzed in more detail. Data generated with public money should, of course, be made available to the public (that is, in the case of sensitive individual or company data subject to data protection restrictions, made available to the research community). However, keeping in mind the overall goal of a data infrastructure, for some projects it may not be a top priority to invest in data sharing, given the high transaction costs and limited value of the data to the scientific community. More pragmatic approaches to secure access to individual project data are being discussed more in the context of “research integrity” than in the context of infrastructure.

The goal of providing data is a markedly different approach from data sharing, and it has become increasingly prominent in the DFG’s funding activities over the past few years. With the term “data provision,” we refer to a type of project or program whose primary aim is not to answer a specific, narrowly delimited research question and to collect data for this purpose, but to collect and/or generate data for wider use and thus act as a “research infrastructure.” The focus and theoretical foundation of this form of data production is not a set of specific hypotheses, but a wider research topic or area. Data production for wider use is the main purpose of the DFG’s projects and programs, which are designed as a service to the scientific community. Increasingly, data production is taking the form of large-scale longitudinal studies.

The DFG has long been regarded as lacking adequate funding instruments for longitudinal studies. In 1995, however, the DFG began considering how to remedy this problem, and held a workshop convening experts from the field of large-scale longitudinal studies and members of the DFG’s committees. The workshop resulted in a paper that specified the criteria that would need to be fulfilled in order for longitudinal studies to seek DFG funding, and that encouraged researchers to develop their ideas for such studies.[6] While this did not produce any significant immediate effect, the situation has changed dramatically in recent years. Large-scale longitudinal studies providing research infrastructure for the social sciences have become a major activity. Various factors have contributed to this change:

(1) Emerging activities in the national research community, closely linked to similar activities in Europe and elsewhere;

(2) Increased attention to these developments in European programs and European institutions;

(3) Adjustments of DFG instruments to foster and promote these activities.

6 The paper was widely published: Kölner Zeitschrift für Soziologie, Psychologische Rundschau, Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, Zeitschrift für Wirtschafts- und Sozialpsychologie, and ZUMA-Nachrichten.

2. Providing data: shaping the instruments

What did the DFG do to bring about this reorientation and why? It all goes back to the workshop of 1995, where the first strategic debate took place on how the DFG could improve opportunities for funding longitudinal studies. The workshop brought together representatives from all disciplines of the social and behavioral sciences. Its recommendations addressed the scientific community as well as the DFG as a funding organization.

This initial input did not produce systematic changes, however, either within the scientific community or at the DFG. This changed, however, with a major strategic initiative launched by the DFG in 2002, called the Funding Initiative for the Humanities (*Förderinitiative Geisteswissenschaften*). This initiative addressed the specific needs of the humanities, but also created new funding opportunities open to both the humanities and the social sciences. One of the four pillars of the strategic initiative was to reshape and modernize the DFG's strategic initiative long-term program or *Langfristprogramm*, whose effects became visible as early as 2003.

The *Langfristprogramm* had been in existence since the DFG was founded, but was initially designed only for the humanities. In 2003, the DFG's *Senat* and Joint Committee resolved to implement a reform of this program with the following elements:

(1) Limits were placed on the formerly open-ended time frame: the program is now only for research activities requiring seven to twelve years of funding.

(2) Only projects of potentially high scientific impact and importance will be funded. A longer-term perspective is necessary, but is by no means the sole requirement.

(3) The *Langfristprogramm* is no longer confined to the humanities, but is now open to both the humanities and the social sciences. The strategic decision to open up the *Langfristprogramm* to the social and behavioral sciences was based, among other things, on the recommendations from 1995. Longitudinal studies are invited to seek funding within the *Langfristprogramm*.

(4) As a consequence of provisions (2) and (3) (aiming at high-impact activities and opening up to the social sciences in general and longitudinal studies in particular), the scale of funding per individual project has been expanded: substantial funding is available depending on the individual project needs. As a consequence, fewer projects will be funded, but they will come from a broader range of disciplines – humanities and social sciences – and with a broader range of budgets.

The first project in which this new funding option was put to use was the European Social Survey (ESS), an internationally comparative study of repeated cross-sections, with more than twenty countries participating. The European Commission provides the core funding for this project, and more than twenty national funding agencies finance the national data collection. The *Langfristprogramm* was essential in making the German part of the European Social Survey possible, and allowed the DFG to fully participate in the European program. When the DFG makes a decision to approve a project as part of the *Langfristprogramm*, this includes a commitment to provide funding for the entire duration of the activity. Because the ESS was part of the *Langfristprogramm*, the DFG was able to stand in for the ESS in the network of national support institutions, the European Commission, and the European Science Foundation, and to formally sign commitments. This provided the groundwork for the ESS to become a truly European infrastructure that eventually became part of the road map of the European Strategy Forum on Research Infrastructure (ESFRI). As a consequence, the ESS may become a "European Research Infrastructure," which will require a new legal form. The aim is to become a kind of international organization. This will certainly have implications for the role of national funding organizations that are still unknown to us.

Just recently, in December 2008, the German Longitudinal Election Study (GLES) was adopted as part of the *Langfristprogramm*, with the potential to be funded for nine years. After that time, and after having gathered data on three successive national elections, it is intended that the GLES will be taken on board at GESIS. Whereas the future perspective for the ESS beyond its funding as part of the *Langfristprogramm* remains open, the future of the GLES is relatively secure: provided that the DFG-funded project proves to be a success in scientific terms, it will be continued under the institutional umbrella of GESIS.

The situation of pairfam, the panel study of intimate relations and family members, is unique in another respect as well: a national research program (SPP) was set up by DFG to develop and implement the study. Normally, national programs aim at rather loose cooperation between projects around a common topic. With pairfam, however, the very idea of the program was to develop a common product. This required a clearly defined division of labor between the individual projects within the program, a high level of coordination, clear leadership, and intense collaboration across the fields of sociological, economic, and psychological research on family and relationships. Although the funding instrument that was used, the SPP, normally aims at supporting a different kind of scientific cooperation, the adaptive use of this instrument was successful, and indeed innovative: the first four years of the SPP were used for the development of the panel study, and the final two years are currently being devoted to carrying out the first two waves of the panel.

Before giving a green light to these final two years and releasing the actual funding for the first two waves, the *Senat* and Joint Committee of DFG carefully considered the future prospects for pairfam. After all, it would not have made sense to finance the first two waves without a perspective on future steps. The deliberation was based on a review panel's assessment of pairfam's plans. Reviewers, the *Senat,* and the Joint Committee came to the conclusion that pairfam should be invited to seek future funding as part of the *Langfristprogramm*. This opened up a perspective of twelve years for pairfam and confirmed the strategic decision to design the *Langfristprogramm* in a way that would allow for substantial funding of individual projects. pairfam, which started as part of a SPP, demonstrates that the new *Langfristprogramm* is not the only instrument in the DFG's portfolio that can be used to support large-scale longitudinal studies.

In principle, all funding instruments should be considered. The German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) is a prominent example. SOEP, which has become a cornerstone of the German research infrastructure in the social and behavioral sciences, was initiated and developed many years ago as part of a collaborative research center (SFB). When this SFB ended (in 1991) after an initial twelve-year funding period, funding for the SOEP was continued under the individual-project funding mode (refinanced by special funds from the German federal and state governments). However, given the importance of the SOEP as a research infrastructure, an institutional solution was needed that could provide long-term stability. A solution was negotiated by the Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*), the respective ministries of the states, and the DFG: after more than twelve years in the individual-project funding mode, the SOEP was established as a special "service unit" at DIW Berlin, a member of the Leibniz Association (WGL, *Leibniz Gemeinschaft*). SOEP's success story – with regard to funding and institutional solutions – is rooted in the adaptive use of several funding instruments and cooperation among the funding institutions (BMBF, DFG, WGL). pairfam, on the other hand, is currently in the process of adapting several funding instruments to its needs.

A final example of both adaptation of funding instruments and cooperation among funding institutions is the National Educational Panel Study (NEPS). The idea for NEPS was first presented and discussed at the symposium in 2004 that was organized by the DFG as part of its "Program on Empirical Research on Education." The symposium brought together researchers from Germany and other European countries as well as representatives of government ministries. At its conclusion, the program's Scientific Board gave advice that formed the basis for the DFG's position on

NEPS. Following the Scientific Board's recommendation, the DFG's Governing Board agreed that the DFG would play an active role in the future process, in close collaboration with BMBF, whereby the funding for NEPS would come solely from the BMBF.

The DFG organized preparatory expert meetings, an international expert workshop to assess the pilot study, and a full-scale international peer review for the full proposal. Based on this peer-review, the BMBF made the formal decision to finance NEPS as a data-providing research infrastructure. The DFG's *Senat* simultaneously decided to allocate a substantial budget for a national research program (SPP) in which projects would be funded that make scientific use of the NEPS data. In other words, the DFG, by implementing its mechanisms for independent assessment of scientific quality, provided the mechanism to firmly root NEPS in the scientific community. For the implementation of peer-review results, the BMBF and DFG agreed on a division of labor: the BMBF finances the research infrastructure, the DFG funds the scientific use of the data through its national research program.

3. The role of the German Research Foundation (DFG)

The major large-scale longitudinal studies that currently serve as the foundation for the data research infrastructure in the social sciences have developed into a major field of activity at the DFG. This development, however, was not the result of a strategic master plan. Of course, there was the policy statement of 1995 and the strategy decision of 2003 to redesign the *Langfristprogramm* specifically geared towards longitudinal studies in the social sciences. Nonetheless, the individual activities and projects that emerged within the scientific community were pursued in a relatively uncoordinated way. This is not surprising, given that the DFG is owned by the scientific community and firmly founded on the principle that strategic initiatives as well as individual funding decisions must be driven by research questions and by researchers themselves. The case of NEPS does not follow this principle to the letter, but nevertheless provides a good illustration of the DFG's role: NEPS was initiated and, at least in its early stage, conceptualized by the BMBF. Furthermore, it is the BMBF, not the DFG, that funds this research infrastructure. Close cooperation and partnership with the DFG was sought to provide scientific quality control through independent peer review and thereby scientific legitimation. The DFG's role in the partnership with BMBF has been to ensure that this externally initiated panel study is and will continue to be essentially science-driven.

Closely related to the principle of being “science-driven” is the fact that the DFG cannot provide institutional funding, but is confined to project funding. The major strategy decision to redesign the *Langfristprogramm*, therefore, meant redefining it as a tool for project funding and introducing the twelve-year limit for each cycle of funding. This means that longitudinal studies can be funded by the DFG under one of the following provisions: (1) the study will come to an end within twelve years; (2) the topic of the study demands a longer perspective than twelve years, but if no continuation can be secured, the scientific outcome of twelve years alone will justify the investment – in other words, the second-best solution can stand alone; or, (3) if the study is planned from the outset as a truly longitudinal one, going beyond twelve years, initial funding by the DFG can be granted if follow-up funding (i.e., institutional funding) can be expected. SOEP (which was not planned as such a long-term project, but rather became one) and GLES (which was planned as such from the very beginning) are examples of the DFG strategy of enabling a potentially long-lasting project to be launched. This brings us to our first conclusion regarding the role of the DFG: projects that seek funding from the DFG have to be driven by the scientific community; that is, they must be well-planned scientifically and they must be organized in a form suited to project funding – at least for the duration of DFG funding period. If these two provisions are fulfilled, the DFG is well-equipped to find adaptive solutions.

Projects like ESS, GLES, and pairfam are data research infrastructures of central importance to the research community; yet, they are expensive and put considerable strain on the budgets available for funding the social sciences. Up to now, these projects have been proposed individually and dealt with on a case-by-case basis. However, if data research infrastructures are going to establish themselves as a major line of activity and funding, some degree of coordination and even strategy might be necessary. The DFG’s elected bodies – the *Fachkollegien* and *Senat* – will be able to provide leadership for this process of addressing key questions within the scientific community: mapping the field, defining thematic priorities, co-ordinating projects and programs in order to maximize effects and economize resources, etc.

Coordination and strategy also pose challenges to the DFG as a funding organization, to the ministries, and to research organizations like the Leibniz Association (WGL) universities. Coordination and collaboration between the institutions have up to now also taken place on a case-by-case basis: SOEP (DFG/BMBF/WGL), GLES (DFG/WGL) and NEPS (DFG/BMBF/University) have each resulted in individual constellations and solutions that we regard as success stories.

Yet, it must also be reiterated, we have witnessed increased activities in this field and the momentum has been building. Not only because of the financial implications, but also in view of the long-term perspective of each individual activity, coordination and collaboration between the major players in the field may become necessary. A division of labor and development of institutional perspectives are the keywords here. The DFG is prepared to play an active role in this coordination process.

Coordination and collaboration between institutions is not only appropriate in view of the division of labor and sharing of responsibilities on the national level, but also in view of the international activities. ESFRI is but one field, however important it may be. If "European Research Infrastructures" come into existence as new legal entities, we as national institutions will have to redefine our position vis-à-vis these new entities as well as in relation to each other. The national institutions will have to cooperate in order to maximize the effects on the European and international level – and of course, in the best interests of the research community.

The International Data Forum (IDF), as a final example, goes beyond the European level. The DFG has supported the initial phase of this idea, together with our partner organizations from the UK, the US, Canada, the Netherlands, and China. The goal of the IDF is to facilitate and coordinate international production and sharing of data for research in the social sciences. It strives to align its aims with the strategic directions and priorities of prominent organizations representing the producers, managers, and research users of data relevant to the social sciences. One of its tasks is to facilitate collaboration and mutual understanding between key data stakeholders in the social sciences. Following the founding conference for the International Data Forum, the next steps are set to establish interagency agreement on the need for IDF and the scale of its operations. Decisions will be sought in 2009.

DFG has nominated the chair of the RatSWD as a member and the German representative of the Founding Committee of the IDF. This is already a concrete example of coordination between national institutions.

4. Summary

The DFG strategy for data research infrastructures will be based on what has already been achieved and on the lessons that can be learned:

(1) The guiding perspective should be on providing data, rather than sharing data. Projects whose primary purpose is to provide a common good should focus on building research infrastructure.

(2) The DFG has powerful programs at its disposal to fund outstanding infrastructure projects. It is up to the scientific community to adapt and utilize the diverse funding instruments of the DFG to its needs.

(3) Strategic cooperation is needed among all interested parties: cooperation within the research community on identifying thematic priorities; co-operation between the research community and funding institutions on the options for funding; and cooperation between the funding institutions on the division of labor, on the national as well as on the international level.

(4) The DFG is prepared to play an active role in this cooperative effort under the leadership of its elected bodies (*Fachkollegien* and *Senat*).

4. Providing Data on the European Level

Peter Elias

Contact:

Peter Elias
University of Warwick
Institute for Employment Research
Coventry CV4 7AL
UK
e-mail: Peter.Elias[at]warwick.ac.uk

Abstract

This paper reviews the potential demand for and the provision of European data for social scientific research. The concept of data provision is defined broadly, covering the ease with which specific types of data can be discovered, interpreted, readily understood and accessed by researchers.

The paper is structured as follows. First, it addresses the issue of why researchers need European (as opposed to national) data resources. This leads in to a short section discussing the potential demand for data at the European level. The main section focuses on the nature of various data resources currently available or under development. Finally, it concludes with an assessment of the need for new and/or improved data infrastructures and suggests where efforts could be focused in order to respond to such needs.

Four areas are identified where there is a clear need for new European research data resources to be developed. These are:

- a European Household Panel
- facilities to encourage comparative analysis of birth and other age cohort studies
- a European organization-based longitudinal survey
- improved access to microdata records held by Eurostat

Keywords: European data infrastructures, social science data needs

1. Introduction

This paper reviews the potential demand for and the provision of European data for social scientific research. The concept of *data provision* is defined broadly, covering the ease with which specific types of data can be discovered, interpreted, readily understood and accessed by researchers.

The paper is structured in the following way. The next section addresses the issue of why researchers need European (as opposed to national data resources). This leads in to a short section discussing the potential demand for data at the European level. The main section focuses on the nature of various data resources. Finally, the paper concludes with an assessment of the need for new and/or improved data infrastructures and suggests where efforts could be focused to realize such needs.

2. Why do we need data at the European level?

There are two main reasons for supporting the development of Europe-wide data infrastructures. The first relates to the need to inform social and economic policies which are pan-European in design or operation. As the European Union continues to integrate its economic and social structures, there is a need to understand how such integration operates across the EU, and to identify both strengths and weaknesses in policy implementation. It is primarily for this reason that the European Union, through its statistical agency (Eurostat), coordinates the production and collection of census, survey, and administrative data across the EU. The second need for European data relates more to the nature of research in the social sciences which, for the most part, cannot make use of the kind of randomized and controlled experiments that typify research in the physical sciences and must rely more on variations across groups and through time to investigate causality. Europe provides wide cultural diversity not simply in the obvious dimensions (language, politics, legal systems, etc.) but also across more difficult to measure traits such as cultural values, traditions, beliefs. To the researcher this provides variations that help inform the research process. “Europe” thus affords the research environment that the physical scientists would otherwise harness in the laboratory.

3. What kinds of data do we need for research in the social sciences at the European level?

European-level research has the same basic needs for data as research at the national level. However, the very nature of the European Union dictates that there will be specific research interests that may not have any national counterpart. For example, research on cross-national migration within the EU or across its external borders. Equally, understanding economic growth and decline within a European context (e.g., transnational investment, impact studies for the location of large-scale infrastructures, economic stability within the eurozone) requires a specific Europe-wide focus whilst drawing upon what are essentially national data resources.

4. Pan-European data resources

This section illustrates the available data resources designed to facilitate European research. No distinction is made here between data resources which are purpose-built for comparative research at the European level (input harmonized) and those which have arisen as research groups have attempted to meld a number of separate resources into a pan-European resource (output harmonized).

To document the variety of data resources that are available, the following typology has been adopted:

- *cross-sectional micro resources* – information which is descriptive of a unit of observation at a single point in time. Cross-sectional microdata observations may be repeated in order to monitor change at the macro-level;
- *longitudinal microdata resources* – information which describes the evolution of a unit of observation (e.g., a person, a family, an organization) through time. Such data resources are powerful instruments in the study of cause and effect;
- *macro databanks* – derived from cross-sectional survey or administrative data sources, "databanks" are repositories of tabulated data, usually providing a wide range of social and economic indicators.

Macro databanks are not covered in detail in this paper. While they constitute important resources for a variety of research interests, access to these resources and their use is relatively easy and uncontroversial.[1] However, for most research purposes, researchers want access to the underlying microdata resources from which the statistical indicators in macro databanks are constructed.

Other typologies are also useful, for example the distinction between administrative and transaction data – the former referring to data generated as a by-product of an administrator process (registration for social security benefits) or the latter from a transaction (e.g., purchase or sale of goods or services). Reference to such data types is made in the concluding section.

1 As an example of a research resource dedicated to providing access to and information about a wide variety of macro databanks, see ESDS International: http://www.esds.ac.uk/international/access/dataset_overview.asp. [Last visited: 03/02/2010].

4.1 Cross-sectional microdata resources

4.1.1 Resources available via Eurostat

Figure 1: Cross-sectional microdata resources available through Eurostat

EU Labour Force Survey	EU Structure of Earnings Survey
The EU-LFS is the longest running EU-wide statistical survey. Conducted by National Statistical Offices in Member States, the LFS has, since 1992, had a common output requirement in terms of the employment-related information it provides on individuals and households. Data currently available covers the period 1983–2006. In Spring 2002 the total sample size was approximately 1.5m persons. Data are available as anonymized micro records.[2]	The EU-SES is a large enterprise-based sample survey designed to provide accurate and harmonized data on earnings across the EU. The survey was held in 1995, 1999, 2002, and 2006. Results for 1995 are not comparable with later years. Data collected includes earnings, age, gender, occupation, sector, hours worked, education, and training for employees of enterprise with 10+ employees. The latest data available for research purposes is the 2002 survey. Access to SES data is through the SAFE Centre in Luxembourg.[3]
EU Community Innovation Survey Community Innovation Statistics are produced in all 27 EU countries, 3 EFTA countries, and candidate countries. Data are collected on a four-year cycle. The first (pilot) survey was held in 1993, the second survey held in 1997/98 and the third survey in 2000/01. The fourth survey, conducted in 2006 with a reference year of 2004 will be available shortly. Anonymized microdata are available via CD-ROM. Access to non-anonymized data is possible through the SAFE Centre facility in Luxembourg.[4] The CIS provides information on the characteristics of innovation at the enterprise level.	**EU Statistics on Income and Living Conditions** The EU-SILC was designed as a successor to the European Community Household Panel which ran from 1994 to 2001. The first release of EU-SILC was in 2004, with a 2003 reference year. Anonymized microdata from 2004 and 2005 are available via CD-ROM. The EU-SILC contains a longitudinal element covering a four year period. The first longitudinal database was made available late in 2007.[5]

2 For further information on access conditions, see EU-LFS: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home. [Last visited: 03/02/2010].
3 See EU-SES: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home.
4 See EU-CIS: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home.
5 For details see EU-SILC: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home.

Since the late 1960s, the European Union (formerly the European Community) has sought to develop comparable microdata resources in order to measure and progress social, political, and economic integration. These efforts have given rise to a number of major data resources. However, access to these resources has, until recently, been severely restricted.

Cross-sectional microdata collected by Eurostat from National Statistical Offices across the EU include:

- EU Labour Force Survey (EU-LFS)
- Community Innovation Survey (EU-CIS)
- Structure of Earnings Survey (EU-SES)
- Statistics on Income and Living Conditions (EU-SILC)

Brief details of each of these sources are shown in the boxes below. Further information can be gained by following the hyperlinks.

4.1.2 Resources available via other data providers

4.1.2.1 Luxembourg Income Study (LIS)

The LIS began in 1983 under the joint sponsorship of the government of Luxembourg and the Centre for Population, Poverty and Policy Studies (CEPS), which became an independent body in 2001. The LIS archive contains two databases, the Luxembourg Income Study database and the Luxembourg Wealth Study (LWS), covering cross-national micro datasets on incomes, wealth, employment, and demography. The LIS database contains nearly 200 datasets organized in six time periods (waves) spanning the years from 1968 to 2005.[6]

With the exceptions of Portugal and Romania for Wave VI (around 2004) and Slovenia for Wave V (around 2000), income microdata are available for all EU countries, North America, Australia, Israel and Taiwan. The newer LWS database (released in December 2007) contains 13 wealth datasets from 10 countries.[7]

No direct access to the micro datasets is permitted. Registered users submit syntax (SAS, SPSS, and STATA) which LIS staff run on their behalf. Planned developments for the period 2008–2013 include a web-based user

6 Microdata held by Eurostat are confidential data about individual statistical units. The release of these data to bona fide researchers is governed by Commission Regulations EU Nos. 83/2002, 1104/2006 and 1000/2007, which implement Council Regulation (EU) No. 322/97. Article 17 allows the EU to grant access to data it has collected from national statistical authorities if the national statistical authority gives explicit permission for such use.

7 Austria, Canada, Cyprus, Finland, Germany, Italy, Norway, Sweden, UK, and the US.

interface for syntax submission, storage of and access to prior programs, and an online tabulator.[8]

4.1.2.2 Council of European Social Science Data Archives (CESSDA)

CESSDA is a network which promotes the acquisition, archiving, and distribution of electronic data. The network now extends to 20+ countries across Europe, providing access to and delivering over 50,000 data collections per annum and acquiring over 1,000 data collections each year. The CESSDA portal provides easy access to the catalogues of member organizations.

Via its multilingual search interface, CESSDA guides enquirers to appropriate datasets at specific data archives.[9] Enquirers can browse datasets by topic and by keywords before linking to specific archive websites to determine access conditions.

In 2007, CESSDA acquired FP7 Preparatory Phase funding to facilitate a significant upgrade in its functionality. This three-year phase will result in a plan to facilitate and coordinate national funding to provide a European research infrastructure. CESSDA also provides access gateways to other important EU-wide data resources, including the European Social Survey,[10] the Eurobarometers,[11] the International Social Survey Programme and the European Values Study (see below for further details about these sources).[12]

4.1.2.3 Integrated Public Use Microdata Series-International (IPUMSI)

IPUMSI is a project funded by the US National Science Foundation, based at the University of Minnesota, dedicated to the collection and distribution of census data from around the world.

To date, 35 countries have donated microdata from 111 censuses, totaling 263 million person records. The eight European countries which have so far contributed to the IPUMSI database are Austria, France, Greece, Hungary, Netherlands, Romania, Spain, and the United Kingdom. Census data for Slovenia will be available in 2009. Plans are also underway for the addition of censuses from Czech Republic, Germany, Ireland, Italy, Switzerland, and Turkey. The IPUMSI website maintains good metadata documentation stan-

8 For further information about LIS see: http://www.lisproject.org/.

9 CESSDA facilitates keyword searches across the following data publishers:
UK Data Archive, SSD (Sweden), SIDOS (Switzerland), NSD (Norway), GSDB (Greece), GESIS-ZA (Germany), FSD (Finland), DDA (Denmark), DANS (Netherlands), ADPSS-Sociodata (Italy), ADP (Slovenia).

10 http://www.europeansocialsurvey.org/.

11 http://www.esds.ac.uk/international/access/I33089.asp.

12 For further information about CESSDA, see: http://www.cessda.org/index.html. [Last visited: 03/02/2010].

dards that allow users to appreciate differences in the ways in which censuses have been carried out, differences in the definition of key variables, etc.[13]

4.1.2.4 European Social Survey (ESS)

The ESS is an academically directed social survey designed to provide information on the attitudes, beliefs, and behaviors of Europe's changing population. Now in its fourth round, the ESS maps long-term attitudinal and behavioral changes in European society. Over 30 European countries now participate in the survey, with sample sizes ranging from 1,000 to 2,000 persons in each country.[14]

A major strength of the ESS is its attention to methodological weaknesses in the generation and its use of cross-national comparative data. Particular emphasis is placed on the interpretation of key concepts in the survey research instruments and their translation into different linguistic and cultural contexts.

4.1.2.5 Eurobarometer

The Eurobarometer surveys were established in 1973, designed to provide the European Commission with data on social trends, values, and public opinion generally, helping in the preparation of EU-wide policy and to inform the evaluation of its work. Surveys are conducted annually, with each survey covering approximately 1,000 face-to-face interviews[15] in each EU country.

Eurobarometer microdata are available from a variety of sources, including the Inter University Consortium for Political and Social Research (ICPSR) at the University of Michigan and the GESIS[16] Data Archive.[17]

4.1.2.6 International Social Survey Programme (ISSP)

Since 1983, the ISSP has promoted cross-national collaboration in the creation of research instruments and methods to generate a wide variety of data about social, economic, and political change, as well as values, beliefs, and motivations. While individual country samples are fairly small, the ISSP

13 Census data are freely available to registered users at: https://international.ipums.org/international/ [Last visited: 03/02/2010].

14 The minimum number of achieved interviews is set at 2,000 persons, except in countries with a population of less than 2 million, where the minimum number is 1,000.

15 Variations are Germany (2,000), Luxembourg (600), UK (1,300 of which 300 in Northern Ireland).

16 Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*).

17 Links to these sources can be made through CESSDA (see above).

devotes considerable resources to ensuring good comparability between countries.[18]

4.1.2.7 European Values Study

The European Values Studies (and its companion, the World Values Surveys) are designed to enable a cross-national, cross-cultural comparison of values and norms on a wide variety of topics and to monitor changes in values and attitudes across the globe. Topics covered include perception of life, family, work, traditional values, personal finances, religion and morale, the economy, politics and society, the environment, allocation of resources, contemporary social issues, national identity, and technology and its impact on society. To date, four waves have been conducted in 1981–1984, 1990–1993, 1995–1997, and 1999–2004. Not all of the earlier surveys employed probability sampling procedures. These survey responses have been integrated into one dataset, to facilitate time series analysis.[19]

4.1.2.8 European Working Conditions Survey (EWCS)

The EWCS series began in 1990–91 and is usually conducted every five years. The survey utilizes a face-to-face questionnaire administered to a random sample of employed people (employees and self-employed), who serve as representatives of the working population in each EU country. The latest survey, held in 2005, covered the EU27 plus Croatia, Turkey, Switzerland, and Norway.

The questionnaire covers many aspects of working conditions, including violence, harassment and intimidation at the workplace, management and communication, work-life balance, and payment systems.

The EWCS datasets for 1991, 1995, 2000, and 2005 are available from the UK Data Archive (ESDS, Economic and Social Data Service). For further information, see EWCS or EWCS at ESDS.

18 Further information about the ISSP is available at: http://www.issp.org/ [Last visited: 03/02/2010].

19 Further information about the EVS can be found at: http://www.europeanvaluesstudy.eu/ [Last visited: 03/02/2010].

4.2 Longitudinal microdata resources

4.2.1 European Community Household Panel (ECHP)

The European Community Household Panel (ECHP) is a panel survey in which samples of households and persons have been interviewed year after year. These interviews cover a wide range of topics concerning living conditions. They include detailed income information, financial situation in a wider sense, working life, housing situation, social relations, health, and biographical information of the interviewed. The total duration of the ECHP was 8 years, running from 1994–2001 (8 waves).[20]

4.2.2 Survey of Health, Ageing and Retirement in Europe (SHARE)

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel database of microdata on health, socio-economic status, and social and family networks of more than 30,000 individuals aged 50 or over. Eleven countries have contributed data to the 2004 SHARE baseline study, ranging from Scandinavia (Denmark and Sweden) through Central Europe (Austria, France, Germany, Switzerland, Belgium, and the Netherlands) to the Mediterranean (Spain, Italy and Greece). Further data have been collected in 2005–06 in Israel. Two "new" EU Member States – the Czech Republic and Poland – as well as Ireland, joined SHARE in 2006 and participated in the second wave of data collection in 2006–07. The survey's third wave of data collection will collect detailed retrospective life-histories in sixteen countries in 2008–09, with Slovenia joining in as a new member.[21]

5. Summary: future needs for a European data infrastructure

Table 1 attempts to briefly summarize this review of available *European* data resources which are likely to be of interest to social scientists. The list covers microdata resources only. Macro databanks, providing indicators of trends and yielding information on country and regional differences across Europe, are useful research resources, but do not provide the flexibility needed for exploring social, economic, and demographic processes in depth, nor are they adequate for most scientific modeling purposes. The table also excludes CESSDA, which (amongst other functions currently under development) acts

20 For further information: http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home
21 For further information: http://www.share-project.org/ [Last visited: 03/02/2010].

primarily as a networked intermediary organization. The facility it offers – to search data catalogs in different ways across a range of archives in various countries for specific sources of data – makes CESSDA a powerful tool for data discovery and for comparative research where data permits. CESSDA also provides links to many of the resources shown in table 1, but it is not, in itself, a producer of pan-European data for research purposes.

The issues that are raised about sample sizes, data accessibility, and/or data quality paint a none-too-inspiring picture of the range and availability of European data resources for research across the social sciences and in related disciplines. Despite the efforts made by individuals, research teams, and by some national bodies, the availability, accessibility, and quality of these data resources are fairly limited. There are a number of notable exceptions here, particularly ESS and SHARE, both of which, like CESSDA, have been recognized by the European Strategy Forum for Research Infrastructures (ESFRI) and the European Commission as major research infrastructures in need of further support and development. However, in a number of EU countries and North America, major advances are being made to facilitate a broader social science research agenda which encompasses research in the fields of environmental sciences (climate change, air soil and water pollution, and crop modification), medical sciences (genetic expression and human behavior, spread of contagious diseases, impact of ageing), and engineering (transport systems and congestion, housing design, personal and collective security). This broader agenda has required new types of data structures that are significantly larger than any of the resources currently available, are longitudinal in nature, and which can be readily enhanced via linkage to administrative and/or transactional data. Simultaneously, new access procedures have been developed which take advantage of technical developments to provide better and more secure access to complex and sensitive data sources, as well as facilitating a more "hands-on" approach to research[22] than has been the case with, say, the Luxembourg Income Study or the Eurostat SAFE access procedures.

Possibly the most disappointing aspect of this review relates to the continued barriers to widespread access by the research community to the purpose-built European statistical databases held by Eurostat. Notwithstanding renewed legislative efforts to improve matters from within Eurostat, access remains slow, costly, and restrictive. No remote access is provided by Eurostat, despite the proven technology, the security this approach offers compared with the proliferation of data via physical media, the reduced

22 http://www.norc.org/projects/data+enclave+project.htm [Last visited: 03/02/2010], and http://www.ons.gov.uk/about/who-we-are/our-services/vml/index.html. [Last visited: 03/02/2010].

Table 1: European microdata requirements: sources and issues relating to access, coverage and quality

Type of microdata required	Source	Data access, coverage, quality issues
Census data (demography and housing)	IPUMSI	Incomplete coverage across the EU.
Labor force, income data, living conditions	Eurostat (EU-LFS; EU-SILC; EU-SES)	High costs of access, complex bureaucracy, some data to be accessed on site, no record linkage possible.
Values, beliefs, attitudes	Eurobarometer; ISSP; European Values Study	Relatively small sample sizes.
Social and political behavior	European Social Survey; ISSP	High quality, but relatively small samples.
Longitudinal individual and household data	Eurostat (ECHP); SHARE	ECHP discontinued after 2001. SHARE has high-quality longitudinal data but is still fairly new.
Organization-based data (structure, pay, working conditions)	Eurostat (EU-CIS); European Working Conditions Survey (EWCS)	Only available at 5-year intervals, no longitudinal measures.

costs, and the convenience it provides to the research community. The costs currently incurred by researchers working on publicly funded research are hardly defensible.[23]

This suggests where efforts should be focused to improve these essential research resources. Four major new initiatives are proposed:

5.1 A new European household panel

This should build upon the latest developments in a number of countries, to establish larger and better household panels than has hitherto been the case. The obvious first step here is to determine how certain countries can take the household panels they currently have under academic direction, and align their activities to facilitate cross-panel analysis. There is nothing new in this approach. Indeed, the demand for cross-national equivalent files based upon the US Panel Study of Income Dynamics, the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*), and the British Household Panel Survey (BHPS) testifies to the values of such resources. However, the new UK Household Longitudinal Study Understanding Society, the SOEP, and the Swiss Household Panel (SHP) are candidates for renewed efforts to build bigger, better, and more comprehensive household panel studies for a number of European countries than has hitherto been possible.

5.2 Comparative birth cohort studies

A number of countries[24] have commenced work to develop new and bigger birth cohort studies than have been available previously. The opportunity to exploit the rich variety of data these studies will provide and the disciplines that must combine to make this happen (genetics, psychology, economics, sociology, education) provide a world-class opportunity that Europe should grasp.

---

23 An example of this is the €8,000 cost for a DVD and CD-Rom(s) containing a set of quarterly/yearly files covering available data in 26 countries and all years from 1983 to 2006.

24 These include the UK (a 2012 birth cohort of up to 60,000 persons), Germany (a proposed national birth cohort beginning in 2011), France (a cohort commencing in 2009), the US (a cohort commencing from 2008 to 2012) and other cohorts in Ireland, Sweden, etc.

5.3 Longitudinal studies of organizations

Comparative longitudinal studies of organizations are required to provide valuable insights into the ways in which enterprises grow, succeed, prosper, and decline in an increasingly risky global business environment. The framework for such a development exists in a number of countries (e.g., the Workplace Employee Relations Studies in the UK, the REPONSE surveys in France, the database of organization data held by the German Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung*)) and could form the core of a proposal to develop such a comparative research resource built upon existing surveys and research expertise.

5.4 Improved access to Eurostat data

Last, but not the least of these proposals, is the need to improve access to data held by Eurostat. In part, the problems of access currently faced by researchers are the responsibilities of the National Statistical Institutes which supply the data to Eurostat and which stipulate conditions for their release. This results in what is termed the "lowest common denominator" problem. For example 26 out of 27 countries stipulated that identifying information on individual records (e.g., names of individuals, names of organizations) should never be made available to researchers. But good research proceeds by allowing researchers to link between data sources, maximizing their utility and facilitating new and important research to be conducted. Concerns about data security can now be addressed via the new forms of control and access that virtual remote access provides. There is a now-pressing need to address these issues and to find innovative solutions to unlock the research potential of these truly European resources that cost the EU taxpayer many millions of Euros to create.

In addition to these specific proposals to develop new or to build on existing research infrastructures at the European level, there is a need to determine the feasibility of promoting access to some less well-established types of data within a European context. The two most obvious sources of information here are *administrative data* sources and *transaction data*. The former are derived from the administration of systems or programs (e.g., social security benefit, school records) and can often be mapped onto other resources to enhance their research potential. As a by-product of systems which are not primarily designed to provide research data, and because they are national in character, potential here may be limited, but further investigation of their research potential is warranted. Transaction data are often held by private-sector organizations and relate to the delivery of services or customer-initiated transactions (e.g., mobile phone records, shopping data). Where such companies are providing services across the European Union, the potential to use such information for Europe-wide research purposes becomes feasible. However, companies are likely to restrict access and to limit the nature of research that can be conducted from such sources. Again, some preliminary work needs to be undertaken to investigate the feasibility of using such data as Europe-wide research resources.

5. Infrastructure for High-Quality and Large-Scale Surveys

Cooperation between Academic Research and Private-Sector Agencies

Peter Ph. Mohler and Bernhard von Rosenbladt

Contact:

Peter Ph. Mohler
University of Mannheim
e-mail: peter.mohler[at]uni-mannheim.de

Bernhard von Rosenbladt
TNS Infratest Sozialforschung, Munich
e-mail: bernhard.rosenbladt[at]tns-infratest.com

Abstract

High-quality data from large-scale surveys provide a solid basis for outstanding research in the social sciences. Because of the unique demands of survey measurement in terms of the resources and skills required, it should be viewed as a specific sector of the research data infrastructure. In Germany, large-scale surveys have been established both within and outside academia, and major new projects are underway. Clearly, the sector is expanding. There is a need to discuss future challenges, not only with a focus on individual large projects, but with a view to the sector of large-scale surveys and high-quality survey measurement in general.

One aspect is the segmentation of large-scale measurement instruments in Germany along institutional lines (statistical offices, government agency research, public research institutions, and the academic community). Here, we recommend that an overall framework be developed covering all sub-sectors. A second aspect is the infrastructure required for large-scale, high-quality data collection. In Germany (outside the sector of statistical offices), this infrastructure is provided by private survey organizations. We argue that these should be recognised as relevant actors within the research data infrastructure. They have to invest in technological and human resources in order to provide the professional services required, and they need conditions and forms of cooperation that encourage this investment.

Keywords: survey research, research infrastructure, Germany

1. The notion of large-scale surveys' measurement

The survey-based analysis of social and economic structures, behaviour, and attitudes is among the great innovations of the social sciences. Today infrastructures exist for surveys of individuals, households, firms and other institutions in all developed countries, although such surveys may differ in size and quality. The present paper focuses on the subset of large-scale, high-quality surveys. This segment of survey research is one of the foundations of excellence in the social sciences. Only with a comprehensive system of large-scale measurement instruments (LMI) will the social sciences be in the position to continue and even expand their work as providers of evidence-based information and advice to citizens, political leaders, and other decision-makers (Mohler 2008). And indeed, the demand for this type of survey measurement is growing. The notion of LMI implies three elements of a social survey:

(1) "Large-scale" refers to sample size. Large samples of respondents are necessary to heighten the survey's statistical power and precision. "Large-scale" also indicates the resulting need for an effective field-force for data collection. The lower limit of "large" is not fixed but may

be illustrated by the German General Social Survey (ALLBUS, *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*) with a sample of 3,500 respondents. Examples of medium-sized samples are the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*), with about 10-12,000 households and the new pairfam Panel (*Panel Analysis of Intimate Relationships and Family Dynamics*) with a starting sample of 12,000 respondents. At the upper end, there is the German Microcensus with a sample of more than 300,000 households surveyed annually.

(2) "High-quality" refers to quality standards for survey methodology. Normally this implies some form of probability sampling and, beyond this, a commitment to quality criteria at all stages of the survey process, in accordance with the Total Survey Error (TSE) quality framework. It also implies "quality costs" in terms of higher financial budgets compared to the normal survey business.

(3) The third element is some form of continuous measurement. This may be implemented by repeated cross-sections or by longitudinal panel surveys. In organizational terms, the survey will normally be part of a medium or long-term research programme with a perspective of observing social trends or individual biographies or other issues of stability and change.

Within the range of these criteria, large-scale measurement instruments may cover different populations, such as households, individuals, enterprises, etc., and may be based on different modes of data collection, such as face-to-face interviewing, telephone interviewing, mail and web surveys, or – increasingly – mixed-mode approaches. We would like to underline that the segment of large-scale surveys discussed here covers a specific though essential part of quantitative research in the social sciences. There are many small surveys, studies of special groups or topics, ad-hoc surveys at a lower budget level – all of them are necessary and may satisfy their respective research purposes. When discussing issues of the research data infrastructure, however, the challenges of large-scale measurement instruments require specific attention.

This paper reconsiders how research needs for large-scale, high-quality survey data can be met in the future within the German social science infrastructure. We argue for treating this issue as one of strategic importance. The agenda of the German Data Forum (RatSWD) in its starting phase mainly aimed at gaining better access for the research community to the microdata collected by Statistical Offices and other public agencies. This initiative was highly successful. The work program may now move to a broader agenda, envisaging the overall architecture of data supply for the social sciences. Large-scale measurement instruments are a core segment of that data supply.

2. Review of large-scale survey measurement in Germany

Large-scale measurement instruments need an effective infrastructure for data collection. Whether this infrastructure exists, how it is organized in institutional terms, and how powerful it is may vary across countries. In order to evaluate the German situation we start with a brief review of large-scale surveys in Germany. Subsequently we look at organizational arrangements and quality standards in a comparative perspective, taking the US and the UK as points of reference.

In general terms, survey-based data can be collected by different kinds of data providers. In most countries there are governmental or semi-governmental agencies (statistical offices) conducting "official" or governmental surveys. Aside from this, in most developed countries there are independent survey organizations. These may be organized within public institutions, e.g., universities, or as professional survey research companies within the private sector. Individual survey organizations may or may not have the capacity for large-scale, high-quality measurement instruments as defined above.

Another aspect of data supply is how large-scale surveys are initiated, funded, and governed. One should be aware of the fact that academic research institutions are only one of several actors here. Government and research institutions within the public administration play an important role as well. The specific needs and institutional arrangements of academia should be discussed in this broader context.

We will now briefly review the main actors initiating LMI in Germany, just mentioning the large surveys under their respective responsibility:[1]

Type 1: Government surveys under specific legal regulation (amtliche Statistik)

In Germany, such surveys are conducted by the Statistical Offices. In organizational terms this means that the Federal Statistical Office acts as a kind of coordinator and clearing agency for 16 autonomous Statistical Offices of the German *Länder*, which normally are the actual data collection agencies. The main surveys are the annual Microcensus, a number of smaller population surveys, and a number of establishment surveys. Continuous population surveys include the Household Expenditure Survey (EVS, *Einkommens- und Verbrauchsstichprobe*, every five years) and EU-SILC (annually) dealing with income and living conditions. Special surveys conducted only once or at longer time intervals include the Time-Budget Survey and the Survey on ICT Usage. For most of these surveys, scientific

1 More information about a range of projects is provided in Rosenbladt (2008).

user files are currently available. Enabling researchers to access these datasets was the main objective of the KVI initiative and the original agenda of the German Data Forum (RatSWD).

Type 2: Government agencies research in particular policy areas (Ressortforschung)

During the last few decades, federal ministries have initiated a number of social surveys that are of general interest for social monitoring in various areas and that meet the criteria of large-scale, high-quality measurement instruments. They are designed as repeated cross-sectional surveys with sample sizes between 5,000 and 20,000 respondents. Examples are the surveys on income of the elderly (ASID, *Alterssicherung in Deutschland*, and AVID, *Altersvorsorge in Deutschland*), on vocational training and adult learning (AES, *Adult Education Survey*), on long-term care (MuG, *Möglichkeiten und Grenzen selbstständiger Lebensführung*), or on volunteering (*Freiwilligensurvey*). All these projects include extensive reporting to the public as well as Scientific Use Files for secondary analysis.

Type 3: Surveys governed by federal and state research institutes

Federal and state research institutes have initiated and funded a number of large-scale, high-quality surveys that are of general interest to the scientific community. Examples are

- the Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung*) with a broad range of projects, such as the annual Establishment Panel Survey (*IAB Betriebspanel*) or, more recently, the household panel on employment and social security (PASS) or the survey on employment biographies, qualification and competences (ALWA)
- the Federal Institute for Vocational Education and Training (BIBB, *Bundesinstitut für Berufsbildung*) with large cross-section surveys such as the survey on employment and qualification
- the Higher Education Information System (HIS, *Hochschulinformationssystem*) with its surveys of a variety of student populations
- the German Youth Institute (DJI, *Deutsches Jugendinstitut*) with its Family Surveys and Youth Surveys, now being redefined to form part of an Integrated Survey starting in 2009

- the Federal Institute for Population Research (BIB, *Bundesinstitut für Bevölkerungsforschung*) with its Gender and Generations Surveys (GGS)
- the Federal Office for Migration and Refugees (BAMF, *Bundesamt für Migration und Flüchtlinge*) with its recent survey of migrants in Germany
- the Robert Koch Institute (RKI, *Robert Koch-Institut*) with its health surveys.

Type 4: Surveys governed by the scientific community (academic research)

In Germany, there are very few surveys created and run by academic research organizations that meet the criteria of large-scale, high-quality measurement instruments. The few that can be mentioned here are ALLBUS, including the incorporated German part of the International Social Survey Programme (ISSP) conducted every two years; the German part of the European Social Survey (ESS) conducted every two years; the SOEP, a household panel with annual interviewing; the German part of the European Survey on Health, Ageing and Retirement in Europe (SHARE), a panel survey with bi-annual interviewing.

The need for such large-scale measurement instruments is recognised increasingly in the scientific community and its funding institutions. This will result in a much broader data supply in the future. In 2008, a new panel survey on family and pair relationships was started (pairfam). In 2009 a series of cohort panel surveys under the common label of the National Educational Panel Study (NEPS) will start. The German National Election Study (GNES), a system of elections surveys, will also start in 2009.

Structures of funding and governing large-scale surveys as reviewed above are related to typical patterns of data collection:

- Surveys of type 1 are designed and conducted by Statistical Offices.[2]
- Surveys of type 2 are tendered by ministries and contracted to survey organizations, which in this case often take over full research responsibility from design to reporting.
- Surveys of type 3 and type 4 are designed and governed by the respective research institutions. These institutions typically do not have their own

2 There are exceptions to this rule. For instance, in case of telephone surveys, data collection is contracted out to survey organizations because the statistical offices do not have their own infrastructure for large-scale CATI operations. The most prominent example was the ILO survey of 2002-2006.

infrastructure for large-scale data collection; therefore, by way of subcontracting, data collection is handed over to a survey organization.[3]

- These survey organizations are private-sector companies. This is to say that, aside from surveys of the Statistical Offices (type 1), data collection for large-scale surveys in Germany relies on the professional services provided by private firms. The scientific community and the public are often not aware of this fact because the publicly known owner of the survey data is the respective public research institution.

There are two conclusions from the review:

(1) In Germany, large-scale measurement instruments are not a vision for the future but an existing, well-developed segment of data supply for public bodies and the social sciences. Based on this, German research groups have been able to play a leading role in social research at the international level as well. It is true, though, that the academic community has lagged behind in establishing large-scale, high-quality measurement instruments of their own. With the major new projects launched recently, the situation is changing: academia is taking a more active role. The objective for the future is to widen the scope of large-sale survey measurement, establishing new surveys and approaching new research questions.

(2) The institutional basis of large-scale measurement instruments in Germany is a combination of public and private organizations. On the public side one finds, besides governmental agencies, research institutions working in different organizational contexts (public administration as well as academia) but all operating as part of the scientific community. On the private side, one finds survey research institutes organized as professional service companies.

The question is how to evaluate this overall structure. One may argue that it has apparently operated quite well so far, as demonstrated by the fact that LMI is well established. In recent years, much progress has been made in survey technology and measurement methods. The question remains, however, whether the existing infrastructure is sufficient for the future. The demand for large-scale surveys that provide high-quality survey measurement is rising. The number of such projects is increasing, accompanied by a tendency towards larger sample sizes, more complex survey designs, and

3 There are two main exceptions to this pattern: 1) RKI organizes data collection for its Health Surveys, which include some medical treatment, on its own. 2) HIS conducts surveys of student populations, normally implemented through mail or web surveys, on its own. Similarly surveys and assessments implemented in schools normally are conducted by specialized agencies or institutes affiliated to the respective ministry of education.

more demanding methodological techniques and standards. All of these developments will require significantly expanded survey measurement capacities.

3. Organization and standards: Germany compared to the US and the UK

Large-scale social science surveys belong to the class of high-precision scientific instruments, similar to those used in the natural sciences. In order to measure social structures, individual behavior, and social change properly, surveys must not only be large-scale but also high-quality. Conducting a large-scale survey at a poor quality level means misallocating money. Quality standards and how to implement them in practice must therefore be part of the discussion on large-scale measurement instruments.

Sample surveys may be viewed as a communication process. They are complex instruments generated in a structured and dynamic interplay of several thousand people. They must be organized in production processes requiring intensive, continuous process quality control.

To understand the enormous task of making a large-scale survey, let us consider the (relatively simple) case of an ALLBUS-type survey, a one-nation cross-sectional survey. After having designed and properly tested a questionnaire, a fielding team of about 200 interviewers (plus back-office staff) will be handed the addresses of about 7,000 selected target persons. Most of them have to be contacted several times to achieve the final net of about 3,200 respondents. Let us assume the average contact rate is 2.5. This means that some 17,500 contacts or contact attempts have to be made. The net sample of 3,200 respondents will, on average, communicate with an interviewer in a face-to-face situation for about 70 minutes (i.e., all in all about 460 working days). The instrument measures about 400 variables per respondent resulting in about 800,000 single data points or measurements, which make up the data file.

To design, implement, and successfully conduct such a survey, a number of quite distinct methods and techniques have to be combined into a single streamlined survey process. Among these are communication and cognition methods which allow the transfer of substantive research questions into appropriate survey items; sample statistics, which govern the design, implementation and assessment of actual samples; logistics and process quality methods, which guarantee transparent fielding processes; content analysis as a special field for all open-ended items and coding; documentation methods which relate numerical information with “what it means,” and statistical analysis combined with other quality measures to assess the validity and reliability of the data obtained, to mention just the major research areas.

In contrast to the sheer endless number of possible errors or distortions that can happen in this process, one can observe a remarkable robustness in many high-quality surveys over longer periods of time. This indicates that the process can be successfully managed – with ample opportunities for improvement. There is, for instance, the issue of declining response rates in combination with higher aspirations to include all strata of a society (i.e., less integrated groups), which must be tackled by future large-scale surveys.

Only the best survey organizations are able to manage this process observing rigorous scientific standards. For obvious reasons, the number of such organizations is small in all countries. Institutional settings may vary between countries; thus, for a compact overview, we will discuss the situation in the United States and the United Kingdom compared to Germany.

United States

The number of private and university affiliated agencies which are able to run large-scale measurement instruments is actually very limited in the US. Apart from the two university-affiliated agencies NORC (*National Opinion Research Center*) and ISR (*Institute for Social Research*), there are two other private sector institutes, namely Westat and RTI (*Research Triangle Institute*).

Centers like NORC or ISR in the United States, though affiliated with universities, organize their data collection in profit centers, whose aim is obviously to earn a profit. As soon as they require subsidies from the university or their head organization, they are either quickly downsized or, as was the case with some smaller survey research centers in the past, simply shut down. As profit centers, they compete for tendered and non-tendered surveys. They carry out surveys in the social sciences as well as government surveys. One can also observe a division of labor within such centers. Often principal investigators and analysts are faculty members, while data collection is dealt with by separate units, which themselves are defined as profit centers. Sometimes, the university data collection organization will compete for contracts from their own university with other agencies, say Westat or RTI.

Concerning standards and methodological rigor, the US has been in the lead for a long time. Discussions about the precision of large-scale measurement instruments (non-response, measurement error, total survey error, etc.) were initiated at US research institutions, which turn down low-quality proposals and are prepared to invest substantial sums in high-quality instruments. In general, considerably larger amounts of money are spent on high-quality surveys in the United States than in Germany.[4]

United Kingdom

The United Kingdom has a highly developed culture of survey research and considerable public spending for all kinds of surveys, whether evaluation studies of government programs or more basic monitoring of social trends. In contrast to the US, but similar to Germany, there are no academic data collection institutions. Instead, large-scale measurement instruments cooperate with private sector institutes in the data collection phase. The number of agencies powerful enough to run large-scale surveys is also very limited. The dominant data collection agency is NATCEN (*National Centre for Social Research*), a private, not-for-profit organization. Others include large survey companies such as BMRB Ltd. (*British Market Research Bureau Limited*), TNS (*Taylor Nelson Sofres*) or Ipsos.

Regarding standards, it is noteworthy that many UK agencies have introduced quality concepts and have been certified according to ISO or other standards. The Economic and Social Research Council (ESRC) has made great efforts to improve instrument quality for quantitative research in the last decade. This has produced several programs, conferences, projects, etc. targeting higher standards and better quality in social surveys. Increasingly, competitive structures are being introduced for long-term projects as well, i.e., calls for bids for long-term surveys at regular intervals. The bidders for these are academic groups, which in turn collaborate on data collection with private-sector data collection agencies. As in the US, the price level for high-quality surveys is considerably higher in the UK than in Germany.

Germany

Like the United Kingdom, Germany has no academically affiliated data collection agency with the capacity to run large-scale surveys such as ALLBUS, SOEP, ESS, EVS, ISSP, etc. Data collection thus has to be delegated to private-sector agencies. One should note here that "data collection" as a catchword covers a wide range of services that can include instrument design, sampling frame, fieldwork, data editing and processing, documentation, websites, and so on.

4 There is little public information on actual survey costs. Krosnik cited the price of $1,000 per interview in a 2006 press statement. Other sources include non-disclosed bids in international surveys and private information. The same holds for the UK.

Similar to the US and the UK, the number of survey agencies that can manage large-scale, high-quality surveys is very limited. This is particularly true for surveys based on face-to-face interviewing. Although there exist a number of well-known survey companies in Germany, a closer look at the list of large-scale surveys reviewed above reveals that in recent years there were mainly two agencies involved in this segment of research: TNS Infratest and infas. Others have reduced or even cut their face-to-face field force entirely, or are not trained for probability sampling or methodological documentation as required for high-quality surveys.

Despite the small number of actors, the market for large-scale, high-quality surveys is fairly competitive. The two survey companies mentioned have demonstrated their ability to conduct demanding social surveys successfully. Both companies provide "full service"; that is, they are able to offer data collection by all interview modes (face-to-face, telephone, mail, Internet) and to provide far more than just fieldwork: their professional services include the complete range of data collection steps as well as methodological consulting or writing research reports for clients requesting such services.

As mentioned above, the price level for conducting surveys is lower in Germany than in countries like the US or UK. Survey companies in Germany have invested a great deal in conducting "lean production" surveys in order to cope with clients' expectations of good quality at low budget levels. This is achieved partly by productivity gains through technology or very tight resource management. And partly it is achieved by cost-saving adaptations of methodological procedures (e.g., variations of random route procedures in face-to-face surveys). Relatively few surveys are budgeted sufficiently to meet the highest quality standards according to state-of-the-art social science methodology. At the same time, academic clients and survey methodologists have not really tried to understand the differences in survey production at different cost levels or to assess the quality achieved in the different types of surveys using the Total Survey Error framework.

Given the trend towards lower response rates in surveys – which is a problem in other countries as well – this has led to some concern in academia about the quality of surveys provided by "commercial" agencies. For many years there have been discussions about potential alternative structures, with a marked preference for academically affiliated survey organizations. We will come back to this issue later. At this juncture, we would like to underscore that the problems addressed in these discussions mainly affect the operation of "normal" surveys, whereas the sub-sector of large-scale surveys is by necessity more quality-driven. The growing demand for large-scale, high-quality measurement instruments makes it all the more important to establish quality standards that can be applied to all the various survey types discussed here.

4. Issues of infrastructure: Assessment and recommendations

When discussing the future of large-scale measurement instruments in Germany (and beyond), various infrastructural issues must be taken into consideration. We suggest discussing such issues on three levels: (1) infrastructure in terms of an overall framework for LMI, (2) infrastructure in terms of resources and know-how for data collection, (3) infrastructure in terms of individual LMI.

(1) Infrastructure in terms of an overall framework for large-scale survey measurement

Large-scale, high-quality measurement instruments must be defined as a core element in the research data infrastructure for the social sciences. Large-scale surveys offer a particular class of data, distinct from others such as administrative statistics on the one hand and survey or observational data for special (often ad-hoc or smaller-scope) studies on the other hand.

A crucial point is to develop an overview of the field as a whole, covering all of the types and sub-sectors of large-scale measurement instruments reviewed above. So far, such a broad view is not common. Instead, large-scale measurement instruments are segmented along institutional lines, that is, statistical offices (*Amtliche Statistik*) (type 1), government agencies research (type 2), state and federal research institutes (type 3) and the academic community (type 4). All these institutions have their specific responsibilities, budgets, and procedures, and will therefore all carry out their own large-scale surveys in the future. Yet one can imagine that there could be some kind of overarching framework or coordination.

Objectives would be to articulate the common interest in strengthening the basis for large-scale, high-quality measurement instruments in Germany; to avoid overlaps or conflicts of interest; to identify problems or needs for action; to develop institutional arrangements for the governance of large-scale measurement instruments; to serve as a platform to discuss issues of technological developments and resources with (public and private) data collection agencies; to support linkages of LMI in Germany with European and international structures, etc.

We will refrain from making organizational proposals here. It is evident that the German Data Forum (RatSWD) forms a kind of institutional nucleus for the representation of all those involved in LMI and high-quality survey measurement. It would be worth discussing how to integrate the private survey companies because of their crucial contribution to an effective research data infrastructure. It would also be useful to have a closer look at comparable institutional arrangements in other countries. A number of countries have established structures for a more coordinated or focused development of large-scale surveys. Among recent initiatives one may mention, in particular, the Survey Resources Network[5] in the UK. The institutional framework here is combined with efforts to promote survey methodology (on an comparative level see the International Workshop on Comparative Survey Design and Implementation (CSDI)).

One may discuss whether the (academic) social science community should focus on its "own" large-scale surveys, which are governed by scientific objectives and academic institutions, or whether this should be part of a broader approach. One argument for a broader approach is social science community's interest in obtaining access to all large-scale survey data, irrespective of their origin in other institutional sectors. A second point is that all institutional sectors use the same "production basis" for large-scale surveys, i.e., the resources and know-how of survey organizations. It should be a matter of common interest to assess future demands for large-scale data collection and to help existing suppliers reach their potential.

(2) Infrastructure in terms of resources and know-how for data collection

Large-scale, high-quality measurement instruments require technical resources and know-how beyond the scope of what universities or research institutions or even most of the existing survey or market research organizations have at their disposal.

As we have described above, conducting large-scale surveys means organizing complex communication processes according to methodological standards, but also as quickly and as affordably as possible. The revolution in communications media and the resulting changes in communication behavior heavily affect how surveys can be conducted today. The logistics of a survey, and partly the interviewing process as such, make use of advanced technology and need streamlined production processes.

Consequently, large-scale measurement instruments are also a matter of economic resources and economic efficiency. To build up and maintain data collection operations of the required scope requires substantial financial resources and continuous operating capacities, as well as ongoing investments and innovations to maintain competitiveness. This includes investment in the highly qualified staff who are necessary to offer comprehensive professional services and research experience.

It does not go without saying that such resources exist. To give an example, one can design a new survey of 10,000 randomly selected respon-

5 http://surveynet.essex.ac.uk/

dents who will take part in a 60-minute personal interview. It is by no means a given that such a survey can simply be "ordered" at some agency.

In Germany, the infrastructure for data collection of this scope does exist. Apart from government surveys conducted by statistical offices, the infrastructure is provided by private survey organizations. Whether the given supply satisfies all needs and whether it is advanced and stable enough to meet future requirements, however, is subject to debate.

In academia, "for-profit" survey companies are sometimes regarded as service providers that belong to the commercial sphere and are not really part of the research process. Potential alternative structures have been discussed. Theoretically, there are two alternatives:

- The first are the statistical offices (the Federal Statistical Office and 16 Statistical Offices of the German *Länder*), which could act as fieldwork organizations. Indeed one may ask whether the present division of labor between the statistical offices and survey organizations will remain the same in the future. Is it conceivable that the statistical offices might take over data collection functions for large-scale surveys in the social sciences? There are no signs indicating this. Statistical offices work under tight legal, budgetary, and organizational restrictions, which make arrangements of this sort unlikely. Moreover, the German statistical offices do not use the survey design preferred for social science projects.[6] One should also mention here the difference between the "enumerators" who work at statistical offices and are trained to list facts, and the "interviewers" in survey research who are trained to facilitate measurement of respondents' characteristics. Given the outcomes of this distinction for survey quality, researchers have criticised how key projects such as EU-SILC are being implemented in Germany.[7]

- A second alternative would be to establish an academically affiliated data collection organization. The vision is to bundle all current and future academically governed surveys to create the critical mass necessary to establish and run a large-scale data collection enterprise profitably. Is this a realistic option for the future? There are numerous obstacles to an academic survey organization. One is the enormous investment required to set

6 Social science surveys normally combine probability sampling with the condition of voluntary participation of respondents. By contrast, population surveys of the statistical offices in Germany either rely on the legal obligation of respondents to take part (Microcensus) or, if participation is voluntary, they use quota sampling.

7 Richard Hauser (2009: 11) recommends that EU-SILC "should be improved by using truly random samples, … face-to-face-interviews with multilingual questionnaires, sole responsibility of the German Federal Statistical Office, and outsourcing fieldwork to a private market research company with a well-trained and permanent staff of interviewers."

up and maintain large-scale survey operations. A second is the segmented institutional structure of the German political, administrative, and research system, which is not favorable to centralized solutions. A third problem is how to organize competition of suppliers in such a structure: would the academic fieldwork organization be protected from competition? Would it be publicly subsidized? Or would it have to act under market conditions like a private company?

Other aspects of this debate deal with the academic world itself. Motivation and career paths within the academic social sciences would have to support an academic fieldwork organization. In the United States, there have been pioneers in survey research who wanted to collect data on their terms, who thus wanted to define standards themselves and who, consequently, spent time and effort in the thicket of fieldwork. To implement the vision of an academic field organization in Germany, one must, first and foremost, create such a culture, which would be a lengthy and thus most unlikely process.

Following from this, our recommendation is to acknowledge the co-operation between public and academic research institutions and private-sector survey agencies as an integral part of the research data infrastructure.[8] Challenges of the future must be met within this framework. This strategy is in line with developments in other European countries.

It may be useful to think about institutional mechanisms to strengthen the public-private cooperation. Basically, however, the economic mechanisms of supply and demand will rule the game. The private economy will supply the required resources insofar as there is sufficient demand and the services are profitable. Investment will be encouraged if there is sufficient planning security and a price level that promises return on investment. Competition among suppliers will be a driving force to improve the effectiveness and quality of the service.

At the same time, the cooperation can take advantage of the professional competence, research experience, and scientific ambitions of many survey managers in those survey agencies that are involved in the large-scale survey business. In fact, there are examples of excellent cooperation between survey managers in data collection agencies and survey directors and their teams in public or academic research institutions. Such cooperation is an important element in the research data infrastructure. Both sides should be aware of their common interest in maintaining and developing a strong infrastructure for data collection. They are both in the same boat, sailing to new horizons.

8 One could envisage such a structure along the lines of the partnership between astronomers and the optical industry: the two work together to design telescopes; the industry produces and maintains the instruments and the astronomers use them for their observations.

5. Infrastructures in terms of individual large-scale survey instruments

There are many aspects to "infrastructure," and the term is used in a number of different ways. The EU, for instance, has set up a European Strategy Forum on Research Infrastructures (ESFRI). Here infrastructure means a large individual project of general importance. In the ESFRI case, the focus is on research infrastructures of pan-European interest. The program is not restricted to the sciences, but also includes the social sciences. The framework is not restricted to survey research projects, but out of a list of six acknowledged infrastructures of European importance there are two large-scale surveys: the European Social Survey and SHARE.

We suggest that, all large-scale measurement instruments are defined as "infrastructures" in this sense. This would emphasize a number of characteristics that are crucial for such projects: their strategic role in enhancing research in the respective fields; the perspective of continuity, including the need for secure funding; governance structures and institutional arrangements for conceptual decisions, data production, and data access; their function of creating communities of researchers in the respective field, at both the national and the international level.

Large-scale measurement instruments must be embedded in an appropriate scientific infrastructure. They should be located in a more comprehensive network of high-quality social surveys, allowing existing elements and missing crucial areas to be easily identified. Moreover, the core instrument – the survey itself – must not work as a closed shop; it should be wide open to its scientific environment. Crucial functions to enable this are R&D for continuous improvement of the core instrument, and outreach to inform the scientific community and the society at large about the potential of the core instrument.

Data collection agencies, which are usually private survey organizations, should be viewed as part of the respective "infrastructure." Selecting the most suitable survey organization will require a competitive procedure. However, after the decision is made, continuity will normally be the most favorable framework for cooperation. Stable working relationships enable learning effects on both sides. Involving survey managers in decisions about methodological design and instrument development can help to optimize the survey. Contract periods of, say, three or six years facilitate investments and returns on investment (both financial resources and know-how).[9]

One might also imagine more innovative forms of cooperation. For instance, imagine that scientists applying for funds for a future large-scale

9 Our arguments put forward here on sustainable knowledge accumulation are similar to those which led to the foundation of ZUMA in 1974.

infrastructure were to form a research alliance with a private-sector agency of their choice. It would then be up to the funders to decide whether the quality and originality of the survey justifies the funds asked for. There may even be competing proposals. Price bargaining would be part of proposal development and not part of ex post subsequent funding decisions. Or, imagine that funders were to accept the need for better quality and in turn be prepared to spend more on methodological improvement and innovation than they have so far. The effects on how surveys are organised and how quality is achieved would be tremendous. Third, imagine that the quality promised were controlled independently by the funding agencies. We leave it to the reader's imagination what a major change in actual survey measurement quality that could be.

6. Conclusions and recommendations

- Large-scale, high-quality survey measurement is a crucial foundation for excellence in the social sciences. Because of its unique demands in terms of resources and skills, it should be viewed as a specific sector of the research data infrastructure.
- In Germany, a range of surveys of this type have been established, inside and outside academia, and new large projects are being created. Clearly, the sector is expanding. While this is no doubt a positive development, there is a need to discuss future challenges not only with a focus on individual large projects, but with a view to the sector of large-scale survey measurement in general. Understanding the various meanings of "infrastructure" may help to conceptualize the issue.
- Large-scale surveys are initiated, funded, and governed in different ways. In Germany, the field is segmented along institutional lines. The key actors include the statistical offices (*Amtliche Statistik*), governmental agencies and research institutes (*Ressortforschung*) and the academic community. It would be reasonable – in terms of resources, quality standards, and access to the data – to develop an overall framework covering all these subsectors. The German Data Forum (RatSWD) is a kind of nucleus for the infrastructure needed for networking and coordination. Institutional arrangements or programs in other countries may provide additional experience and models.
- Large-scale, high-quality measurement instruments require a well-developed infrastructure for data collection. In Germany, this infrastructure

exists. For surveys other than those of the statistical offices, it is supplied by private-sector survey organizations. It may be discussed whether this structure is stable and effective enough for future demands. However, alternative structures are not realistically in sight. Private survey organizations should, therefore, be recognised as relevant actors within the research data infrastructure. They have to invest in technological and human resources in order to provide the professional services required, and they need conditions and forms of cooperation which encourage this investment.

- Single large-scale survey measurement instruments may be defined as "infrastructures" in line with the use of the term at the European level (ESFRI). Compared to normal research projects, such programs need a more highly developed institutional infrastructure and must be embedded in the scientific environment. A well-defined map of such infrastructures is a prerequisite for the long-term coherent planning of a national social science infrastructure in Germany and beyond.[10]

10 "Beyond" mainly refers to the European level, which is of particular importance for large-scale survey measurement. A vision for a European System of Social Science Instruments was set out by Mohler and Wagner (2004).

References:

Hauser, R. (2009): Household income, poverty and wealth. RatSWD Working Paper No. 53.
Mohler, P.P. (2008): Große sozialwissenschaftliche Infrastruktureinrichtungen für das 21. Jahrhundert. Manuskript.
Mohler, P.P. and Wagner, G.G. (2004): ESSI – The European System of Social Science Instruments. Foundations of Continuous Measurement on a European Level using Survey Technology. Working Paper.
Rosenbladt, B. v. (2008): Die Rolle der Umfrageinstitute in der sozialwissenschaftlichen Dateninfrastruktur. RatSWD Working Paper No. 36.

6. The Availability of Market Research Data and its Potential for Use in Empirical Social and Economic Research

Erich Wiegand

Contact:

Erich Wiegand
ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V.
Langer Weg 18
60489 Frankfurt am Main
Germany
e-mail: wiegand[at]adm-ev.de

Abstract

The great usefulness of market research data for secondary analyses lays mainly in the fields of consumer behavior, consumption patterns, and media usage. Making this data available for secondary analysis in empirical social and economic research depends on the professional codes and regulations of the market research industry, and on the readiness of private agencies and their clients to make the data available.

Many market research projects focus on specific target groups. Their potential use for secondary analyses resides in deriving representative insights from these specific groups as well as in basic and methodological research. In most cases, public access to market research data must be contractually agreed upon with the client of the research project.

For methodological reasons, access to official statistics is also important for a number of market research projects. Therefore, private research agencies should have the same privileges and access to official data as academic research institutions. As long as this access has not been established, it is unlikely that these private agencies will be eager to make their market research data publicly available.

1. Introduction

Although market research projects are commissioned and conducted in response to the problems and questions of individual clients, they can be of great interest for empirical social and economic research. In fact, many market research data contain significant potential for secondary analyses. The availability of market research data for empirical social and economic research depends not only on the relevant legal provisions and professional regulations, but also on the readiness of the market research agencies themselves and their clients to make these data available for secondary analyses. Therefore, it is critical to create a win-win situation for market research providers and clients on one side, and for researchers in the relevant fields on the other.

2. The market of market research in Germany

In developed countries, research-based information is gaining importance for decision making; the German market research industry is growing continuously – both economically and in its social reach. An empirical expression of the current state of market research can be gleaned from the following:[1]

- In 2007, the total revenue of market and social research agencies in Germany was approximately 2.1 billion Euros.
- Less than half of this revenue was associated with German-based activity, which demonstrates that market research has become a global business.
- More than 90 percent of the revenue came from quantitative research projects which have a higher potential for secondary analyses than qualitative projects.
- About two-thirds of the revenue was achieved by so-called "ad hoc studies" specially designed and conducted to solve a single problem.
- The most important clients of the market research agencies come from the consumer goods industry, the pharmaceutical industry, and the media. Clients from the public sector play a relatively minor role.
- More than 40 percent of all quantitative interviews have been conducted by telephone.

These figures indicate that the potential of market research data for secondary analyses in empirical social and economic research lays specifically in the broad field of consumer behavior, patterns of consumption (including goods and services as well as time use), and media usage.

3. Potential uses of market research data in empirical social and economic research

The potential of market research data for secondary analyses in empirical social and economic research depends on several key factors:

3.1 Information and knowledge about market research

Despite the close cooperation of associations representing market and social research interests respectively (see clause 5, below), many social and economic researchers have relatively limited information and knowledge about the market of market research and its current and future developments. This

1 http://www.adm-ev.de. [Last visited: 03/02/2010].

lack of information limits the potential uses of market research data for secondary analyses. Therefore, the situation looks like a treasure hunt: “Study the map carefully, dig deeply and you will find the treasure!” The situation is also true in reverse – the knowledge gap also applies to many market researchers with regard to the potential of empirical social and economic data from academic and non-profit research institutions.

3.2 Representativeness of market research data

As a consequence of the individualization of lifestyles and of consumption patterns, an increasing number of market research projects are focused on specifically defined and sometimes relatively small target groups. Accordingly, the sample designs of these research projects are not intended to be representative for the whole population in Germany. This does not significantly reduce the potential for using such market research data for secondary analyses in social and economic research. The research data of market research projects on specific target groups still allows for structural insights into a large number of social and economic issues regarding these target groups, as well as for basic and methodological research.

3.3 Space of time between data collection and public availability

In many cases, market research clients need research results at short notice for fast and ever-accelerating decision making. As a consequence, the “half-life” of market research data for the private-sector clients is an ever-shorter time period (i.e., market research results become outdated in an ever-shorter period of time). In principle, this development increases the potential of market research data for the purpose of social and economic research. Faster outdating of market research means a shorter time from the collection of research data to their availability for secondary analyses – provided that market research agencies and clients are prepared to make them available publicly.

3.4 Infrastructure and documentation

Academic survey research in Germany has a well-established infrastructure for secondary analyses of survey data. But for the “outside” world – including parts of the market research industry – this is all but unknown. Since the public availability of market research data for secondary analyses is not only the responsibility of the market research agencies and their clients, the

GESIS[2] Data Archive, the primary institution in charge of the data, should work to increase awareness of its activities and raise its visibility outside of the research community. For example, the introduction of a GESIS exhibition stand at the annual market research trade fair in 2008 was an important step in gaining recognition and raising the profile of this public institution within the market research industry.

In order to assess the possibilities and limitations of the research results from a single project, and to determine their comparability with the research results of other projects, detailed methodological information about the project should be made available. According to DIN ISO 20252:2006, "Market, Opinion and Social Research – Vocabulary and Service requirements," the technical description of a quantitative research project shall comprise the following details (where applicable):

- client name
- research service provider
- research objectives
- target group
- actual sample size versus projected sample size and reasons, if relevant, for not obtaining the projected sample
- date of fieldwork
- sampling method, including the procedure for selecting respondents
- data collection method
- response rate (in the case of probability samples) and the definition and method of calculating it
- type of incentives
- number of interviewers
- interviewer validation methods
- questionnaires, visual exhibits, or show cards, in addition to other relevant data collection documents
- documents, materials, or products used as part of the research project
- weighting procedures
- estimating and imputation procedures
- the reliability of the results, including (when probability samples are used) estimates of sampling variance and estimates of non-sampling errors or indicators thereof
- results based on subgroups and the number of cases used in subgroup analysis

It is the responsibility of market, opinion, and social research service providers and their clients to establish and promote these standards for documentation of research projects and research results.

2 Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*).

4. Availability of market research data for empirical social and economic research

Besides the factors described above, the potential for the use of market research data in empirical social and economic research is affected by the availability of these data for secondary analyses. The actual availability of market research data depends on the professional rules of market research and the applicable legal provisions, as well as on the readiness of market research agencies and their private and public clients to make the data available. This readiness, in turn, depends on the benefits that are connected to public availability of privately commissioned market research data.

In order to exhaust the potential of market research data it is necessary to create a mutually beneficial situation for the market research agencies and their clients as data providers on the one hand, and for the empirical social and economic researchers as data users on the other.

4.1 Professional rules and legal provisions

The already mentioned international quality standard for market, opinion, and social research (DIN ISO 20252:2006) does not contain specific requirements regarding public availability of research data. With regard to the publication of research results, the following is stipulated:

> "Research service providers may publish research results for scientific or other purposes if they have conducted the research project at their own expense, or if such publication has been contractually agreed with the client commissioning the research project, or if the latter has consented to such publication" (see DIN ISO 20252:2006, clause 4.8.4).

If this international quality standard had also addressed the public availability of research data, the corresponding requirements would probably have been defined as fully as those pertaining to the publication of research results.

The "ICC/ESOMAR International Code of Market and Social Research" (last revised in December 2007) does not contain specific professional rules regarding public availability of market research data. With regard to the publication of research results, however, it points out the mutual responsibilities of researchers and clients. Both shall "ensure that published results are not misleading" (see Article 11b).

However, it belongs to the professional responsibilities of market research agencies to safeguard the confidentiality of their clients and all documents and materials that have been provided to or produced by them in the context of research projects. This requirement applies to the research data, too. In the international quality standard for market, opinion, and social research (DIN ISO 20252:2006), the following is stipulated:

"Every effort shall be made to store records in a manner adequate to ensure…that their confidential nature is not compromised. Unless otherwise agreed, all research records shall only be available to the client…" (see DIN ISO 20252:2006, clauses 4.9.3 and 4.9.4).

Since nearly all market research industry research is commissioned by private or public clients, it is crucial that the availability of market research data is contractually agreed with the clients. Without such an agreement, the market research agencies in most cases are not allowed to make their data available for secondary analyses as part of their professional responsibility to their clients.

According to professional principles and industry rules of market research, data must be processed and transmitted to clients and any other third parties in an anonymized form in order to safeguard respondent privacy. Such research data are no longer personal data. That is, analyzing them does not allow for re-identification of single respondents and the data protection laws do not apply. However, when making research data available for secondary analyses, market research agencies shall undertake specific efforts to check and to avoid any potential problems with regard to re-identification of single respondents, especially since secondary analyses might be conducted by foreign researchers abroad.

Whether the intention to make research data publicly available for secondary analyses is something that must be integrated into the required consent of the data subjects (i.e., the respondents from whom the research data are collected), must be considered from a legal point of view. According to the Federal Data Protection Act (BDSG, *Bundesdatenschutzgesetz*), the data subjects shall be informed of the purpose of collection, processing, and use of their data. The important question is: does this legal provision mean that respondents also shall be informed of and must agree to subsequent secondary analyses when asking them to participate in a certain market research project?

4.2 Readiness of the market research agencies

Without doubt, making market research data available for social and economic research by way of secondary analyses contributes to an increased awareness and perhaps reputation of market research agencies – provided they are quoted in publications in accordance with the professional rules of the scientific community. But is this possible increase in awareness and reputation alone perceived as a (relevant) benefit from their point of view?

For a number of market research projects, the data collected by official statistics are not relevant. For other projects, however, access to the individual – of course anonymized – data collected by the statistical offices are important for methodological reasons (e.g., to calculate benchmarks or weighting factors). In these cases, the private market research agencies should have the same privileges and access to the official statistical data as academic research institutions. As long as this access is not established, it is unlikely that there will by any great willingness on behalf of the market research agencies to make their data publicly available for secondary analyses. In principle, however, this willingness already exists – in both market research agencies as well as their clients – a fact that is evident from numerous examples and illustrated by the following:

In 2005, a conference on data fusion and data integration was organized jointly by the Federal Statistical Office, the Working Group of German Market and Social Research Institutes (ADM, *Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute*), and the Working Group of Social Science Institutes (ASI, *Arbeitsgemeinschaft Sozialwissenschaftlicher Institute*). A presentation was given by Heiner Meulemann and others on the potential of media consumption data for secondary analyses. These data have been collected since 1954 in order to provide a reliable empirical basis for the commercial purposes of media planning. These data have been archived at the Central Archive for Empirical Social Research at the University of Cologne (ZA, *Zentralarchiv für Empirische Sozialforschung*) from the very beginning. Therefore, they comprise a valuable source for secondary analyses, especially in the fields of media usage, social structure, and social change in addition to research methodology.

4.3 Readiness of the market research clients

The readiness of private-sector market research clients to share research data from projects they have commissioned and to make it available for the secondary analyses of social and economic research depends largely on the value of the data. As long as market research results are contributing to the success of a client's business, this readiness does not exist at all. Only when the research data no longer provide a competitive business advantage are clients potentially willing to make research data available to the broader scientific community. At this point, the readiness of market research clients to make privately purchased research data available largely depends on hearing a persuasive argument that it is of mutual benefit for both sides.

4.4 Establishing the win-win situation

For the market research industry, regular access to data from official statistics is a key factor influencing the readiness of research agencies as well as their clients to make their data available for secondary analyses. In order to create a truly win-win situation, access to market research data might be attached to certain conditions; for example, a commitment to making the research data from single projects publicly available for subsequent analyses might be granted where individual statistical data from official sources have been used in order to conduct the research project.

It is clear that in order for private market agencies to have regular access to official statistical data, the Federal Statistics Law (BStatG, *Bundesstatistikgesetz*) would have to be revised, since § 16 Abs. 6[3] stipulates that access to individual data is restricted to institutions that carry out independent scientific research. Since the Basic Law for the Federal Republic of Germany (*Grundgesetz für die Bundesrepublik Deutschland*) does not differentiate between basic and applied research but protects the freedom to conduct both types of research in the same manner, however, the restriction of access to individual statistical data for “independent scientific research” seems inappropriate.

Of course, in the long term, it is the responsibility of ADM to negotiate with the relevant political bodies in order to revise the BStatG accordingly. However, the political representation of interests in this field probably will not be successful without a strategic alliance with the associations and institutions representing empirical social and economic research.

5. The role of the associations

In Germany, the close cooperation of professional and trade associations representing market, opinion, and social research has a long tradition and is more intensive than in many other countries. This cooperation is focused mainly on self-regulation in a broad sense by defining professional rules and developing quality standards, including the formation of a joint disciplinary body as well as organizing common conferences on a regular basis. In the case of the latter, the Federal Statistical Office is involved, too. In the future, the comparatively few contacts between associations representing the private-sector and the academic research community and their respective bodies need to be intensified.

In terms of the potential use and availability of market research data for social and economic research, these points of contact between the associations representing private-sector and academic empirical survey research respectively provides the following advantages:

3 The citations to German legal sources have been left in German to guarantee accuracy.

- They help increase mutual understanding between market research agencies and the social and economic research community, a precondition for exhausting the potential of market research data for secondary analyses.
- They help maintain the research data infrastructure of empirical survey research and thus improve its mutual benefit for research data suppliers and users as well.
- They are a precondition for organizing joint conferences, meetings, and workshops – offering important means for ensuring the above two bullet points.
- They intensify relations between suppliers and users of research data, an important step toward ensuring the public availability of data for secondary analyses.
- They help establish the strategic alliance in political representation of interests in order to create the win-win situation described above.

6. Summary

The potential for market research data to contribute to the field of empirical social and economic research lies mainly in research areas dealing with consumer behavior, consumption patterns, and media use. The practical availability of market research data for secondary analyses is affected by professional rules within market research, legal provisions, and the readiness of market research agencies and their clients. A key factor determining the readiness of the market research industry to make its data publicly available is whether it has regular access to official statistical data for private market research agencies, similar to the privileges academic research institutions enjoy. The cooperation of professional and trade associations which represent market, opinion, and social research interests in Germany will play a major role defining the future possibilities for secondary analyses of market research data.

7. Recommendations

(1) There should be continued effort on both sides and by various means (e.g., joint conferences, workshops, newsletters, etc.) to improve understanding and knowledge between market research agencies and the research community.

(2) The data infrastructure of empirical survey research should be promoted more actively outside the academic scientific community to increase its visibility for market research agencies and their clients.

(3) Existing standards for documentation of the methodological and technical details of research projects both by the market research industry as well as by empirical social and economic research need to be more strongly reinforced.

(4) The availability of market research data for secondary analyses for purposes of empirical social and economic research should be agreed upon contractually when a research project is commissioned.

(5) In order to improve the readiness of private market research agencies and their clients to make market research data publicly available for secondary analyses, the bodies representing academic social and economic research should actively support the efforts by private market research agencies to acquire regular access to statistical data.

(6) The bodies representing the empirical economic research community should be included in the forms of cooperation that already exist between the associations representing market, opinion, and social research in Germany and public-sector agencies.

References:

ADM Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e.V. (www.adm-ev.de, 2008).

DIN ISO 20252:2006 Market, opinion and social research – Vocabulary and service requirements.

ICC/ESOMAR International code on market and social research (last revised December 2007).

Meulemann, H./Hagenah, J. and Akinci, H. (2005): Die Media-Analysen. Synopse des Datenbestands und Nutzungschancen für Sekundäranalysen des sozialen Wandels in Deutschland seit 1954. In: König, Ch./Stahl, M. and Wiegand, E. (Eds.): Datenfusion und Datenintegration. 6. Wissenschaftliche Tagung.

PROGRESS SINCE 2001 AND CURRENT STATE

1. The Recommendations of the 2001 KVI Report and Their Realizations up to 2008

Gabriele Rolf-Engel

Contact:

Gabriele Rolf-Engel
Business Office of the German Data Forum (RatSWD)
c/o DIW Berlin
10108 Berlin
Germany
e-mail: grolf-engel[at]ratswd.de

In 1999 the German Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*) appointed a commission to examine the information infrastructure in Germany and to make proposals for its improvement. The central objective was to improve the cooperation between the scientific community and official statistical agencies. The German Commission on Improving the Information Infrastructure between Science and Statistics (KVI) presented its final report in 2001 (KVI 2001).

The Commission's report consisted of:

- a stocktaking of deficits and data needs in different fields (e.g., population, employment, income and wealth data, etc.),
- an overview of the major data producers, data providers, and statistical databases (e.g., official statistics, social security statistics, government ministry data, administrative data, scientific data, private market data, and data from social research institutes and commercial providers) and the practice of providing access to anonymized microdata,
- an international comparison of the best statistics and best practices of statistical analysis, and
- recommendations on improving the performance of the information infrastructure for empirical economic and social research in Germany.

The Commission developed 36 recommendations on:

- improving cooperation between the scientific community and official statistical agencies,
- expanding participation of the scientific community in developing survey and data processing programs (by official statistical agencies as well as by ministries and non-statistical institutions conducting surveys),
- priorities for continuing and developing important statistics,
- supporting research on data collection, processing, and archiving,
- higher education and training,
- data access, especially access to microdata,
- confidentiality of research data, and
- implementation and funding.

The following synopsis gives an overview of the Commission's recommendations. The second column lists the objectives of the recommendations, some of which are formulated explicitly in the Commission's report, and some of which I have deduced from the proposed solutions listed in the third column. The traffic signals in the fourth column illustrate the extent to which results of the recommended action are already evident: green indicates that the objectives have been fully achieved; yellow indicates that work is still in progress; and red indicates that there remains significant further work to be done. Since these conclusions may be in dispute, the fifth column provides additional explanatory comments.

As the traffic signals in the following synopsis show, many of the Commission's recommendations have already been put into effect, some of the most important being:

- the establishment of the German Data Forum (RatSWD) and its predecessor the Founding Committee of the Council for Social and Economic Data,
- the establishment and evaluation of several Research Data Centers and Data Service Centers that are working to improve access to microdata and facilitate data analysis,
- new means of data access. In addition to the distribution of Scientific Use Files and Public Use Files, controlled remote data access is provided. Furthermore, workplaces are being provided for guest researchers in the Research Data Centers,
- improved cooperation and information exchange between the scientific community and official statistics through:
  - the German Data Forum (RatSWD), as a platform for communication,
  - the biennial Conference for Social and Economic Data (KSWD, *Konferenz für Sozial- und Wirtschaftsdaten*),
  - dialog workshops in the fields of media data, crime statistics, household statistics, and globalization,
  - joint research projects on data access, statistical development, and methodological development,
  - the appointment of three working groups by the German Data Forum (RatSWD) dealing with crime statistics, new means of access to microdata for Germany, and preparation of a German National Death Index, and
  - the establishment of the Census Commission,
- access to anonymous firm-level data and opening up longitudinal microdata, and

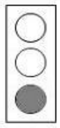
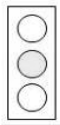
- capacity-building (young scholar workshops, expert report competitions for young researchers, CAMPUS-Files using realistic datasets).

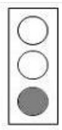
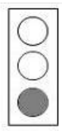
Nevertheless, there is still a substantial need for action, especially when implementing the Commission's recommendations, in terms of:

- legislative action,
- international activities,
- coordination within and between organizations on a voluntary basis and/or without sufficient budget (e.g., facilitating dialog within the scientific community).

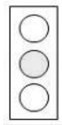
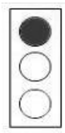
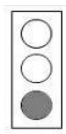
Last but not least, continuous funding of the existing infrastructure remains a problem. This applies both to the permanent institutionalization of the German Data Forum (RatSWD), which has been financed up to now as a pilot project of the BMBF, and to permanent funding for the Research Data Center of the Statistical Offices of the German *Länder*.

Synopsis: Recommendations of the Commission

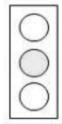
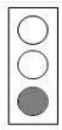
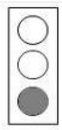
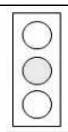
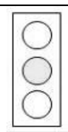
	Task completed or on schedule green		Room for improvement yellow		Project structure not yet visible red

Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
Improved cooperation between science and statistics is necessary for data users and data producers			
1	To improve cooperation between the scientific community and official statistical agencies based on the traditional model of a division of labor	Adopt and enforce institutional regulations	
2	To assess and to improve the information infrastructure based on input from data producers and data users To develop a platform for structured dialog between data producers and data users	Establish a Council for Social and Economic Data, whose main functions are: ▪ assessing and improving the data infrastructure and advising the federal and state governments on programs of science-based statistics and their funding ▪ promoting social and economic reporting ▪ recommending the establishment of Research Data Centers and Data Service Centers and evaluating their activities ▪ suggesting how project funds should be allocated These tasks need coordinators in the group of data producers, in the group of data users, and between the two groups.	

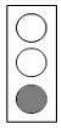
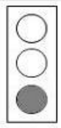
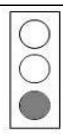
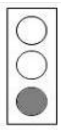
Status
Comment
Several important activities have been carried out, especially the establishment of the German Data Forum (RatSWD), which offers a platform for dialog between data providers and data users (see recommendation 2) and the KSWD, which takes place every two years.
Measures In 2004 the German Federal Ministry of Education and Research (BMBF) set up the RatSWD. This Council includes empirical researchers from universities, colleges, and other independent research institutions as well as data producers and representatives of service centers. The predecessor of the RatSWD, the Founding Committee of the Council for Social and Economic Data (GA, *Gründungsausschuss*), was founded in 2001. The GA and the RatSWD have undertaken a great number of activities to improve the research data infrastructure in Germany (Rolf et al. 2008; Solga and Wagner 2007), particularly by offering a platform for dialog between data providers and data users and by advising Germany's federal and state governments on the establishment of Research Data Centers and Data Service Centers and by evaluating their work. Additionally, the GA and the RatSWD have contributed to improving the research data infrastructure by assessing projects in terms of data access and methodological developments in the social and economic sciences. In its first few years of work, the RatSWD's activities have focused on improving data access for empirical research. Now, the need to improve survey development and processing programs has moved to the forefront of the RatSWD's agenda. *To be done:* Ensure permanent funding for the RatSWD, which has funding from the BMBF for an initial period. The German Council of Science and Humanities (*Wissenschaftsrat*) has evaluated the RatSWD positively and recommends its permanent funding (Wissenschaftsrat 2009a).

Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
Participation by the scientific community in developing survey and processing programs			
3	To involve the scientific community both in improving the survey and processing programs of the official statistical agencies and in special hearings by German parliament on this subject	Adopt institutional regulations Improve coordination in the scientific community (in discussions of the Council for Social and Economic Data in cooperation with the relevant scientific associations)	
4	To make survey and processing programs of the official statistical agencies more flexible	Reduce strict legal regulations and expand the scope for statistical offices and their advisory bodies to shape survey programs	
5	To expand the influence of the scientific community in proposing modifications of official statistical programs	▪ Broaden the definition of the Statistical Advisory Committee's tasks (including medium-term program planning)	
		▪ Achieve fuller representation of the scientific community in the Statistical Advisory Committee (increase the number of representatives of empirical social and economic research)	
		▪ Hold mandatory hearings as part of the legislative process on official statistics	

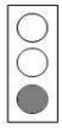
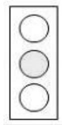
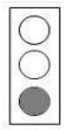
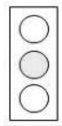
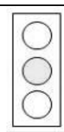
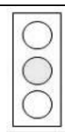
Status
Comment
Official statistical agencies are open to advice, but the scientific community has still not made sufficient use of this opportunity. *Measures* Appointment of the Census Commission (*Zensuskommission*), a scientific commission that advises the federal government and official statistical agencies on preparing, processing, and analyzing the 2011 Census. Nomination of the Census Commission's members on recommendation from the RatSWD No institutional regulations are in place, but a number of joint activities are underway, such as a series of workshops *"Dialog von Wissenschaft und amtlicher Statistik"* dealing with the 2011 register-based census, household surveys, and globalization. *To be done:* Fostering dialog in the social, economic, and behavioral sciences and mobilizing the respective scientific associations to improve their information infrastructure. Holding regular hearings with the scientific community as part of the legislative process.
Not yet visible The Statistical Advisory Committee (*Statistischer Beirat*), an organization of the users, respondents, and producers of federal statistics, has called for more flexibility in designing the statistical system (Statistisches Bundesamt 2003).
Not yet visible The Statistical Advisory Committee has recommended medium-term program planning (Statistisches Bundesamt 2003). The scientific community has attained greater influence on the Statistical Advisory Committee through an additional representative of empirical social and economic research appointed by the RatSWD. Not yet visible

Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
6	To increase the influence of the scientific community on surveys conducted by ministries and non-statistical authorities (e.g., Deutsche Bundesbank, Federal Employment Agency, and the social security institutions)	Provide structured opportunities for scientific advice	
Priorities in continuing and developing important statistics			
7	To continue collecting important official statistics	Conduct a population census	
8	To enhance and to develop important official statistics	Further develop the German Microcensus by ▪ introducing a rotating panel ▪ developing an access panel ▪ presenting exact data on gross earned income ▪ providing Scientific Use Files	
9	To enhance and develop important official statistics	Improve the sample survey of income and consumption (EVS, *Einkommens- und Verbrauchsstichprobe*) by ▪ reducing the time intervals between the sample surveys ▪ introducing a rotating panel ▪ presenting detailed wealth data	
10	To bridge serious gaps in business sector statistics	▪ Further develop statistics on the service economy ▪ Achieve better statistical coverage of business modifications	

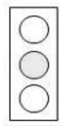
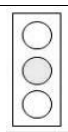
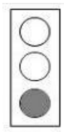
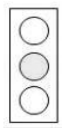
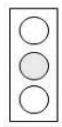
Status
Comment
As of yet there exist no structured opportunities for science to exert greater influence over official surveys, but informal steps have been taken, for example, by including scientific advisory councils in survey planning.
The 2011 Census is on schedule: http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Navigation/Zensus/Zensus,templateId=renderPrint.psml__nnn=true
The Microcensuses are available as Scientific Use Files. The joint project "Preparation and Provision of the Microcensus as a Panel Sample" has been carried out with participation of the German Federal Statistical Office, the Research Data Centers of the German Federal Statistical Office and the Statistical Offices of the German *Länder*, the Freie Universität Berlin, and the Centre for Survey Research and Methodology (ZUMA) funded by the BMBF and the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*). Today, the 1996-1999 Microcensus panel and the 2001-2004 Microcensus panel are available for research as Scientific Use Files. As of yet, exact data on gross earned income are not available.
The proposed measures have not been implemented so far. This must be seen within the overall context of household surveys: a discussion is underway between the scientific community and official statistical agencies concerning new concepts of household surveys (both in Germany and abroad). A workshop, entitled "*Dialog von Wissenschaft und amtlicher Statistik zum Erhebungsprogramm der amtlichen Haushaltsstichproben in Deutschland,*" has been organized by the RatSWD and ZUMA: http://www.ratswd.de/ver/mannheimWS.php
The research potential of firm-level data has been improved through data matching (see recommendation 27).

Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
11	To maintain and develop important science-based statistics	▪ Ensure permanent institutionalization and funding of the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) ▪ Extend the sample	
12	To continue important science-based statistics	Continuing the ▪ German General Social Survey (ALLBUS, *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*) ▪ International Social Survey Programme (ISSP) and ▪ Welfare Surveys (*Wohlfahrtssurveys*)	
13	To provide stronger support for cohort studies such as longitudinal studies of human development	Continue existing cohort studies and generate new cohort studies that cover early childhood, adolescence, and early adulthood	
Supporting research on data collection, processing, and archiving			
14	To improve university-level teaching on the methodologies of empirical social and economic research	▪ Set up a special commission for the German Council of Science and Humanities on the current state of affairs in higher education and research regarding the methods of empirical social and economic research (or defining this area as a task of the High Commission on Empirical Economic Research) ▪ Establish professorships or research centers at universities to focus on methodological problems of survey and official statistics	

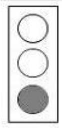
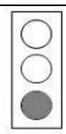
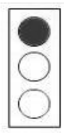
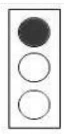
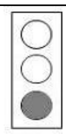
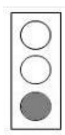
Status
Comment
Since 2003 the SOEP has been receiving ongoing funding through the Bund-Länder Commission for Educational Planning and Research Promotion (BLK, *Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung*) by Germany's federal and state governments. Several additional subsamples have expanded the possibilities for studying small societal groups. The SOEP has proposed to considerably enlarge the sample. The German Council of Science and Humanities approves of this proposal (Wissenschaftsrat 2009b).
The ALLBUS and the ISSP are conducted regularly: the ALLBUS is a continuous biennial survey, the ISSP a continuous annual program. Welfare Surveys were conducted from 1978 to 1998. Since 2002, the European Social Survey (ESS) has taken place every two years.
Existing cohort studies are continuing, such as: ▪ the SOEP, an annual survey conducted since 1984 (see recommendation 11), and ▪ the IAB Establishment Panel, an annual survey conducted since 1993. Examples of new panel studies are: ▪ the project "Educational Processes, Competence Development and Selection Decisions in Pre- and Primary School Age" (BiKS, *Bildungsprozesse, Kompetenzentwicklung und Selektionsentscheidungen im Vor- und Grundschulalter*), which is funded by the DFG, ▪ the National Educational Panel Study (NEPS), measuring competencies of children, adolescents, and adults over an extended period, which is funded by the BMBF, ▪ the Panel Analysis of Intimate Relationships and Family Dynamics (pairfam) funded by the DFG, and ▪ the Survey of Health, Ageing, and Retirement in Europe (SHARE).
In 2002, the German Council of Science and Humanities published the report "*Empfehlungen zur Stärkung wirtschaftswissenschaftlicher Forschung an den Hochschulen*" (Wissenschaftsrat 2002). The Priority Program of the DFG "Survey Methodology" has been launched: http://www.survey-methodology.de/de/projekte.html Examples of further activities enhancing higher education and research in methods of empirical social and economic research are: ▪ workshops for young researchers dealing with technical and methodological problems with complex datasets provided by the RatSWD in conjunction with official statistics and non-university research institutes, and ▪ the "European Data Watch" section of *Schmollers Jahrbuch*, which offers descriptions and discussions of micro databases that are of interest to empirical researchers: http://www.ratswd.de/publ/datawatch.php.

Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
15	To support methodological research in official statistics	▪ Strengthen the involvement of the scientific community in the further development of methodological instruments	
		▪ Include sustained methodological research in the tasks and budgets of official statistics	
		▪ Expand joint research projects by scientific and official statistical agencies	
16	To ensure the long-term preservation of statistical data	Commission the Council for Economic and Social Data to deal with the problem of archiving statistical data	
17	To promote the subject of Empirical Economic Research and to make it more visible	Recommend that the DFG establish the subject of Empirical Economic Research as a sub-discipline (or as an extension of the sub-discipline statistics)	
Higher education and training			
18	To improve education in areas such as statistics, econometrics, applied computer science, empirical methods, data collection, data	Recommend that universities and faculties improve education for ▪ undergraduates (statistics, econometrics, and applied computer science by using realistic datasets) ▪ graduate studies (statistics, econometrics, data collection, data	

Status
Comment
See recommendation 14 Efforts have been made to assign the task of "research" to official statistics in the Law on Statistics for Federal Purposes (BStatG, *Bundesstatistikgesetz*) (Hohmann 2007). Several joint research projects have been funded by the BMBF or the DFG (see recommendation 27 for an example).
The problem of archiving primary research data is currently being debated. The issues being discussed include ▪ *Rundgespräch "Forschungsprimärdaten"* of the DFG, Bonn, January 2008 http://www.dfg.de/download/pdf/foerderung/programme/lis/forschungsprimaerdaten_0108.pdf ▪ The Priority Initiative "Digital Information" of the Alliance of German Science Organizations, Berlin, June 2008 http://www.dfg.de/download/pdf/foerderung/programme/lis/allianz_initiative_digital_information_en.pdf There is consensus on not attempting to establish central rules for data archiving. Each discipline should develop its own individual solution to the problem.
The recommendation listed in column 3 has not been taken up by the DFG or the respective scientific associations. But this does not mean that the DFG does not promote empirical economic research, as the Priority Program of the DFG "Flexibility in Heterogeneous Labor Markets" shows http://kooperationen.zew.de/en/dfgflex/home.html
Curriculum development is difficult to assess because of changes in the German educational system (bachelor, master, doctorate). Positive development can be observed in the fields of post-graduate programs and teaching appointments to the staff of Research Data Centers. Measures improving the education for students and young researchers mostly taken by non-university stakeholders include:

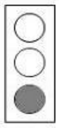
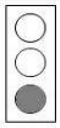
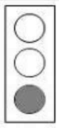
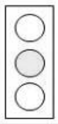
Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
	editing, data protection, and data analysis To improve statistical knowledge transfer to students	editing, data protection, and data analysis) ▪ post-graduate-programs (new empirical methods and more in-depth study of statistics and econometrics) Increasing teaching posts on the staff of official statistical agencies	
19	To make working in empirical social and economic research, statistics, and econometrics more attractive	Recommend that universities and ministries of science ▪ increase the number of professorships in empirical social and economic research, statistics, and econometrics ▪ upgrade existing associate professorships to full professorships	
20	To bring together universities, non-university research institutes, and official statistical agencies	Organize seminars, advanced training courses, and interdisciplinary summer schools in cooperation between universities, non-university research institutes, and official statistical agencies	
Economic aspects of data access			
21	To provide low-cost access to aggregated data of official statistics	Enable low-cost access to aggregated data of official statistics via Internet	
22	To provide low-cost access to Scientific Use Files and Public Use Files	Enable low-cost access to Scientific Use Files and Public Use Files Follow the example of the BMBF-funded pilot projects (providing flat-rate financing for the fixed costs of anonymization and covering the marginal costs of data delivery to the researcher)	
Access to aggregated data			
23	To promote convenient access to regionalized data via Internet	Set up a joint database system of official statistics that contains data from all federal statistical sources, broken down by region	

Status
Comment
▪ Supply of CAMPUS-Files (free Public Use Files for teaching purposes) ▪ "European Data Watch" section of *Schmollers Jahrbuch*, presenting micro databases (see recommendation 14) ▪ Expert report competitions for young researchers announced by the RatSWD ▪ Supplying a teaching module which focuses on data protection in the social sciences http://www.ratswd.de/publ/ratswd_dokumente.php ▪ Organizing young scholars' workshops (see recommendation 20)
In the social, educational, and behavioral sciences, an empirical focus seems to play a major role in professorship appointments.
Workshops on technical and methodological problems in dealing with complex data are being offered to young researchers by the RatSWD in conjunction with official statistical agencies, universities, and non-university research institutes.
Free or low-cost access to aggregated data is being provided by official statistical agencies via Internet.
Low-cost access to a large number of Scientific Use Files for scientific purposes is available; CAMPUS-Files can be downloaded for free. Costs of combining several complex datasets or of analyzing panel data are rather high.
Microdata: recommendation implemented as far as possible Macrodata: GENESIS-Online, room for improvement

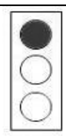
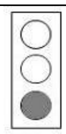
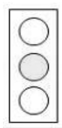
Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
Microdata access and data protection			
24	To ensure respondents' trust in data protection and to enable unlimited re-analyses	Use of different ways of access to micro datasets depending on the kind of data	
25	To guarantee confidentiality of data To ensure data protection and privacy	Periodically revise the list of technical measures developed as part of the German Anonymization Project (University of Mannheim)	
		Develop of a code of conduct describing the obligations of scientists and research institutions under data protection regulations. The code of conduct should be developed jointly by the disciplines concerned.	
		Provide certification of institutions that would benefit from the *"Wissenschafts-privileg"* (§ 16 Abs. 6 BStatG)[1]	
26	To improve access to confidential microdata	▪ Enhance the development of Scientific Use Files ▪ Provide Scientific Use Files of older data to allow analysis of social change ▪ Provide similar files such as regional Microcensus files and Microcensus panel files	
27	To permit access to business microdata	Develop anonymization strategies for data on businesses and local bodies (joint research project of the scientific and official statistical communities)	

1 The citations to German legal sources have been left in German to guarantee accuracy.

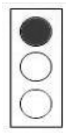
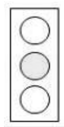
Status
Comment
Various options of data access exist depending on the degree of anonymity of the data: ▪ dissemination of Public Use Files (absolutely anonymous microdata files) ▪ dissemination of Scientific Use Files (factually anonymous microdata files) ▪ workplaces for guest researchers in the Research Data Centers ▪ controlled remote data access
Work is underway in the field of anonymization and data protection, e.g., Wirth (2006) and several anonymization projects (see recommendation 27). Several discipline specific codes of conduct, but no common code of conduct (e.g., *Ethik-Kodex der Deutschen Gesellschaft für Soziologie (DGS) und des Berufsverbandes Deutscher Soziologen (BDS)*; *Erklärung für das Gebiet der Bundesrepublik Deutschland zum ICC/ESOMAR Internationalen Kodex für die Markt- und Sozialforschung* http://www.soziologie.de/index.php?id=19 http://www.adm-ev.de/fileadmin/user_upload/PDFS/Erklaerung_2008.pdf It has proven difficult to find a common solution for the Research Data Centers (see recommendation 29) because of different legal foundations (BStatG, SGB). *To be done:* Develop a list of criteria for identifying institutions with the task of independent scientific research under § 16 Abs. 6 BStatG
Since their foundation, the first four Research Data Centers (see recommendation 29) have provided a large number of Scientific Use Files. For an overview, see: http://www.forschungsdatenzentrum.de/datenangebot.asp http://fdz.iab.de/de/FDZ_Overview_of_Data.aspx http://forschung.deutsche-rentenversicherung.de/ForschPortalWeb/contentAction.do?key=main_fdz_forschung The BMBF has financed the creation of Scientific Use Files by other data producers, too, through pilot projects such as the SUF HIS-Absolventenpanel: http://www.his.de/abt2/ab22/archiv/abs12
A number of projects (finished, in progress, or planned) have been supported by the BMBF: ▪ “Factual Anonymization of business microdata” (FAWE) ▪ “Anonymization of business panel data” (FAWE-Panel: *Anonymisierung wirtschaftsstatistischer Paneldaten*) Combining data from different surveys (and from different data producers) ▪ “Official Firm Data for Germany“ (AFiD, *Amtliche Firmendaten für Deutschland*) ▪ “Combined Firm Data for Germany“ (KombiFiD, *Kombinierte Firmendaten für Deutschland*)

Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
28	To improve access to microdata and to learn how to analyze microdata	Develop Public Use Files ▪ to train students, ▪ to meet commercial users' needs, ▪ to enable foreign scholars to access German microdata	
29	To improve and facilitate access to microdata	Establish Research Data Centers with controlled remote data access to enable use of microdata that is difficult to anonymize (i.e., when factual anonymization would impair the information in the data) and in the case of matching various datasets	
30	To improve and facilitate access to microdata	▪ Establish workplaces for guest researchers in the Research Data Centers ▪ Develop transparent procedures for the selection of guest researchers	
Using international microdata			
31	To improve the situation for research in economic and social sciences at the international level	Here a great number of measures are necessary, including ▪ developing and passing on Eurostat databases to the scientific community in the form of anonymized Scientific Use Files	

Status
Comment
Absolutely anonymous Public Use Files are provided for teaching purposes (CAMPUS-Files). See: http://www.forschungsdatenzentrum.de/campus-file.asp An internationally integrated microdata-orientated infrastructure for census research has been established: “Integrated Public Use Microdata Series – International” (IPUMS-International). See: https://international.ipums.org/international/
Measures Four Research Data Centers have been established and evaluated by the RatSWD and its predecessor, the GA: ▪ Research Data Center of the German Federal Statistical Office ▪ Research Data Center of the Statistical Offices of the German *Länder* ▪ Research Data Center of the Federal Employment Agency at the Institute for Employment Research ▪ Research Data Center of the German Federal Pension Insurance The Research Data Centers offer different means of data access, including controlled remote access. For the datasets provided, see the relevant homepages (see recommendation 26). To ensure the quality of the Research Data Centers, the RatSWD has developed a list of criteria to be met by Research Data Centers. For example, Research Data Centers should not evaluate the content of research projects applying for data access, and data producers should not maintain exclusive access to their data: http://www.ratswd.de/download/publikationen_rat/RatSWD_FDZKriterien.PDF Meanwhile, nine further Research Data Centers have adopted these standards and further data centers are scheduled to do so: http://www.ratswd.de/eng/dat/fdz.html *To be done:* Funding of the Research Data Center of the Statistical Offices of the German *Länder* on a permanent basis
Tasks completed.
A network of centers is to be established in Europe that allows access to microdata. At the end of the process, Eurostat will aim to provide remote data access to the statistics community. Examples of international projects harmonizing data from different countries: ▪ “Integrated Public Use Microdata Series – International” (IPUMS-International)”; construction of an internationally integrated microdata-orientated infrastructure for census research https://international.ipums.org/international/

Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
		▪ harmonizing data from different countries	
32	To support data exchange with research institutions in non-EU Member States	Recommend that the federal government implement a “Safe Harbor” mechanism	
Demand for services and service agency for microdata			
33	To enhance the efficiency of using microdata for research purposes	Maintain research service institutions in Germany in the future as part of the information infrastructure	
Data linkage			
34	To reduce the costs of data acquisition and the burden on respondents	Develop legal provisions on the possibility of precisely linking microdata for statistical purposes without the explicit consent of each respondent (matching only in completely shielded research and statistics areas)	

Status
Comment
▪ "Survey of Health, Ageing and Retirement in Europe (SHARE)"; cross-national panel database of microdata on health, socio-economic status, and social and family networks http://www.share-project.org/ CESSDA: One of the objectives of the Council of European Social Science Data Archives (CESSDA) is to promote the integration of the European database. http://www.cessda.org/doc/cessdaconstitution20040402.pdf ESFRI: The objective of the European Strategy Forum on Research Infrastructures (ESFRI) is to support a coherent approach to policy-making on research infrastructure in Europe. http://cordis.europa.eu/esfri/ IDF: There is an initiative to establish an International Data Forum (IDF) to facilitate the production and dissemination of social and economic data at the international level. http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/IDF%20Conference%20report%202007_tcm6-21126.pdf Establishing a European Data Forum is in discussion.
Not yet visible, but progress has been made below the level of a law.
Two Data Service Centers have been established and evaluated by the RatSWD and its predecessor, the GA, to make data analysis more convenient. These are the: ▪ German Microdata Lab, which is part of the Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*), and the ▪ International Data Service Center at the Institute for the Study of Labor (IZA, *Forschungsinstitut zur Zukunft der Arbeit*) http://www.gesis.org/das-institut/wissenschaftliche-arbeitsbereiche/dauerbeobachtung-der-gesellschaft/german-microdata-lab/ http://www.iza.org/ New developments to be mentioned here are: ▪ MISSY "Microdata Information System" http://www.gesis.org/en/services/data/official-microdata/microcensus/missy/ ▪ JoSuA "Job Submission Application" http://idsc.iza.org/metadata/
Not yet visible The project "Biographical data of selected social insurance agencies in Germany" (BASiD: *Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland)* is in its early stages. The project's aim is to construct a combined dataset for research purposes based on data from the German Pension Insurance, the Federal Employment Agency, and the Institute for Employment Research. Other approaches (statistical matching) are under discussion or in use.

Recommendation			
No.	**Objective**	**Solution**	**Traffic signal**
Confidentiality of research data			
35	To avoid trade-offs between the freedom of science and the need for data protection	Recommend that legislators introduce the principle of “research data confidentiality”: the scientist’s privilege to refuse testimony as a witness on research data and prohibition of seizure (*Zeugnisverweigerungsrecht und Beschlagnahmeverbot*)	
Implementation and funding			
36	To provide sufficient funds to implement the Commission’s recommendations	Recommend that the institutions responsible for research and science funding sponsor the activities mentioned above	

Status
Comment
Not yet visible
The BMBF has offered financial support for many of the recommended activities for a starting phase (pilot project financing). *To be done:* Permanent funding of the RatSWD and of the Research Data Center of the Statistical Offices of the German *Länder*

References:

Hohmann, E. (2007): Zur Weiterentwicklung der Forschungsdatenzentren der amtlichen Statistik. Verankerung der Aufgabe „Forschung“ im Statistikrecht von Bund und Ländern, 22. Februar 2007. http://www.ratswd.de/download/veranstal tungen/Nachtrag_Hohmann.pdf.

Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Ed.) (2001): Wege zu einer besseren informationellen Infrastruktur. Baden-Baden.

Rolf, G./Zwick, M. and Wagner, G.G. (2008): Fortschritte der informationellen Infrastruktur in Deutschland: Ein Überblick über die Jahre 1998 bis 2008. In: Rolf, G./Zwick, M. and Wagner, G.G. (Eds.): Fortschritte der informationellen Infrastruktur in Deutschland. Festschrift für Johann Hahlen zum 65. Geburtstag und Hans-Jürgen Krupp zum 75. Geburtstag. Baden-Baden.

Solga, H. and Wagner, G.G. (2007): A Modern Statistical Infrastructure for Excellent Research and Policy Advice. Report on the German Council for Social and Economic Data during Its First Period in Office (2004 – 2006). Schmollers Jahrbuch 127 (2), 315-320.

Statistisches Bundesamt (2003): Statistischer Beirat empfiehlt Novellierung des Bundesstatistikgesetzes. Pressemitteilung vom 18. Juni 2003.

Wirth, H. (2006): Anonymisierung des Mikrozensuspanels im Kontext der Bereitstellung als Scientific-Use-File. Arbeitspapier 11 des Methodenverbundes ”Aufbereitung und Bereitstellung des Mikrozensus als Panelstichprobe“. http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/Internet/DE/Content/Wissenschaftsforum/MethodenVerfahren/Mikrozensus/Arbeitspapiere/Arbeitspapier11,property=file.pdf.

Wissenschaftsrat (2002): Empfehlungen zur Stärkung wirtschaftswissenschaftlicher Forschung an den Hochschulen. Drs. 5455-02. Saarbrücken, 15. November 2002. http://www.wissenschaftsrat.de/texte/5455-02-1.pdf.

Wissenschaftsrat (2009a): Stellungnahme zum Status und der zukünftigen Entwicklung des Rates für Sozial- und Wirtschaftsdaten (RatSWD), Berlin. Drs. 9504-09. Aachen, 13.11.2009. http://www.wissenschaftsrat.de/texte/9504-09.pdf.

Wissenschaftsrat (2009b): Stellungnahme zum Status und der zukünftigen Entwicklung des Sozio-oekonomischen Panels (SOEP), Berlin. Drs. 9503-09. Aachen, 13.11.2009 http://www.wissenschaftsrat.de/texte/9503-09.pdf.

2. Access to Microdata from official statistics

Stefan Bender, Ralf Himmelreicher, Sylvia Zühlke and Markus Zwick

Contact:

Stefan Bender
Research Data Center of the Federal Employment Agency at the Institute for Employment Research
Regensburger Str. 104
90478 Nuremberg
Germany
e-mail: stefan.bender[at]iab.de

Ralf Himmelreicher
Research Data Center of the German Pension Insurance
Ruhrstr. 2
10709 Berlin
Germany
e-mail: ralf.himmelreicher[at]drv-bund.de

Sylvia Zühlke
Research Data Center of the Statistical Offices of the *Länder*
c/o IT North Rhine-Westphalia
Mauerstraße 51
40476 Düsseldorf
Germany
e-mail: sylvia.zuehlke[at]it.nrw.de

Markus Zwick
Research Data Center of the Federal Statistical Office
65180 Wiesbaden
Germany
e-mail: markus.zwick[at]destatis.de

Abstract

The four publicly funded Research Data Centers in Germany – the Research Data Center of the Federal Employment Agency, the Research Data Center of the German Pension Insurance, the Research Data Center of the Statistical Offices of the German *Länder* and the Research Data Center of the Federal Statistical Office – have made a significant improvement to the data and services available to researchers over the past few years. Their services are widely used, empirical findings lead to refereed publications and the state of research in rendering microdata anonymous has made great leaps. Many policy decisions are now planned and evaluated on the basis of data originating from the Research Data Centers. Germany has gone from the bottom of Europe's league with regard to the use of individual data to an innovative provider of new ideas, such as on access to individual data for teaching purposes and linked employer-employee datasets.

In 2007, the Research Data Centers developed criteria for their specific design in conjunction with the German Data Forum (RatSWD).

The aim of this paper is to describe the key criteria for a common working basis for the Research Data Centers, detailed descriptions of the four Research Data Centers and an outlook over future German developments.

Keywords: Research Data Center, data access, data protection, microdata

1. Introduction

The four publicly funded Research Data Centers in Germany – the Research Data Center of the Federal Employment Agency at the Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung* within the BA, *Bundesagentur für Arbeit*), the Research Data Center of the German Pension Insurance (RV, *Deutsche Rentenversicherung*), the Research Data Center of the Statistical Offices of the German *Länder*, and the Research Data Center of the Federal Statistical Office – have made significant improvements to the data and services available to researchers over the past few years.[1] Founded on the recommendation of the 2001 KVI report, and funded in the project phases by the Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*), the centers have developed in a way that was not initially anticipated. Their services are widely used, empirical findings lead to refereed publications, and the state of research in rendering microdata anonymous has made great leaps. Many poli-

1 Two Data Service Centers – the German Microdata Lab (GML) at ZUMA and the International Data Service Center (IDSC) at the Institut for the Study of Labor (IZA, *Forschungsinstitut zur Zukunft der Arbeit*) – were also set up as part of this initiative, and have also worked very successfully, see Schneider and Wolf 2008.

cy decisions are now planned and evaluated on the basis of data originating from the Research Data Centers. Germany has gone from a position at the bottom of Europe's league with regard to the use of individual data produced by empirical research with public funding, to a role as an innovative provider of new ideas, such as providing access to individual data for teaching purposes and linked employer-employee datasets.

In 2007, the Research Data Centers, in conjunction with the German Data Forum (RatSWD), developed criteria specifically focused on the design of Research Data Centers in Germany.[2] These criteria are based on the experiences of the four Research Data Centers mentioned above, which have now all been positively evaluated according to the regulations of the Leibniz Association (WGL, *Leibniz Gemeinschaft*). The criteria catalogue is designed as a guideline for other data producers planning to set up a Research Data Center.

Section 2 of this report presents the key criteria for a common working basis for Research Data Centers. Section 3 consists of more detailed descriptions of the four existing Research Data Centers as they are today. These include the respective data provided alongside further services and usage intensity. The article closes with an outlook over future developments.

2. The RatSWD criteria for Research Data Centers

Research Data Centers are institutions with the main purpose of providing simple, transparent, and high-quality access to microdata suitable for statistical analysis, while maintaining data protection and data security. Moreover, the Research Data Centers are intended to contribute to improving cooperation between the data users from the scientific community and the respective data producers. The Research Data Centers are thus an interface between the data producers' supply of data and the demand for these data from the research side. Strictly observing data protection regulations, they enable the following individual data access:

- anonymous microdata files
- controlled remote data access
- workplaces for guest researchers in the Research Data Centers

In order to provide these central services, the four publicly funded Research Data Centers have developed the following basic characteristics as criteria, in conjunction with the German Data Forum (RatSWD):

2 http://www.ratswd.de/download/publikationen_rat/RatSWD_FDZKriterien.PDF

(1) The data made available to the scientific community arise for statistical purposes as part of public administration processes, research, and evaluation, and are produced using public funding.

(2) Access to the data is subject to the legal provisions of data protection and data security in the specific area. The task of the Research Data Centers is to provide easier access through specific regulations.

(3) Access to microdata is subject to legal regulations guaranteeing equal treatment of data users. Correspondingly, the Research Data Centers ensure transparent and standardized access regulations. This also includes the regulation that no application for use shall be privileged or disadvantaged on grounds of its content. The Research Data Centers do not undertake any evaluation of the content of the research projects applying for access, but merely check for data protection or contractual permissibility. Should there be contractual or legal restrictions on the analysis of the data, these shall be published simultaneously with the provision of the data. Evaluations that give no cause for concern on grounds of data protection (i.e., are contractually permissible) may be published independently and autonomously by the users.

(4) As well as providing access possibilities, the Research Data Centers also produce data products for easy analysis and comprehensive data documentation. Moreover, information is provided via the available data and via the Research Data Centers in standardized form through websites, data, and method reports, as well as through individual consulting. The Research Data Centers' tasks also include organizing and participating in academic events in order to present the available data material, and presenting the available data and access to potentially interested parties (particularly non-university research institutes, specialized colleges of higher education, and universities). The Research Data Centers actively participate in academic discussion on the potential for analysis of existing data and in dialogues on use and development possibilities of the data infrastructure for scientific purposes.

(5) A specific amount of research must take place within the Research Data Centers. Practical research is essential to become familiar with the data and the latest methodological and content-related discussions, and thus to be able to provide users with adequate advice and instructions. The work within the Research Data Centers must not be restricted to service activities, as this would ultimately be equivalent to an exit from the scientific stage. Scientific research within the Research Data Centers enables access to further skills and qualifications and participation in scientific events, and also the publication of own work in the relevant journals.

(6) Research within the Research Data Centers is not coupled with exclusive access for the data producers. The anonymous microdata is made available simultaneously to all researchers, at least via controlled remote data access or at workplaces for guest researchers.

3. The four publicly funded Research Data Centers

From 1999 to 2001, the KVI report developed proposals for improving the data infrastructure between the scientific and statistics communities, on behalf of the BMBF. One of the commission's central recommendations was to set up Research Data Centers. There are currently four Research Data Centers in Germany that were recommended by the German Data Forum (RatSWD), which are described in detail below.

3.1 Research Data Center of the Federal Statistical Office

The Research Data Center of the Federal Statistical Office was the first center to be set up on the recommendation of the 2001 KVI report, and it was given a positive evaluation in summer 2004. The core activity of the Research Data Center of the Federal Statistical Office, processing user requests, is now funded completely from original in-house sources. The Research Data Center also receives funding from the BMBF within scientific projects, for example for rendering panel data on economic statistics anonymous.[3]

The most important official statistics are now available in the Research Data Centers of the Federal Statistical Office and the Statistical Offices of the German *Länder*, as a joint service. Access to the data, which is growing in volume, is possible in four forms, differing with regard to the type of anonymity and form of data provided. Absolutely anonymous Public Use Files (PUFs) and factually anonymous Scientific Use Files (SUFs) can be used outside of the statistical offices (off-site use). Data rendered less anonymous and containing less reduced information are made available at workplaces for guest researchers on the premises. Moreover, researchers may also work with formally anonymous individual data using their own syntax via controlled remote data access (on-site use).[4]

The most intensively used form of data are the SUFs. Approximately 710 standardized SUFs have been provided for 328 different projects since mid-2004, when the Research Data Center of the Federal Statistical Office was

3 For the problem of permanent establishment see Zwick (2006).
4 www.forschungsdatenzentrum.de

first evaluated. The most frequently used dataset is the Microcensus. Overall, however, demand for standardized SUFs is stagnating, whereas demand for individual datasets at the workplaces for guest researchers and via controlled remote data access is increasing. Controlled remote data access is now widely popular as a form of access available to researchers abroad and to commercial users. Thirty researchers have visited the Research Data Center since 2004, with controlled remote data access used in 55 projects. Eighteen further projects are currently taking place using the two forms of access.

The *Research Data Center Working Papers* series was initiated to present the wide use of the official microdata. To date, nearly thirty such working papers have been published in this series, available at the website. The Federal Statistical Office's book series, *Statistik und Wissenschaft*, also includes various volumes of articles reflecting the dialogue between the Research Data Center and the scientific community.

The Research Data Center of the Federal Statistical Office developed the series *CAMPUS-Files* especially for teaching at the university level. These files consist of absolutely anonymous microdata, allowing students to learn methodological skills and analyze sociological and economic issues. These data are available free of charge via the website of the Research Data Centers of the German Federal Statistical Office and the Statistical Offices of the German *Länder*.[5]

The Research Data Center's work focuses on further development of the access routes, anonymity methodology, and conceptional development. In order to strengthen its anchoring in the scientific community, the Research Data Center of the Federal Statistical Office is strongly present at relevant conferences (e.g., *Statistische Woche*, *Jahrestagung des Vereins für Socialpolitik*, *Kongress der DGS*). It also offers its staff the possibility to gain PhDs using the Research Data Center's resources, via two-thirds contracts.

3.2 Research Data Center of the Statistical Offices of the German Länder

The Research Data Center of the Statistical Offices of the German *Länder* took up work on 1 April 2002. Up to 2003, it focused on solving basic issues concerning funding, data access, and conditions for use. The Research Data Center has been funded by the BMBF since the beginning of 2004. Its core task is to provide easier access to the individual data of the Statistical Offices of the German *Länder*, for scientific research. In order to realize this task, a regional infrastructure was set up, enabling nationwide access to official microdata for the scientific community in sixteen regional locations. Moreover, a centralized data administration was established, which simplifies in-

5 For further information, see Zwick (2007).

terregional use of the microdata of the Statistical Offices of the German *Länder*.[6]

The Research Data Center of the Statistical Offices of the German *Länder* was evaluated in October 2006. The assessors gave a positive evaluation of the services and recommended extending the project funding and establishing the facility on a permanent basis. The project's term was thus extended up to the end of 2009 on the basis of a new funding application. The second project phase focuses on integrating economic and environmental statistical data, implementing knowledge transfer at university level, and improving ease of access to the regional locations by setting up branch offices at universities and other scientific institutions. The Research Data Center is also working towards establishing its services on a permanent basis.

The activities of the Research Data Center of the Statistical Offices of the German *Länder* have led to a broad range of microdata on various subjects for the scientific community. A total of over sixty statistics are currently available for use in academic research projects from the fields of social issues, the economy, agriculture, the environment, justice administration, and taxation. The range of data is continually extended to meet research needs. Demand in the field of economic and environmental statistical data has shown a particularly dynamic development. The demand for integrated datasets based on different statistics and survey years presents particular challenges. There are various access routes available for users.[7]

Use of the Research Data Center has increased steadily since 2004, with the number of applications for use multiplying fourfold by 2007 – from 31 to 133. On average, each application requires access to six different datasets, so that more than 2721 datasets have been provided for research purposes to date. Due to the decentralized infrastructure, most data use takes place at the workplaces for guest researchers or via controlled remote data access, and is thus particularly labor-intensive for the Research Data Center.

The previous work of the Research Data Center of the Statistical Offices of the German *Länder* has shown that the official statistical microdata are an important basis for innovative scientific analyses[8] and the users are very satisfied with the new range of services. The Research Data Center is therefore working very hard to establish its services on a permanent basis.

6 For further information, see Zühlke et al. (2004), and Zühlke et al. (2007).
7 www.forschungsdatenzentrum.de
8 Selective datasets are discussed in detail by Kaiser and Wagner (2008), Wirth and Müller (2006), Zühlke and Christians (2006).

3.3 *The Research Data Center of the Federal Employment Agency at the Institute for Employment Research*

The Research Data Center of the Federal Employment Agency at the Institute for Employment Research[9] was founded in December 2003, as there had been no systematic access to social data up until this point. Following a positive evaluation by the German Data Forum (RatSWD) in April 2006, the Research Data Center was permanently established as an independent Research Data Center of the BA at the IAB.[10] An evaluation by the German Council of Science and Humanities in 2007 confirmed that the Research Data Center was an internationally unique institution:

> “The Research Data Center (focusing on methods and data access) is an internationally visible, indispensable service institution, unique in Europe and a prime example to other institutions, possessing large datasets of scientific importance” (Report of the German Council of Science and Humanities for the IAB 2007: 55).

The Research Data Center prepares individual datasets developed in the sphere of social security and in employment research and makes them available for research purposes – primarily for external researchers. Through documentation and working tools such as the “FDZ Datenreport” and “FDZ Methodenreport” that are available online[11] and its workshops and user conferences, the Research Data Center makes it easier for external researchers to work with datasets.

The Research Data Center micro datasets include the IAB Establishment Panel, die IAB Employment Sample (IABS), the BA Employment Panel (BAP), the Integrated Employment Biographies Sample (IEBS), the Establishment History Panel (BHP), the Linked-Employer-Employee Data for the IAB (LIAB), the cross-sectional survey “Life Situation and Social Security 2005” (LSS 2005) and the first wave of the panel study “Labor Market and Social Security’” (IAB-PASS, *Panel “Arbeitsmarkt und soziale Sicherung“*).[12]

Before the Research Data Center data can be used for the first time, researchers must submit a request to use the data. Following approval by the Federal Ministry for Labor and Social Affairs (BMAS, *Bundesministerium für Arbeit und Soziales*), a use agreement is concluded between the scientist and the IAB. The number of approvals for dataset and data access has increased continuously from 81 (in 2005) to 116 (in 2007). It should also be

9 More information on the Research Data Center is available in Kohlmann (2005), Bender et al. (2008).

10 The Research Data Center has basic financing for a Head (exempt from collectively agreed terms), five positions for (senior) researchers, and three for non-academic staff and student assistants (40 hours per week).

11 http://fdz.iab.de

12 There is an English documentation on the website for nearly every dataset and a publication in the data watch section of *Schmollers Jahrbuch*.

noted that the projects normally last for over a year and thus projects from 2005 and 2006 were also continued in 2007. Two other very important parameters are the number of cases of remote data access and the number of guest stays (normally lasting several days) at the Research Data Center for on-site use. Both figures have approximately quadrupled or almost quintupled as compared to 2005 (on-site use rising from 22 in 2005 to 133 in 2007; remote data access from 359 in 2005 to 1328 in 2007). Up to 2007, researchers had published, for example, 246 articles or papers on the basis of the IAB Employment Sample, 82 using the LIAB and 1,999 using the IAB Establishment Panel, within and outside the IAB.[13]

The Research Data Center serves not only the national but also the international market. One important step towards internationalization in 2007 was the online publication of web pages in English and the translation of nearly every data documentation. The use of the Research Data Center by researchers abroad has thus increased.[14] In 2006, the Research Data Center had 16 contractual partners based abroad, including two who visited as guest researchers. In 2007, the Research Data Center counted 34 contractual partners based abroad and welcomed nine guest researchers from abroad. Guest researchers from abroad can access Research Data Center data relatively easily. It is no more difficult for them than for researchers from Germany. Since the cost of a stay in Nuremberg for visitors from abroad is higher than it is for locals, the Research Data Center established a grant to aid guest researchers in 2007. In 2007, four visitors made use of this service. The establishment of this grant was evaluated positively by the German Council of Science and Humanities in its report.

The Research Data Center is now networking more strongly with Research Data Centers in other countries. This ensures that new and innovative developments can be applied more quickly in the Research Data Center. These include, for example, anonymization of datasets through multiple imputation (Drechsler et al. 2008) or metadata databases.

For the quality of the data supply and the advisory service it is crucial, however, for Research Data Center employees to carry out empirical research themselves. The Research Data Center's research activities are well documented by its publication record. In both 2006 and 2007, Research Data Center employees published a total of ten research articles. These also include two publications in top scientific journals listed in the Social Sciences Citation Index (SSCI). This picture has been completed by numerous lectures about their research activities given in Germany and abroad.

13 The figures refer to all publications with the relevant dataset since the dataset first became available in the IAB. Some of the datasets were already available within and outside the IAB long before the existence of the Research Data Center.

14 The categorization of researchers abroad refers to their place of work, not to their nationality.

In addition, the Research Data Center of the BA participates in a number of externally funded projects, co-financed by the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*), the BMBF or the Leibniz Association and carried out in cooperation with universities, research institutes and, of course, with the other Research Data Centers. Each of these externally funded projects also included funding for personnel.

3.4 Research Data Center of the German Pension Insurance

The Research Data Center of the German Pension Insurance was founded in January 2004 and is now based in Berlin and Würzburg. During its initial setup phase – from 2004 to 2008 – the Research Data Center was funded by the BMBF.

The core task of the Research Data Center is to recover the data treasures of the German Pension Insurance (Rehfeld and Mika 2006). Alongside the microdata itself, the Research Data Center provides methodological information and commentaries intended to help simplify analyses using data from the German Federal Pension Insurance.[15]

The Research Data Center of the German Pension Insurance has realized the projects it agreed on with the funding institution (BMBF). Firstly, it has established an infrastructure within the Research Data Center, and secondly, it has taken the Research Data Center from a pilot project to a permanent institution. Thirdly, the range of data and the use possibilities have been extended considerably. In both Berlin and Würzburg, the micro datasets of the German Pension Insurance are processed in cooperation with the respective departments and the data users, to make them available to researchers particularly in the form of user-friendly Scientific Use Files.

Figure 1 gives an overview of cross-sectional and longitudinal microdata from the German Pension Insurance in the fields of retirement, insured persons and rehabilitation, with the corresponding names of the microdata. This data, highlighted in grey, is generated from the Research Data Center as anonymous SUFs, which scientists working in research units may access free of charge, the only requirement being a signed contract with the Research Data Center.

The statistics of the German Pension Insurance can be divided into datasets that focus on biographical information in combination with retirement and insurance and in a special dataset for rehabilitation. The datasets listed with a reference period of one day mean that this day represents the monitoring date in a specific year. Some statistics have both daily and annual reference periods.

15 Current information on the range of data, access routes, workshops and publications is available at www.RDC-rv.de.

The data range offered by the Research Data Center now also includes the SUFs of the two longitudinal datasets Completed Insured Life Courses 2005 (VVL, *Vollendete Versichertenleben*) and the insurance account sample 2005 (VSKT, *Versicherungskontenstichprobe*) (Himmelreicher and Stegmann 2008). Please note that it is particularly complicated to prepare the longitudinal data as SUFs, since many modifications have to be undertaken in order to render the longitudinal information anonymous (Stegmann and Himmelreicher 2008).

Figure 1: Micro datasets of the German Pension Insurance

Topic of the micro dataset				
Retirement			**Insurance**	**Rehabilitation**
Retirement inflow	Retirement stock	Retirement cash-stock		
Cross-sectional data				
Pension awarded within a certain period/ cessation of pension payment (reference period 31.12.)	**Pension payments** (reference period 31.12.)	**Pensioners with one or more pension payments** (reference period 30.06.)	**Insured persons** (reference period 31.12. and within reference year)	**Medical and occupational rehabilitation** (reference period 31.12. and within reference year)
Longitudinal data				
Completed Insured Life Courses (reference period 31.12.)			**Insurance account sample** (reference period 31.12.)	**Longitudinal dataset for rehabilitation**

Dataset available as SUF via Research Data Center of the German Pension Insurance (February 2010)

Source: Following Himmelreicher and Radl (2006).

The extended data range provided free of charge by the Research Data Center, which now also includes longitudinal data, represents a considerable improvement in usage possibilities for research. As the Research Data Center data have now been used in numerous scientific disciplines by more than two hundred and fifty researchers and an increasing number of presentations and publications are being written on the basis of the data, the Research Data Center is becoming increasingly well known in the scientific community. However, once the BMBF funding expires at the end of 2008, it will become difficult to provide the familiar standard of services to an increasing number of researchers with fewer staff.

The services of the Research Data Center and its plans for the future clearly show that the Research Data Center has recovered the large data treasures of the German Pension Insurance for research use. The newly created institution is thus on the right path. It has considerably extended possibilities for scientific analysis, while deepening the empirical knowledge in the fields of pensions, demography, and above all employment biographies.

4. Outlook

At the end of the phases funded by the BMBF, the Research Data Centers are facing new challenges. The Research Data Center of the Federal Employment Agency at the Institute for Employment Research has meanwhile been integrated into the Federal Employment Agency with all capacities from the funding phase, and now carries out its work as an organizational unit of the Federal Employment Agency at the Institute for Employment Research. The Research Data Center of the German Pension Insurance has been established as a permanent institution in the German Pension Insurance, equipped with basic funding to meet the key infrastructural needs. Additional third-party funding has to be obtained for research projects. For the Research Data Centers of the Federal Statistical Office and the Statistical Offices of the German *Länder*, possibilities for establishment on a permanent basis are still under discussion.

The Research Data Centers coordinate basic issues of data access for research purposes among each other, and work in close conjunction on various projects to extend the range of data available and the access routes. Further development of the data range will focus on integrating statistics in the near future. The projects Official Firm Data for Germany (AFiD, *Amtliche Firmendaten für Deutschland*) and Combined Firm Data for Germany (KombiFiD, *Kombinierte Firmendaten für Deutschland*) will extend the range of data in two directions: AFiD will bring together economic and environmental data from the Statistical Offices by means of the German Company Register (URS, *Unternehmensregister*) on the microdata level. KombiFiD goes one step further, uniting company data across the boundaries of the individual data producers as part of a feasibility study on a joint dataset. In addition, processes for statistical matching of survey and process-produced data are being tested, for example, between longitudinal data of the Research Data Center of the German Pension Insurance and the DIW's Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) (Rasner et al. 2007) or the IAB's longitudinal data or in several of the IAB's own projects (Bender et al. 2009, to quote one example).

In the field of data access, the Research Data Centers are looking into the procedure of remote data access. This access route has already been put to successful use in other European countries. The researchers are provided with direct access to the microdata at a specially set-up workplace in their own institutions via a secure internet connection. The Research Data Centers are currently checking the requirements for introducing this access route in Germany.

References:

Bender, S./Hartmann, B. and Herrlinger, D. (2008): FDZ Annual Report 2007. RDC Methodenreport 02/2008. Nuremberg.

Bender, S./Fertig, M./Görlitz, K./Huber, M. and Schmucker, A. (2009): WeLL - unique linked employer-employee data on further training in Germany. In: Schmollers Jahrbuch 129 (4), 637-643.

Drechsler, J./Dundler, A./Bender, S./Rässler, S. and Zwick, T. (2008): A new approach for disclosure control in the IAB Establishment Panel. Multiple imputation for a better data access. Advances in Statistical Analysis 92, 439-458

Kohlmann, A. (2005): The Research Data Center of the Federal Employment Service in the Institute for Employment Research. Schmollers Jahrbuch 125 (3), 437-447.

Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Ed.) (2001): Wege zu einer besseren informationellen Infrastruktur. Baden-Baden.

Himmelreicher, R. and Radl, J. (2006): Zusammenfassung und Ausblick auf die weitere Entwicklung des FDZ-RV. DRV-Schriften 55/2005, 281-286.

Himmelreicher, R. and Stegmann, M. (2008): New possibilities for socio-economic research through longitudinal data from the Research Data Center of the German Federal Pension Insurance (RDC-RV). Schmollers Jahrbuch 128 (4).

Kaiser, U. and Wagner, J. (2008): Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten. Perspektiven der Wirtschaftspolitik 9, 329-349.

Rasner, A./Himmelreicher, R./Grabka, M. and Frick, J. (2007): Best of Both Worlds: Preparatory Steps in Matching Survey Data with Administrative Pension Records – The Case of the German Socio-Economic Panel and the Scientific Use File Completed Insurance Biographies 2004. SOEP Data Documentation.

Rehfeld, U. and Mika, T. (2006): The Research Data Center of the German Statutory Pension Insurance (RDC-RV). Schmollers Jahrbuch 126 (1), 122-127.

Schneider, H. and Wolf, C. (2008): Die Datenservicezentren als Teil der informationellen Infrastruktur. In: Rolf, G./Zwick, M. and Wagner, G. (Eds.): Fortschritte der informationellen Infrastruktur in Deutschland. Baden-Baden.

Stegmann, M. and Himmelreicher, R. (2008): Aufbereitung der prozessproduzierten Daten der gesetzlichen Rentenversicherung im FDZ-RV. DRV-Schriften 79, 7-13.

Wirth, H. and Müller, W. (2006): Mikrodaten der amtlichen Statistik – Ihr Potenzial in der empirischen Sozialforschung. In: Dieckmann, A. (Ed.): Methoden der empirischen Sozialforschung. Wiesbaden.

Zühlke, S./Zwick, M./Scharnhorst, S. and Wende, T. (2004): The research data Centers of the Federal Statistical Office and the statistical offices of the *Länder*. Schmollers Jahrbuch 124 (4), 567ff.

Zühlke, S. and Christians, H. (2006): Datenangebot und Datenzugang im Forschungsdatenzentrum der Statistischen Landesämter. In: Landesamt für Datenverarbeitung und Statistik (Ed.): Statistische Analysen und Studien Nordrhein-Westfalen. Düsseldorf.

Zühlke, S./Christians, H. and Cramer, K. (2007): Das Forschungsdatenzentrum der Statistischen Landesämter – eine Serviceeinrichtung für die Wissenschaft. Wirtschafts- und sozialstatistisches Archiv 1 (3-4), 169-178.

Zwick, M. (2006): Forschungsdatenzentren – Nutzen und Kosten einer informationellen Infrastruktur für Wissenschaft, Politik und Datenproduzenten. Wirtschaft und Statistik 12, 1233-1240.

Zwick, M. (2007): CAMPUS-Files, Free Public Use Files for Teaching Purposes. Schmollers Jahrbuch 127 (4), 655-668.

3. Publicly Financed Survey Data: Access and Documentation

Wolfgang Jagodzinski and Christof Wolf

Contact:

Christof Wolf
GESIS – Leibniz Institute for the Social Sciences
B2, 1
68159 Mannheim
Germany
e-mail: Christof.Wolf[at]gesis.org

Abstract

This paper will address the issue of access to and documentation of survey data financed through public funds. We distinguish between four types of publicly financed survey data: (1) academic survey data from the national or international research infrastructures, (2) data from projects funded by the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*) or similarly funded projects, (3) survey data collected in research projects funded by the federal government and the German states, or *Länder* (*Ressortforschung*), and (4) population and household surveys from national and international statistical agencies. For each of these types of data we describe the current situation and present recommendations for future development.

Keywords: survey data, data access, data documentation, data archive

1. Introduction: Four data types

Our recommendations refer to four data types: (1) academic survey data from the national (such as ALLBUS[1] or SOEP[2]) or international (such as ESS,[3] SHARE,[4] ISSP,[5] European Values Study, or CSES[6]) research infrastructure; (2) data from the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*) projects or similarly funded projects; (3) data collected in research projects funded by the federal government and the German states (*Ressortforschung*); (4) population and household surveys from national and international statistical agencies. We will briefly describe the current situation and make suggestions for the future development of each of these data types. We do not attempt to give a comprehensive overview of all existing survey programs, however. We also do not address problems concerning register data.

1 German General Social Survey (ALLBUS, *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*).
2 German Socio-Economic Panel (SOEP, Sozio-oekonomisches Panel).
3 European Social Survey.
4 Suvey on Health, Ageing and Retirement in Europe.
5 International Social Survey Programme.
6 Comparative Study of Electoral Systems.

2. National and international research infrastructure

2.1 Present situation

Surveys conducted in connection with academic research are part of the research infrastructure (national and international survey programs) and provide the main source of comparative studies either in a longitudinal or in a comparative perspective. In Germany, national programs such as ALLBUS and SOEP are seen as part of the research infrastructure for the social sciences and thus they are fully funded. With regard to international surveys the situation is more heterogeneous. As far as ISSP is concerned, the costs for the German survey as well as a large share of the costs for processing the international dataset are seen as investments in the international research infrastructure and publicly funded. The European Values Study has recently reached a similar status. The last wave has been publicly funded and the costs of data processing are divided between Tilburg and the GESIS[7] Data Archive.[8]

Panel studies like SOEP are optimally suited for analyzing individual change over time. They are not only expensive, however, but also require a highly developed infrastructure for data collection and data processing. It is therefore difficult to organize multi-wave panel studies on an international level. Apart from very few exceptions, like SHARE, the large international survey programs are therefore still cross-sectional. In the meantime, most of them have built up sequences of cross-sections that permit cohort studies for the analysis of change. Standards for international surveys have recently been published by the Institute for Social Research in Michigan.[9]

There is a high demand for these studies. This is evident from the large number of data downloads and distributed copies as well as from the numerous citations of the datasets in publications. Almost all survey programs publish their own bibliography.

This demand justifies a larger investment in data documentation and data improvement. There has been some progress made in the standardization and harmonization of data. The ESS has set new standards for the documentation of international studies. Several programs have started to add contextual data to the microdata files.

Both the continuous growth and improvement of the database as well as the high demand for data in the scientific community guarantee the application of the most recent technologies in data processing and therefore an almost optimal access to the data. Although some of these programs are based on a mixed funding they largely follow the recommendations of the

7 Leibniz Institute for the Social Sciences (GESIS, Leibniz-Institut für Sozialwissenschaften).
8 http://www.europeanvaluesstudy.eu
9 http://ccsg.isr.umich.edu

OECD for fully publicly funded research data.[10] In a few survey programs the time point of general data access is still a point of discussion. As long as primary investigators are also responsible for the national funding they sometimes postpone the open data access in time. However, the situation has improved considerably over the last years. This problem would immediately be solved on a contractual basis if an international infrastructure for academic survey programs could be established. To our knowledge ESS and SHARE are so far the only science driven survey programs which receive the funding of the overhead costs from an international organization.

The other restrictions come from data protection laws. Datasets which are offered for free download on the internet therefore usually do not include fine-graded regional or occupational variables. A reduced version of the ALLBUS (ALLBUScompact) is freely accessible. Larger versions of the ALLBUS and of international social surveys like ESS, the European Values Study, or ISSP can be downloaded for free for scientific use. If data protection requires a special contract between the researcher and the user, data are distributed individually. The scientist has only to pay handling charges for data delivery.

2.2 Recommendations

It would be highly desirable if the data quality of other international survey programs could reach the quality of the ESS in the future. This would require, however, larger budgets for the international research infrastructure. The ESS has also set new standards for the documentation of sampling and data collection which should be gradually adopted by other programs. Furthermore, the translation process as well as its documentation can be improved. Until recently the translation of international surveys was under the responsibility of the national teams and largely terra incognita for secondary analysts. They could only get the final questionnaire which often did not even include interviewer instructions. Recent developments attempt to reach a higher degree of standardization and transparency.[11]

Other activities would require the institutionalization of a larger international infrastructure that would not only advise researchers in data collection and data processing but also coordinate different survey programs. In particular, the input standardization of socio-demographic variables should be achieved. It would also be desirable to improve comparability by includ-

10 http://www.oecd.org/dataoecd/9/61/38500813.pdf

11 Thus, the European Values Study 2008 has recently used the web-based translation module WEBTRANS developed by Gallup Europe for reaching a centralized control of the translation process, better comparability of the translations in different languages, more uniformity of the final questionnaires, and better documentation for comparative analyses.

ing sub-modules of items from time to time into different research programs or by integrating different surveys into a common database.

3. DFG projects and other scientific projects

3.1 Present situation

While the data access to publicly funded national and international survey programs that belong to the research infrastructure is fairly satisfying, the access to data of singular scientific projects funded by the DFG and other comparable foundations still leaves quite a lot to be desired.[12] GESIS has recently attempted to identify DFG projects from the years 2003 to 2005 that are likely to meet the acquisition criteria of the GESIS Data Archive.[13] Due to the limitations of the project documentation, however, it cannot be decided in all instances whether the project meets the criteria or not. What can be safely said, however, is that more than half of the studies which almost certainly meet the criteria are not sent to the Data Archive.[14]

Basic rules for scientific conduct require that data have to be made accessible for replication. However, they do not require delivering the data to an archive. On the one hand, in light of the cost of archival work, it is debatable whether all project data should be deposited in an archive. On the other hand, there are serious doubts whether empirical data – even if they have been stored on floppy disks or tapes years ago – still are accessible. The serious limitations of meta-analyses clearly show that access to the original data is always preferable over confining oneself to published results of statistical analysis.

3.2 Recommendations

In our view, modern information technologies allow for a substantial improvement of the present situation in two directions.

First of all, we propose the definition of a minimum standard of data accessibility that must be guaranteed by all publicly funded scientific projects: all data must be stored in a digital repository provided by the social science infrastructure. The researcher does not store the data on a disk in the

12 For a detailed description of the perspective of the German Research Foundation on the development of social science infrastructure, see Nießen and Kämper in this volume.

13 In principle, the GESIS Data Archive only accepts representative studies of populations or larger subpopulations which are relevant to social science research. It does not acquire experimental studies, for instance.

14 The results can be obtained from the authors.

university but in a domain that is maintained by a publicly funded institution. The obvious advantage of this solution for the researchers is that they do not have to concern themselves with backups and data transfer onto new computers. All these tasks are in the responsibility of the institution hosting the data repository. Special agreements between data producers and the hosting institution will address all questions concerning data ownership, data access, and data distribution. The data producer is free to choose between different options; that is, the rights to the data do not automatically go to the data host. The advantages offered by such a system would be an incentive for storing the data in a central place.

Second, we should distinguish at least between two different types of project data: those which are only relevant to a small group of scientists and data of broader interest. For the former type of data, a mode of *self-archiving* should be established. This is based on a clear division of labor: the data are stored in a central location such as the GESIS Data Archive in Cologne, but data processing and documentation are done by the primary investigator. The social science infrastructure should provide the researchers with attractive self-storage tools which help them to document and preserve the data. These tools could allow for lower and higher standards of data processing. They could also enable the researcher to build up both simple and more sophisticated databases as well as to combine data and publications. However, the project has the main responsibility for data deposition and the Data Archive should not be involved to a larger extent in this process.

Clearly, a number of questions have to be clarified before a mode of self-archiving can be established. What exactly is the division of labor between the social science infrastructure and the primary investigators? Who is responsible for the migration of data to new computer systems? Who protects the primary investigator against the violation of laws, in particular laws of data confidentiality? What kind of facilitating tools for data processing should be developed?

Self-archiving and self-documentation are not sufficient for datasets that will be of probable interest for a larger group of researchers. These data should not only be stored in the data archive but should be processed in accordance with the most advanced standards of data processing and documentation. It is advisable to consult the archive in the early stages of the project, a standard practice in all important international survey programs. The involvement of an archive requires additional resources. These resources should be included in the budget calculations of the research project from the very beginning.

One immediate objection that will be made to our proposal is that the distinction between data of restricted and broader interest is artificial and vague. For example, hasn't it sometimes turned out that a study like the election study of 1953[15] – almost forgotten in the 1950s – became extremely important for the analysis of long-term change in later decades? Yes, this happens from time to time. We would counter, however, that reviewers of project applications have good judgment and can determine whether a dataset has the potential for secondary analyses or not. Collaborative research units, for instance, will usually produce datasets that are highly salient for the scientific community at large. Moreover, if half a million or more Euros are granted for a representative national sample, it is often at least implicitly assumed that these data will not be used exclusively by the primary investigators. Details of this procedure have of course to be further elaborated.

We recommend a pilot project that will further clarify the terms and modalities of assisted self-archiving within a central data repository and professional data archiving. Such a project should also come up with proposals for self-archiving tools.

4. Research projects funded by federal or state governments (*Ressortforschung*)

4.1 Present situation

Research in this category is largely carried out by Governmental Research Agencies (GRA) and in smaller part by external researchers. GRAs have recently been evaluated by a research committee of the German Council of Science and Humanities (*Wissenschaftsrat*).[16] Besides containing evaluation reports on twenty-eight institutes, this committee has published a comprehensive report, "Recommendations on the Role and Future Development of Governmental Research Agencies with R&D Activities," in May 2006, January 2007, May 2008, and November 2008.[17] Further reports and additional recommendations were published in 2009. As far as the service of research and development infrastructure (R&D infrastructure) and data access is concerned, the recommendations from 1 April 2007 on page 11 state:

"All Federal Ministries and their agencies should avoid installing redundant and expensive R&D infrastructure. The R&D infrastructure should instead be subject to use by scientists from all kinds of R&D establishments. Such joint use requires that information on the infrastructure be readily available. Therefore, within the next two years, the BMBF in cooperation with all other federal ministries should compile a compendium listing all R&D infrastructure in GRAs (especially instruments and data). This compendium should be made available to all universities and research establishments in Germany. The Govern-

15 ZA-Study number S0145, so called Reigrotzki-Study.
16 http://www.wissenschaftsrat.de/engl_rechts.htm#EVAL
17 http://www.wissenschaftsrat.de/texte/7854-07.pdf

ment is also advised to release scientific use files to research data centers, thus granting external scientists access to specific data collections. If such data centers cannot be created, other instruments such as work places for visiting scientists should be used to facilitate access."[18]

The establishment of Research Data Centers at a subset of the GRAs will improve the accessibility of data to smaller or larger extents. Some institutes – such as the German Youth Institute (DJI, *Deutsches Jugendnstitut*) – already routinely deliver their data to the GESIS data archive. In these cases, the scientific community will benefit from new Research Data Centers mainly by having access to single and cumulative data files that so far have not been made accessible. In other instances, however, the establishment of Research Data Centers will lead to more dramatic improvements.

The research committee of the *Wissenschaftsrat* so far has focused primarily on the research of GRAs; however, quite a number of its recommendations either directly address or also apply to research projects carried out by external researchers. Therefore, we do not need to go into additional detail here but can confine ourselves to two minor issues which to our knowledge have not been systematically addressed.

The first is the Scientific Use File (SUF). It is expensive to produce and requires technical and methodological skills that often are not available at a GRA. It is more difficult to provide SUFs to the scientific community continuously than it is to establish one or two work places for visiting scientists. As a result, SUFs might actually be set at a low priority in the emerging Research Data Centers. At the same time, work places for scientists are not substitutes for SUFs, because the latter allow for more flexible and less time-consuming data analysis. SUFs therefore act as a much lower barrier against secondary analysis than workplaces in remote institutions. The report of the *Wissenschaftsrat* neither lists potential SUFs nor defines selection criteria; it does not discuss the cost-effective production of SUFs. It is particularly ambiguous in the latter respect: while the second-to-last sentence in the upper quotation can be interpreted as an indication that externally produced SUFs should be released to the new Research Data Centers, the German version by contrast defines the production of SUFs as a task of the Research Data Centers themselves.[19]

18 http://www.wissenschaftsrat.de/texte/7854-07.pdf

19 "Im Rahmen von Forschungsdatenzentren sollen 'scientific use files' erstellt werden, die externen Wissenschaftlern die Auswertung ausgewählter Datensammlungen erleichtern sollen. Wo 'scientific use files' nicht möglich sind, sollen die Forschungsdatenzentren mit Hilfe anderer Instrumente (z.B. Fernrechnen und Gastwissenschaftlerarbeitsplätze) Daten auf geeignete Weise zugänglich machen." (Within the framework of these Research Data Centers, "scientific use files" are to be created, so as to make it easier for external researchers to evaluate selected data collections. Where these files are unable to be provided, Research Data Centers are to make data appropriately accessible with the assistance of

The second problem concerns the release of data from projects which are funded by the federal or state governments. While some government departments, in particular the Federal Ministry of Family Affairs, Senior Citizens, Women and Youth (BMFSFJ, *Bundesministerium für Familie, Senioren, Frauen und Jugend*),[20] follow a fairly open policy, others are more restrictive. There is no general regulation so far.[21] If research projects of this type become visible in the media, the GESIS Data Archive directly approaches the primary investigators. Sometimes these attempts are successful and the data are acquired by the archive. Quite a few datasets, however, never become accessible for the scientific community.

4.2 Recommendations

Our recommendations focus on the two topics mentioned above. As far as SUFs are concerned, we share the preference of the *Wissenschaftsrat*. In order to secure an optimal number of SUFs, experts should first ascertain the demand for SUFs and define priorities. If the SUF is a sufficiently high priority, the most cost-efficient mode of file production has to be determined. SUFs can be produced by the Research Data Center alone, in close cooperation with an experienced external organization, or by an external organization alone. It can be distributed by the Research Data Center, by the external organization, or by both. The "Recommendations" of the *Wissenschaftsrat* and its English translation suggest two different modes of SUF production: while the German text aims at the creation of SUFs by a Research Data Center of the GRA, the English translation alludes to SUF production by the external agency. Both interpretations are correct insofar as cost-efficient solutions will differ from GRA to GRA. Presumably there is no general solution to the problem, but in any case it is highly desirable that the cost-efficient production of SUFs in this area is tackled as quickly as possible.

The question of data release should be investigated more systematically by the committee of the *Wissenschaftsrat*. In our view, the previous considerations should hold: if data from *Ressortforschung* are in the interest of the scientific community, they should in general be accessible. Data confidentiality regulations, often seen as an obstacle to data access, actually are rarely a reason for withholding a complete dataset. More often, they only require

other means, such as, for example, the allocation of visiting research positions or remote computing).

20 Negotiations between the Zentralarchiv (now: GESIS Data Archive) and the BMFSFJ have resulted in the decision that data of research projects which are funded by this government department are regularly delivered to the GESIS Data Archive at the end of the project. The datasets which the archive obtains are usually of high quality and well documented.

21 The Eurobarometers are another example of publicly funded surveys which are regularly delivered to the GESIS Data Archive.

the cutoff of some information and variables. In addition, access to sensible data may be offered in safe data centers. Free access to data for scientific purposes, in any case, should be the general rule and exceptions should be allowed only in a few, well-founded instances.

5. Household surveys from official statistics

Large-scale data collections produced under the auspices of national statistical agencies have specific strengths that make them especially interesting for social and economic research. With respect to population or household surveys, the large sample sizes and the usually very low non-response rates make these data a valuable source for economic and social-structural investigation.[22] They are regularly used for purposes of social monitoring – such as the *Datenreport* (Statistisches Bundesamt et al. 2008) – or for the construction of social indicators – as in the "Education at a Glance" (OECD 2007) or the Social Indicators Monitor SIMon.[23] However, these data are also used for a wide range of different analytical purposes, evident in the extensive bibliographies of articles based on the Scientific Use Files of the German Labor Force Survey, for example, or the German Income and Consumption Survey.

5.1 Present situation

The most important household surveys for socio-economic research from official statistics in Germany are the Microcensus, the German Income and Consumption Survey (EVS, *Einkommens- und Verbrauchsstichprobe*), and the German Time Budget Survey (*Zeitbudgeterhebung*).

The Microcensus – Germany's Labor Force Survey – is an annual random sample survey of one percent of the German population. It has been carried out in West Germany since 1957 and in reunified Germany since 1991. Integrated into the Microcensus is the German part of the European Labor Force Survey. Because participation in the Microcensus is obligatory, response rates are close to 100 percent. With over 800,000 individuals it is the largest population survey in Europe.

The EVS has been conducted every fifth year since 1963. The survey is based on a quota sample and participation is voluntary.

The Time Budget Survey is Germany's time use survey. It was conducted for the first time in 1991/92 and repeated 10 years later in 2001/2002.

22 Other data from official statistics include business surveys and process-produced data; these are dealt with in other chapters in this volume.

23 http://gesis-simon.de

The Time Budget Survey is a quota sample of over 12,000 individuals living in 5,400 households. The questionnaire of the survey complies with Eurostat's recommendations for time-use surveys and participation in the survey is voluntary.

In addition to these databases, microdata from the Censuses of 1970 and 1987 (West Germany) and from 1981 (East Germany) are currently available or will shortly be available for academic research.

In general, there are four different ways to access German microdata from official statistics:

- In the case of most of the surveys mentioned above, Scientific Use Files (SUFs) can be ordered from the Federal Statistical Office by academic or research institutions for the purpose of predefined scientific research purposes. Usage within these institutions is not restricted to German nationals, although each individual working with a SUF has to be registered as data user with the Statistical Office. SUFs are microdata files that have been reasonably anonymized. According to the Law on Statistics for Federal Purposes, this means that the files have been anonymized in such a way that any identification of individuals is only possible by excessive expenditures of time, costs, and personnel (Wirth 2008). This is typically achieved by providing only a subsample of the original dataset. In the case of the Microcensus, for instance, only a 70 percent sample is provided, deleting most of the regional information and collapsing categories with small frequencies (see also Müller et al. 1995).[24] For the Microcensus, a total of 21 SUFs are currently available, the earliest coming from 1973, the latest from 2006.[25] For the Income and Consumption Survey there are currently data from seven years, the first from 1962/1963, the latest from 2003. The data from the two waves of the Time Budget Survey are also available as SUFs.

- A second option for accessing data from official statistics is offered by the Research Data Centers of the Federal Statistical Office and the Satistical Ofices of the German *Länder*. Both offer facilities for on-site use.

- Thirdly, official microdata can be accessed remotely. In this case, the analyst provides syntax to the Research Data Centers of the Federal Statistical Office and the Satistical Ofices of the German *Länder*, the Research Data Centers execute the syntax and check if the output complies

24 Alternatively, when detailed regional information is kept, other attributes such as occupation, industry or nationality are recoded into larger categories (see Wirth et al. 2005).

25 The SUFs are created by the statistical agencies in close cooperation with the German Microdata Lab at GESIS in Mannheim (see Lüttinger et al. 2004; Schneider and Wolf 2008).

with data confidentiality requirements. This form of access is especially valuable if direct access to microdata cannot be granted due to problems of data confidentiality. This kind of problem, however, is mainly only relevant to establishment data and does not usually pose a problem for the use of household or population data. If, however, a researcher does not have the option of obtaining a SUF, for example because he or she is not working at a national research organization, then remote access might be a helpful service.

- Finally, the statistical agencies provide so called CAMPUS-Files which are Public Use Files (PUFs). These files are absolutely anonymized and can therefore be used without restriction. They are especially useful for training purposes. With respect to household surveys there are currently four CAMPUS-Files for different waves of the Microcensus and the Microcensus panel file available from the website of the Research Data Centers of the Federal Statistical Office and the Statistical Ofices of the German *Länder*.[26]

According to a recent survey among users of German microdata from official statistics, scientists clearly prefer the SUF as mode of data access. *All* respondents have used SUFs. In addition, one-fifth of users has made use of remotely processing the data and 10 percent have accessed the data in at least one of the Research Data Centers of the Federal Statistical Office and the Satistical Ofices of the German *Länder* (Lüttinger et al. 2007).

More and more researchers are interested in international comparative research. Regarding this growing demand, official microdata provided by Eurostat – the Statistical Office of the European Union – comes into focus. Eurostat currently provides access to microdata of four household surveys. These are the European Community Household Panel (ECHP), the European Union Labour Force Survey (EU-LFS), and the European Union Statistics on Income and Living Conditions (EU-SILC) (for a broader overview of European data, see Elias in this publication).

The ECHP is a panel survey that started in the twelve Member States of the European Union in 1994 and continued on an annual basis until 2001 (8 waves; some additional countries joined the survey after its initial launch). The survey covers a wide range of topics concerning living conditions including detailed income information, financial situation in a wider sense, working life, housing situation, social relations, health, and biographical information of the interviewed. The ECHP was Eurostat's attempt to create a comparative database following the principal of input harmonization (for the

26 http://www.forschungsdatenzentrum.de/campus-file.asp

different harmonization strategies see below and Ehling 2003; Granda, Wolf and Hadorn 2010).

The European Union Labour Force Survey is a rotating random sample survey covering the population in private households in currently thirty European countries. The sampling units are dwellings, households or individuals depending on the country-specific sampling frames. The collection of microdata, or individual data, started in 1983. Since 1998, the EU-LFS has developed into a continuous quarterly survey. The EU-LFS is conducted by the national statistical institutes across Europe and is centrally processed by Eurostat. The main aim of the EU-LFS is to provide comparable information on employed, unemployed, and inactive persons of working age (15 years and above) in European countries. The definitions of employment and unemployment used in the EU-LFS closely follow the guidelines put out by the International Labour Organisation. However, it follows an ex-ante output harmonization approach.

EU-SILC is an annual statistic and was launched in 2004 in thirteen Member States. From 2005 onwards the data are available for all EU25 Member States plus Iceland and Norway. Romania, Bulgaria, Turkey and Switzerland launched the EU-SILC in 2006. The EU-SILC provides cross-sectional and longitudinal microdata on income, poverty, social exclusion, living conditions and health. It can be viewed as a successor of the ECHP, though it employs an ex-ante output harmonization approach. The reference population of EU-SILC is defined as all private households and includes all persons aged 16 and over within a private household residing in the territory of the Member States at the time of data collection.

Other datasets initiated by the European Union or coordinated by Eurostat are either not available as an integrated microdata file or they are not distributed by Eurostat even though these data may be of great interest for social research (for details see the next section).

5.2 Recommendations

Among the manifold challenges we face with respect to further development in the area of population and household surveys from official statistics, there are three that seem especially pertinent from the perspective of socio-economic research: (1) continued improvement of data access, (2) adjustment of procedures to anonymize new data sources, and (3) enhancement of inter-temporal and cross-national comparability of data.

The improvement of data access can be divided into two main areas: improvement in documentation in order to ease access to data already available to the research community and the generation of access to new data sources. As is true for all secondary research, analyses of official microdata also depend on extensive documentation of the data and the data generation process. In addition, to be useful, this information has to be formatted in a standardized form and organized in such a way that it can be accessed seamlessly (e.g., a document that is stored under a pile of other documents and that can be only read with a pair of "magic glasses" obviously is of no use). An example for a very thoroughly documented statistic is the German Microcensus. The microdata information system MISSY[27], developed by GESIS, combines all available metadata for this survey and offers them in a coherently organized form through a web-based system (see Janßen and Bohr 2006).

Data access should also be improved with respect to information on field procedures. Compared to what we know about the process of data collection in social surveys such as the European Social Survey, the field work procedures utilized by the different Statistical Offices of the German *Länder* or in the different national offices of the EU are mostly terra incognita, paradata is mostly missing. The situation has improved somewhat over the last ten years, at least for the Labour Force Survey. Today we at least know the mode of interviewing (self-administered, CAPI, or CATI), the date of the interview, and if the interview is a proxy interview.

A significant problem that remains is the difficulty of access to data sources collected under the regulation of or at least coordinated by the European Union. Currently, only microdata from the above mentioned EU-LFS, ECHP, and EU-SILC are available for research outside of Eurostat. Other data such as the Adult Education Survey, the Time Use Survey, Household Budget Survey, Statistics on Information and Communications Technologies (Household Survey) or Europe's Health Survey are currently not available for comparative research. If the Lisbon goal of the European Council is to be met, namely Europe becoming the "most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion," then research monitoring this progress is mandatory and this research requires access to the relevant data.

A new challenge for data access is posed by register data that will become increasingly important over the next years. In this context, problems of integrating data from different registers and from registers and surveys has to be solved (Alda et al. 2005). Furthermore, the currently applied methods of data anonymization have to be adapted to these new data sources. However, this is not totally new terrain.

A final issue we would like to address concerning the most critical improvements to micro databases from official statistics is that of inter-temporal and especially cross-national comparability. At present, EU data is collected on the basis of regulations detailing the variables that Member States have to provide to Eurostat. This approach, called ex-ante output harmonization (Eh-

27 http://www.gesis.org/MISSY

ling 2003), leaves the concrete process of data collection to the data producer (i.e., each country has its own questionnaire and applies their own field procedures). This flexibility of data collection makes it easier for the national statistical offices to integrate the data collection process into their national programs. The comparability of data for demographic and socio-economic variables yielded by this approach is generally satisfactory. This is especially the case where international standard classifications such as the International Standard Classification of Occupations (ISCO) or the Nomenclature générale des activités economiques dans les Communautés Européennes (NACE) are available and the countries agree on their interpretation and application. However, even with such "factual" information as highest educational degree (Schneider 2008) or supervisory status (Pollak et al. 2009), output harmonization may lead to incomparable data. Naturally this is much more true for subjective data such as health status, life satisfaction or happiness, all of which are included in the EU-SILC program.

The analytical potential of microdata collected under EU regulations and integrated by Eurostat could be improved without greater cost if the following three recommendations were applied: first, although it might not be feasible and for some variables even impossible to strictly apply input harmonization, we believe that these pan-European programs have to move in this direction. Even if, as can be assumed, not all Member States agree on a blueprint for a questionnaire or on a set of data collection procedures, Eurostat could propose such a blueprint and develop a set of best practice rules for data collection.[28] Although these documents would not be legally binding, their existence would lead to them being adopted by many countries because doing so will save time and money. Second, to be able to assess data quality in more detail, all survey documents should be made available. In addition to questionnaires, these would ideally include interviewer instructions and data on the data collection process as is common practice in social surveys. Third, the harmonized and integrated datasets distributed by Eurostat should also contain the original country-specific measures at least for variables for which the harmonization process necessarily leads to a high information loss. The availability of these data would enable researchers to assess the quality of the harmonized measures and it would allow the construction of alternatively harmonized variables.

28 This strategy has been already applied with respect to the ICT Business Survey (Eurostat 2007).

6. Conclusions

In this section we have dealt with selected problems of data documentation and data access. We have not addressed the data exchange on the international level that has by and large positively developed in Germany. Foreign scientists currently have a variety of opportunities to analyze German data. International research and data centers would be a further step for improving cooperation in research and teaching.

We have only briefly touched on the progress that has been made in broadening the bases of empirical research. A number of activities aim at the generation of complex databases which combine different data types. The typical micro-macro dataset is only one example of a large variety of new sources for analysis. Empirical data can be combined with literature and publications, survey data can be combined with regional information, media data, etc. In order to create these new databases, metadata standards, in particular the DDI standard, have to be further developed (see Heus et al. in this publication). New tools enabling the linkage of different meta-databases are necessary. Some of these tools are currently developed in the context of the Preparatory Phase Project of the Council of European Social Science Data Archives (CESSDA). Interoperable meta databases finally will help to combine datasets from different years and/or different countries, thereby enlarging our resources for inter-temporal and comparative research.

References:

Alda, H./Bender, S. and Gartner. H. (2005): The linked employer-employee dataset created from the IAB establishment panel and the process-produced data of the IAB (LIAB). Schmollers Jahrbuch 125 (2), 327-336.

Ehling, M. (2003): Harmonising Data in Official Statistics: Development, Procedures, and Data Quality. In: Hoffmeyer-Zlotnik, J.H.-P. and Wolf, Ch. (Eds.): Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables. New York.

Eurostat (2007): Methodological Manual for Statistics on the Information Society. Survey year 2007, v2.0. Luxembourg. http://europa.eu.int/estatref/info/sdds/en/isoc/isoc_metmanual_2007.pdf. [Last visited 09/30/2008].

Granda, P./Wolf, Ch./Hadorn, R. (2010): Harmonizing Survey Data. In: Harkness, J. et al. (Eds.): Survey Methods in Multicultural, Multinational, and Multiregional Contexts. New York. [In print].

Janßen, A. and Bohr, J. (2006): Microdata Information System MISSY. Iassist Quaterly 30 (2), 5-11.

Lüttinger, P./Köhne-Finster, S. and Urban, J. (2007): Ergebnisse der dritten Befragung von Nutzern der Mikrozensus Scientific Use Files. GESIS Methodenbericht Nr. 1/2007. Mannheim.

Lüttinger, P./Schimpl-Neimanns, B./Wirth, H. and Papastefanou, G. (2004): The German Microdata Lab at ZUMA: Services provided to the scientific community. Schmollers Jahrbuch 124 (3), 455-467.

Müller, W./Blien, U. and Wirth, H. (1995): Identification Risks of Microdata. Evidence from experimental studies. Sociological Methods & Research 24 (2), 131-157.

OECD (2007): Education at a Glance 2007. OECD Indicators. Paris.

Pollak, R./Wirth, H./Weiss, F./Bauer, G. and Müller, W. (2009): Issues in the Comparative Measurement of the Supervisory Function using the examples of the ESS and the EU-LFS. In: Pfau-Effinger, B./Magdalenić, S.S. and Wolf, Ch. (Eds.): International vergleichende Sozialforschung: Ansätze und Messkonzepte unter den Bedingungen der Globalisierung. Wiesbaden.

Schneider, H. and Wolf, Ch. (2008): Die Datenservicezentren als Teil der informationellen Infrastruktur. In: Rolf, G./Zwick, M. and Wagner, G.G. (Eds.): Fortschritte der informationellen Infrastruktur in Deutschland. Festschrift für Johann Hahlen zum 65. Geburtstag und Hans Jürgen Krupp zum 75. Geburtstag. Baden-Baden.

Schneider, S.L. (2008): Suggestions for the cross-national measurement of educational attainment: refining the ISCED-97 and improving data collection and coding procedures. In: Schneider, S.L. (Ed.): The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries. Mannheim.

Statistisches Bundesamt, GESIS, and WZB (Eds.) (2008): Datenreport 2008. Ein Sozialbericht für die Bundesrepublik Deutschland. Bonn.

Wirth, H. (2008): Microdata access and confidentiality issues in Germany. Presentation at the meeting "Census Microdata: findings and futures", University of Manchester, 1 - 3 September 2008.

Wirth, H./Zühlke, S. and Christians, H. (2005): Der Mikrozensus als Datenbasis für die Regionalforschung. In: Grözinger, G. and Matiaske, W. (Eds.): Deutschland regional. Sozialwissenschaftlichen Daten im Forschungsverbund. München.

Wissenschaftsrat (2007): Recommendations on the Role and Future Development of Governmental Research Agencies with R&D Activities. http://www.wissenschaftsrat.de/texte/7854-07.pdf. [Last visited 02/03/2010].

4. Teaching and Statistical Training

Ulrich Rendtel

Contact:

Ulrich Rendtel
Freie Universität Berlin
Institut für Statistik und Ökonometrie / Fachbereich Wirtschaftswissenschaft
Gary Str. 21
14195 Berlin
Germany
e-mail: Ulrich.Rendtel[at]fu-berlin.de

Abstract

The availability of well-educated researchers is necessary for the fruitful analysis of social and economic data. The increased data offer made possible by the creation of the Research Data Centers has resulted in an increased demand for PhD students at the master's or *Diplom* levels. Especially in economics, where we find intense competition among the various individual subjects within the course of study, survey statistics has not been very successful in laying claim to a substantial proportion of the coursework and training. The situation is more favorable in sociology faculties.

This article argues that the creation of new CAMPUS-Files would help foster statistical education by providing Public Use Files covering a wider range of subjects. It also presents some suggestions for new CAMPUS-Files along these lines. Additionally, it argues for the establishment of master's programs in survey statistics to increase the availability of well-trained statisticians. An outline of such a master's program is presented and current PhD programs are evaluated with respect to training in survey statistics.

Training courses are also offered outside the university that promote the use of new datasets as well as expanding the knowledge of new statistical methods or methods that lie outside standard education. These training courses are organized by the Research Data Centers, (i.e. the data producers), the Data Service Centers, or by GESIS (*Leibniz Institute for the Social Sciences*). The current tendency to strengthen ties and collaborate with universities should be supported by making it possible to earn academic credit for such courses.

Keywords: master's programs, survey statistics, CAMPUS-Files, statistical training.

1. Introduction

A major issue identified by the German Commission on Improving the Information Infrastructure between Science and Statistics (KVI, *Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik*), is the relationship between data access and the ability to analyze these data competently. For this reason, the original KVI proposal voted for the creation of CAMPUS-Files, free Public Use Files (PUFs) to support academic teaching, as well as new training courses on Scientific Use Files (SUFs) (KVI 2001: 32). In this paper I review the current state of statistical teaching and training in Germany with respect to the use of new information sources that became available during the first phase of the German Data Forum (RatSWD).

Several aspects of university training in statistics will be addressed. First, both economics and the social sciences are affected by the transition from the educational model of the German *Diplom* to the bachelor's and master's program model. This transition, which is ongoing, has strong implications for university curricula. The impact of this change was not foreseen by the 2001 KVI report and is analyzed in section 4. One new feature of German university education today is the obligatory emphasis on structured doctoral programs and graduate schools. The relationship between these new branches to the present availability of statistical training is discussed in the section 5. In this context, I will also introduce the role of training courses held outside the university, namely courses run by GESIS and the Research Data Centers. Finally, some concluding remarks will be made with respect to international comparisons.

2. Consequences of improved data access

During the first phase of the RatSWD there was a strong emphasis placed on data access; namely, the development and production of SUFs and their deployment by the Research Data Centers. Corresponding to the federal structure of Germany, there is a total of sixteen state agencies and one federal state agency that offer SUFs as well as on-site access to datasets where no SUFs exist, such as firm-level data or household data with detailed regional information. This expanding data supply has resulted in a sharp increase of users. From the beginning of 2004 to 2007, the number of new data contracts rose by a factor of seven. Given that this process of improved data access has not yet come to an end, one may reasonably predict an additional dramatic increase in the number of data user contracts.

A further consequence of the increased number of research contracts at the Research Data Centers has been an increased number of job openings in the area of applied data analysis. As a register of this increased demand, I have looked to the SOEP mailing list, a forum for advertising job openings in the field of applied data analysis.[1] In this venue, the number of job offers (including academic research) has risen from eleven (in the second half of 2004) to thirty (in the first half of 2008). The positions offered are mainly part-time jobs (half or two-third positions) that include the opportunity of writing a doctoral thesis.[2]

The principal qualifications required for these positions include: competence in handling data generated by complex surveys, background in statisti-

1 The mailing list has existed in its present form since March 2004. The results reported here should be interpreted with some caution. Other effects, such as a potential increase of list subscribers, may also have induced a larger number of job offers. Help from the SOEP group, especially Uta Rahmann, in providing this information is gratefully acknowledged.

2 Compared to job offers from the private sector, the income earned in these positions is quite unattractive.

cal methodology, the ability to run statistical analysis packages (i.e., STATA, SPSS, R, or SAS), and some familiarity with a special substantial topic, such as labor economics or gender diversity. In these areas, however, representatives of the Federal Statistical Office in Germany have complained of a lack of skills and education, especially among economists who have recently left the university (see Rendtel 2008). Specifically, it is said that young economists aren't familiar with the important surveys in official statistics, that they don't know the framework of survey methodology, and have limited experience handling empirical data – for example in dealing with item nonresponse or coding errors. Sociologists, on the other hand, are regarded as better trained. They seem to profit from mandatory courses on empirical methods in surveying in their field, which are not included in the standard program of economic study.

To summarize, there is a gap created by the increased demand for young researchers with a sound knowledge of important surveys and data handling and an insufficient amount of statistical training. This observation, however, is more characteristic of university programs in economics than it is of sociology departments.

3. CAMPUS-Files

One measure taken to narrow this gap is the use of CAMPUS-Files in academic teaching. These files are created for use in statistical training. Because of the lack of controls in their use by students, the level of anonymization should be higher than in the case of SUFs. In general, they are regarded as absolutely anonymous PUFs, which restricts their power for analysis (see Zwick 2008).

At the moment (August, 2008), there are eight CAMPUS-Files offered by the Research Data Center of the Federal Statistical Office: two from the German Microcensus (1998, 2002), two taxation data files (*Lohn- und Einkommenssteuerstatistik* 1998, 2001), a file of employee and firm-level data on wages (*Gehalts- und Lohnstrukturerhebung* 2001), a file of the German subsample of the European firm-level panel data on the impact of job training, a file of social aid recipients (*Sozialhilfestatistik* 1998), and finally, a file of small and medium-sized firms on cost structure (*Kostenstrukturerhebung kleine und mittlere Unternehmen* 1999).[3] Since they allow no identification of units in the files, there is no control over what is done with the file data.

3 These files can be downloaded from the website of the Research Data Center (http://www.forschungsdaten zentrum.de/campus-file.asp).

A more restrictive use of data is offered by the Research Data Center of the German Pension Insurance (RV, *Deutsche Rentenversicherung*). In this case, the instructor must apply for a CAMPUS-File and notify each student who receives a copy of the file.[4] There are four of these files offered for teachers: two on the stock of the retired persons (2003, 2005) and two on recently retired persons (2003, 2004). For the social sciences, the German General Social Survey (ALLBUS, *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*) is also offered as a CAMPUS-File, although this title is not explicitly used.[5] Other surveys are offered for a modest fee by the GESIS Data Archive and Data Analysis for use in teaching.

It is relatively surprising that one of the most frequently analyzed data files, the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*), is not represented as a CAMPUS-File. Teachers are allowed to use a 50 percent subset of the SOEP that they must construct themselves. However, this file cannot be given away to the students, which makes it unsatisfactory as a teaching option. Moreover, the SOEP is a collection of more than one hundred flat files across six subsamples. The complexity of this data structure is overwhelming for untrained students. Thus it seems desirable to have a SOEP file that can be distributed to students and that has a simpler structure than the full SOEP.[6] Nevertheless researchers who run analyses with the SOEP need to be trained on a SOEP version that has the full complexity of a long-running household panel. A CAMPUS-File version of the SOEP would arguably present an educational tool at a level somewhere between the full complexity of the original file and that of a collection of mere analysis files to demonstrate the syntax and outcome of statistical program packages.[7]

In response to the broader range of data sources that can be analyzed now, the topics covered by CAMPUS-Files should be correspondingly enlarged. For example, the German Income and Consumption Survey (EVS, *Einkommens und Verbrauchsstichprobe)* is a basic source of poverty research. Also the German Microcensus, which has followed a continuous sampling scheme since 2005, is not represented by a CAMPUS-File, nor is it used as a rotating panel over three years.[8]

---

4 See www.RDC-rv.de

5 The ALLBUScompact Cumulation 1980-2006 covers 13 biannual cross-sectional surveys, see http://www.gesis.org/en/services/data/survey-data/allbus/

6 The complexity of the data structure is to some extent buffered by the retrieval system, the SOEPinfo meta analysis program (see http://www.diw.de/de/soep). One easy way to reduce the complexity of the full SOEP might be to put aside all subsamples with the exception of the Subsample F, which was started in 2000.

7 See for example the collection of SOEP files in STATA format used to support Kohler and Kreuter's textbook, Datenanalyse mit Stata (www.stata.com/datenanalyse/).

8 For more information, use the search tool on the website of the Federal Statistical Office, (http://www.destatis.de) for the Microcensus Panel Project (MZ-Panel).

Generally speaking, a good infrastructure should offer a CAMPUS-File for each subject area. For example, the data of the Federal Employment Agency have become a must for a labor economist, yet there is no CAMPUS-File offered by the Research Data Center of the Federal Employment Agency at the Institute for Employment Research (IAB, *Institut für Arbeitsmark- und Berufsforschung* within the BA, *Bundesagentur für Arbeit*). Alternatively, European datasets such as the EU Statistics on Income and Living Conditions (EU-SILC), a rotating panel started in 2005, would offer another good candidate for a European CAMPUS-File.

For sociologists, the European Social Survey (ESS) is an important international data source. Compared with EU-SILC, the situation is again much more advantageous. The ESS EduNet is an internet-based analysis training program developed by Norwegian Social Science Data Services. It not only provides data access but also a teaching environment.[9] Further areas that deserve more attention include health surveys and educational data.[10]

The downloadable format is very convenient for teachers and students. However, other formats may be equally attractive for the dissemination of data for seminars and projects. For example, the British Economic and Social Data Service (ESDS) offers a data sharing option that allows the teacher to distribute data to his or her students under the condition that the students are registered and have signed an agreement concerning the terms of data usage.[11] More information on this can be found on the ESDS website.

4. After Bologna: The situation of statistical education in Germany

The most important outcome of the Bologna Process is the transition from a single phase *Diplom* curriculum to a two-phase scheme with a three-year bachelor's and a two-year master's phase. Compared with the German *Diplom* and its four-year schedule, the bachelor's phase is significantly shorter.[12] However, this has given rise to competition between the individual subjects within a faculty over their representation in the shorter bachelor's framework.

9 See http://essedunet.usd.uib.no/cms/edunet/about.html

10 The large-scale Educational Panel Study (*Bildungspanel*), for example, might be a good candidate for demonstrating the difficulty of analyzing school data.

11 http://www.esds.ac.uk/ordering Data/sharing Data.asp

12 This effect is reinforced by mandatory general occupational skills training, comprised of languages, internships, or word-processing. At the Freie Universität Berlin this block of required study amounts to 30 credits, equal to the workload of a semester.

There are two models for coping with such a situation: (a) all courses are required to cut their curriculum by approximately 30 percent; and (b) a narrower selection of courses lead to a more specific BA (*Bachelor of Arts*) exam. The decision in favour of either model depends on the individual preferences and composition of local faculties. An empirical analysis of the economic curricula in various faculties was presented by Rendtel (2008). This study compared 117 BA degree programs at economic faculties in universities and universities of applied sciences (*Fachhochschulen*). In one quarter of the cases the percentage of mandatory credits to be earned from quantitative courses was found to be less than 5.5 percent! At the other extreme, a small number of some BA programs required 25 percent or even up to 30 percent of course credits from mandatory quantitative courses.

In addition to changes in course requirements, the format of the written diploma thesis, a final year project, has changed from one that would have been finished within four to six months, to one that must be completed in nine weeks.[13] Such a short time frame excludes examination topics requiring substantial empirical data analysis.

The large disparity in required quantitative credits illustrates the extreme diversity of different subjects within economic faculties. In these faculties, business administration recruits the majority of students and often has interest in subjects that do not use statistical inference or survey data. For example, in the bachelor's program in Business Administration at the Freie Universität Berlin, statistical inference is no longer a mandatory course. As a result, one may expect a large variation in the statistical skills of new BAs graduating from different economic faculties. In the case of sociology, university departments seem more homogeneous. In this discipline the role of survey data and empirical statistical analysis in the educational program seems to be well recognized.

Nonetheless, it does seem that at the moment students with a bachelor's degree are not qualified for research projects in empirical data analysis. The usual qualifications that are listed in the job descriptions correspond rather to the *Diplom* or the master's level of study. Thus, unless there is a substantial progression of students from the bachelor's into the master's level, one may predict a decrease in candidates qualified for high-level data analysis.

The heterogeneity of qualifications increases at the master's level. There is a trend toward highly specialized master's degrees. Again the diversity of master's degrees seems to be much greater in economics than in other faculties. This trend towards tailored master's degrees has given rise to highly specialized courses in the curricula, such as "Quantitative Methods in Finance." These replace statistical courses of general relevance, such as "Multivariate Analysis." Unless the master's program is geared specifically

13 12 credit points equal a total of 12 x 30 = 360 working hours. With a weekly workload of 40 hours one obtains 360/40 = 9 weeks.

toward statistics and data analysis there will be no possibility of obtaining a sound education, for example, in survey statistics.

This article therefore proposes the establishment of master's programs tailored to the needs of empirical data analysis with a special emphasis on survey data. Such a program should cover the framework of design-based statistics, i.e., sampling from a finite population with known inclusion probabilities, since most of the Research Data Center files come from surveys with informative sampling. The calibration of survey data – often simplified as "weighting," which is the standard routine in official statistics – should also be given more attention. Furthermore, the issue of nonresponse and some strategies to cope with it is an important topic for everyone who utilizes survey data. In fact, missing data not only occur as nonresponse but they also occur in evaluation studies as one missing observation in treatment-control pairs.[14] Measurement error is another important issue for everyone who analyses survey data.[15] Measurement errors overlap with survey techniques and questionnaire design. This is an area in which social scientists are well-trained but it is much less familiar to economists. Last but not least, there should be extensive training in basic skills (i.e., data management, model selection, data presentation, and interpretation). This can be supplemented, for example, with internships at the Research Data Centers or other research institutes, such as the Institute for Employment Research or the German Institute for Economic Research (DIW, *Deutsches Institut für Wirtschaftsforschung*).

One of the big statistical events in the near future, i.e. the German Census in 2011, will be a mixture of data from different sources. Such a design, which is complicated by overcoverage (*Karteileichen*) and undercoverage (*Fehlbestände*), is a methodological challenge for the Federal Statistical Office and it will be a challenge for those who analyze a SUF based on this census. Moreover, as regional counts are one of the most important issues of the census, the use of small area estimates is on the agenda. Whether it is accessed on-site or via the installation of platform for remote access, small area estimation will become a topic for data users.

However, none of these topics are the focus of a master's program in Germany. Neither has survey statistics been prioritized at the two German statistical faculties in Dortmund and Munich. Almost no graduates from the Dortmund program, of the some 1000 *Diplom* statisticians the department has produced, are working in official statistics (see Thöne and Weihs 2008). Here biometrics, computational statistics and, not least, the facilities own demand for doctoral candidates were the largest fields where the graduates were employed.

14 See Rässler (2006) for an overview of this in the context of data from the Federal Employment Agency (Bundesagentur für Arbeit).

15 For illustrative examples see Raghunathan (2006) and Durrant (2006)

There are obvious problems in terms of teaching capacity with trying to establish this type of intensive focus on surveys statistics at one site. To run a master's program like this would require the teaching power of at least three chairs in statistics who had a close affiliation to survey statistics. At the moment no German university offers such a concentration of energies in survey specialization. However, one might assemble the teaching resources residing at different universities in a joint master's program as a second-best solution. There are still problems with teleteaching from different sites, but given the technical possibilities that exist, teleteaching survey statistics at the master's level seems a feasible solution.

5. After the Master's: Vocational training and PhD programs

It is clear that the new datasets that have been generated by the Research Data Centers require some introduction for interested users to acquaint them with the potentials and risks of the dataset. In general, this type of training units last about three days and includes practical exercises with the data. The standard clients are young researchers who are at the beginning of some empirical project and/or their thesis project. Most participants have just finished their *Diplom*. The level of statistical proficiency is quite mixed. Very often researchers lack even an elementary knowledge of the design-based approach, and models beyond the linear regression model (e.g., Logit model or Loglinear models) are unknown. To my knowledge there is no systematic test of the statistical knowledge given to participants of such training courses.

The need for data training courses was recognized early on by the SOEP project, which has offered an annual training course at the German Institute for Economic Research (DIW) in Berlin since 1989. This opportunity has now been expanded by its integration into the university framework and for the two years they have organized a workshop series, SOEP@campus, in collaboration with other universities.[16] The participation is partly sponsored by the Federal Ministry of Education and Research (BMBF) as a part of the KVI process. The Research Data Center of the IAB within the BA has offered a workshop on spell data on the basis of some of its test data. Again with sponsorship from the BMBF, the Research Data Center of the Federal Statistical Office and the Data Service Center of the German Microdata Lab at GESIS have been offering workshops on newly released data files. Here the Microcensus and the Microcensus Panel have played an important part.

16 See, for example, http://www.uni-due.de/soziologie/-soepatcampus/index.php

Often data producers have established user groups that convene for regular meetings (annual or biannual), where results can be presented and discussed. The user group can be regarded as an academic network for the exchange of knowledge and experience. Therefore it can support statistical training in multiple ways.

Within the framework of statistical training, GESIS plays an important part. GESIS is a member of the Leibniz-Gemeinschaft and provides statistical education on subjects that are not routinely offered at university. Thus it supplements university education, for example, by presenting courses on latent class analysis, multilevel models or mixture models. Their "Spring Seminar" is devoted to an intensive training on special methods, usually presented in a sequence of three blocks of one week each.

The ZUMA (*Zentrum für Umfragen, Methoden und Analysen*) branch of GESIS offers workshops in different fields (new datasets, interview techniques and questionnaire design, sampling, editing, and statistical methods). The workshops are presented in a three-day format at Mannheim. The participants have to pay a moderate fee. The number of participants is limited (14–18), so waiting lists have been created. The demand is such that the average waiting list is as long as the number of participants, numbering about 400 per year. There is nothing similar to GESIS in economics. One reason is probably the greater heterogeneity of the research areas.

At the international level, there are similar bodies that offer training and statistical instruction in survey statistics and the analysis of survey data; however, their organization differs. The National Survey Research Center in the US is affiliated with the University of Michigan and involves the development, refinement, and propagation of the scientific method of survey research through teaching and training.[17] The National Center for Research Methods (NCRM) in the UK is a network of research groups, each conducting research and training in an area of social science research methods.[18] It acts under the auspices of the Economic and Social Research Council (ESRC), the British funding organization that integrates research activities in this field.[19] The network is organized according to a "hub-model," where the Southampton School of Social Science serves as the hub that connects six nodes. These nodes are located at separate universities and each specializes in certain topics or methods. The whole project runs under a four-year funding scheme.

Under the auspices of the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*) several PhD programs are offered for statistical teaching and training. However, these programs are only open to those few students who were accepted in the program. Moreover, most of these

17 See isr.umich.edu/src/

18 See www.ncrm.ac.uk

19 In Germany this council would cover the activities of DFG, the BMBF and the RatSWD.

graduate schools, such as the Mannheim Graduate School of Economic and Social Sciences or the Bremen International Graduate School of Social Sciences are integrated into the German Universities Excellence Initiative.[20] Thus they are not oriented toward broader participation like the GESIS training courses.[21]

A different approach was proposed in the DFG-funded Priority Program on Survey Methodology which was started in 2007. Here the intention was to establish survey methodology as an independent subject. For this purpose the program plans to establish a "German School of Survey Methodology."[22] The proposal incorporates international experts in survey methodology as teachers and includes a nationwide recruitment of students. This proposal is similar to the proposal for the establishment of a master's program in survey statistics.

A few comments need to be made concerning the relationship of university teaching within the bachelor's and master's scheme and those training programs that lie outside this scheme:

(1) The two levels should be adapted to each other. It is my impression that sometimes the participants of training courses lack both an elementary knowledge of statistics and experience with empirical data analysis.

(2) Quite often the motivation of students to participate in a training course is low because they cannot earn credits toward their master's degree. The credit system is very flexible, however, which makes it easy to grant credit for participation in training courses. A necessary prerequisite to this, of course, would be some kind of examination of the attendees by the trainers.

6. Conclusions and recommendations

The availability of well-educated researchers is necessary for the fruitful analysis of social and economic data. The increased data offer made possible by the establishment of the Research Data Centers has resulted in an increased demand for PhD students at the master's or *Diplom* level. Even today it is not an easy task to recruit young researchers with a sound education in the methods of data analysis who also have some practical experience in this business. Especially in economics, where we find intense competition among

20 Gess.uni-mannheim.de; www.bigsss-bremen.de

21 One may regard low admission numbers as intrinsic to excellence. However, with respect to the need of a higher number of well trained researchers this might be also regarded as a kind of luxury.

22 See www.survey-methodology.de/

the various subjects within the course of study, survey statistics has not been very successful in laying claim to a substantial proportion of the course work and training. The situation is more favorable in the sociological faculties.

This article proposes the creation of new CAMPUS-Files, free PUFs that would help foster statistical education by covering a wider range of subjects. It also advances some specific suggestions for new CAMPUS-Files along these lines. Additionally, it argues for the establishment of master's programs in survey statistics that can help increase the availability of well-trained statisticians, and provides an outline of such a master's program.

There is also a widespread network of training courses that address the needs of young researchers. These programs provide training in the introduction of new datasets as well as in non-standard analysis techniques. These training courses are organized by the Research Data Centers (i.e., the data producers), the Data Service Centers or by GESIS. Recently, there has been a greater tendency toward collaboration with universities. In order to attract students before their exam – and thus enlarging the number of applicants for research projects – one should investigate the possibility of granting academic credit for bachelor's and master's students.

The close cooperation between the SOEP group and universities is regarded as a fruitful model of this approach. Likewise, the Research Data Centers offer not only data but also support for the analysis of these data. They should be encouraged to reinforce and expand their training activities. This will not only improve statistical education in the university but will help widen the scope of official statistics from a mere data producer to an information provider.

References:

Durrant, G. (2006): Missing data Methods in Official Statistics in the United Kingdom: Some Recent Developments. ASTA Advances in Statistical Analysis 90 (4), 577-593.

Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Ed.) (2001): Wege zu einer besseren informationellen Infrastruktur. Baden-Baden.

Rässler, S. (2006): Der Einsatz von Missing Data Techniken in der Arbeitsmarktforchung des IAB. ASTA Advances in Statistical Analysis 90 (4), 527-552.

Raghunathan, T. (2006): Combining Information from Multiple Surveys for Assessing Health Disparities. ASTA Advances in Statistical Analysis 90 (4), 515-526.

Rendtel, U. (2008): Statistikausbildung und Amtliche Statistik. Kritik und Perspektiven. ASTA Wirtschafts- und Sozialstatistisches Archiv 2 (1/2), 5-20.

Thöne, M. and Weihs, C. (2008): Vielseitig und gefragt. Absolventen und Absolventinnen der Dortmunder Studiengangs Statistik. ASTA Wirtschafts- und Sozialstatistisches Archiv 2 (1/2), 75-92.

Zwick, M. (2008): Campus-Files – Kostenfreie Public-Use-Files für die Lehre. ASTA Wirtschafts- und Sozialstatistisches Archiv 2 (1/2), 175-188.

5. e- Infrastructure for the Social Sciences

Ekkehard Mochmann

Contact:

Ekkehard Mochmann
e-mail: E.Mochmann[at]web.de

Abstract

When the term “e-Science” became popular, it frequently was referred to as “enhanced science” or “electronic science.” More revealing, however, is the definition, “e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it” (Taylor 2001). The question arises to what extent can the social sciences profit from recent developments in e-Science infrastructure?

While available computing, storage, and network capacities have so far been able to accommodate and access social science databases, new capacities and technologies will support new types of research, such as linking and analyzing transactional or audio-visual data. Increasingly, collaborative work among researchers in distributed networks is efficiently supported by information technology and new resources have been made available for e-learning. Whether these new developments will be transformative or merely helpful will very much depend on whether their full potential is recognized and creatively integrated into new research designs by theoretically innovative scientists.

Progress in e-Science was closely linked to the vision of the Grid as “a software infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources” with virtually unlimited computing capacities (Foster et al. 2000). In the social sciences there has been considerable progress made in the use of modern information technologies (IT) for multilingual access to virtual distributed research databases across Europe and beyond (e.g., NESSTAR, CESSDA-Portal), data portals for access to statistical offices, and for linking access to data, literature, project, expert, and other databases (e.g., Digital Libraries, VASCODA/SOWIPORT). Whether future developments will require Grid enabling of social science databases or can be further developed using WEB 2.0 support is currently an open question. The challenges that must be met are the need for seamless integration and interoperability of databases, a requirement further mandated by internationalization and transdisciplinary research. This goes along with the need for standards and harmonization of data and metadata.

Progress powered by e-Infrastructure is, among other things, dependent on both regulatory frameworks and human capital well trained in both data science and research methods. It is also dependent on a sufficient critical mass of the institutional infrastructure to efficiently support a dynamic research community that wants to “take the lead without catching up.”

1. Introduction

Are advances in socio-economic research driven by data or technology? Claims in one direction and inspired deliberations pondering these alternatives are not new. While Norman Nie asserted without reservation “that all science is fundamentally data driven” (Nie 1989: 2) others have argued “that progress in science rather depends on formal modelling" (Rockwell 1999: 157). More recently "methodological and substantive rigour"[1] have been emphasized as necessary preconditions for the creation of reliable sources of knowledge about social change. Both information technology and the social science database have developed remarkably over the past few decades – from poverty of data to the rapidly expanding production of all kinds of empirical evidence beyond the survey and statistical microdata. These now include, for example, electronic texts, event databases, videos, geo-information, and new kinds of data, as in the case of transaction data (Lane 2010; Engel 2010) or biomarkers (Schnell 2010). Access to comprehensive databases and advanced data analysis increasingly allow modeling of complex social processes.

To efficiently support future empirical research

"[t]he present major task is [...] to create pan-European infrastructural systems that are needed by the social sciences [...] to utilise the vast amount of data and information that already exist or should be generated in Europe. Today the social sciences [...] are hampered by the fragmentation of the scientific information space. Data, information and knowledge are scattered in space and divided by language, cultural, economic, legal and institutional barriers" (ESFRI 2006).

2. e-Science, e-Social Science, the Grid and Web 2.0

Though there has already been evident progress fuelled by new kinds of measurement, expanding databases, and technological support for the past few decades, new and revolutionary systematic approaches can now be used to analyze research challenges. Based on the results of the resulting analyses, comprehensive technological infrastructures can be implemented to facilitate innovative research. These "e-Science" approaches were initially referred to as "enhanced science" or "electronic science." More revealing, however, is the definition "e-Science is about global collaboration in key areas of science and the next generation of infrastructure that will enable it" (Taylor 1999). Basically, e-Social Science follows these ideas, with emphasis on providing advanced IT services to "enable" social research. The National Centre for e-Social Science at Manchester (NCeSS) states:

"e-Social Science is a term which encompasses technological developments and approaches within Social Science. We are working with Social Scientists and Computer Scientists on tools and research which Social Scientists can take and use to help their research. These tools might either allow a Social Science Researcher to conduct new research or else conduct research more quickly These tools can be used across a variety of

1 http://www.europeansocialsurvey.org/

Social Science domains. […] Within NCeSS, we refer to the 'e' in e-Social Science as 'enabling'."[2]

Progress in e-Science was closely linked to the vision of the Grid as "a software infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources and virtually unlimited computing capacities" (Foster et al. 2000). As such, it was based on multi gigabit broad band width fiber cables connecting distributed and loosely coupled computing resources, using open standards in the Grid. In coordination with the National Research and Educational Networks (NRENs), they would provide a globe-spanning net with virtually unlimited computing capacity, intelligent middleware to support interoperability of network services, and control of access and authentication. To support information handling and support for knowledge processing within the e-scientific process, future developments point toward the Semantic Grid (De Roure et al. 2003: 9).

The Enabling Grids for E-SciencE (EGEE) project is a prominent and globally expansive example of the impetus to build a secure, reliable, and robust Grid infrastructure with a light-weight middleware solution intended to be used by many different scientific disciplines. It is built on the EU Research Network (GÉANT), and exploits Grid expertise generated by many EU, national, and international Grid projects, including the EU Data Grid.[3] Just to show the new dimensions: at present, it consists of approximately 300 sites in 50 countries and gives its 10,000 users access to 80,000 CPU[4] cores, 24 hours a day, 7 days a week. This project came to the conclusion that the state of computer and networking technology today facilitates extensive computing grids that integrate geographically distributed computer clusters, instruments, scientific communities, and large data storage facilities. The resulting benefits include a large increase in the peak capacity, the total computing available, and data management power for various scientific projects, in a secure environment.[5] Critics, however, point to the fact that these new developments cannot be used outside high energy physics so far.

The Grid idea followed the computer scientists' blueprint for a perfectly designed distributed infrastructure. Lessons learned from early developments emphasize that it is very important to have application scientists collaborate closely with computer scientists.

"Successful projects were mostly application and user driven, with a focus on the development of standard and commodity components, open source, and results easy to understand and to use" (Gentzsch 2007: 17).

2 http://www.ncess.ac.uk/about_eSS/
3 http://eu-datagrid.web.cern.ch/eu-datagrid/
4 Central Processing Unit.
5 http://www.eu-egee.org

It is significant that the German Grid initiative (D-Grid)[6], which started in 2005 with six science projects, now also includes Text Grid[7] from the humanities but none from the social sciences. Over the past few years more than ten new projects from the sciences have been added. In this area, at least, the social sciences certainly do not belong among those who adopted technology early. This pattern can be observed in most other countries, with the exception of the UK and the US, where the social science communities have made particular efforts to boost their e-Infrastructure. In this context it may also be worth noting that the first attempt to support retrieval of data by machine was actually conceived in the context of a social science project already described in 1964 (Scheuch and Stone 1964). Ideas for researcher dialogue with interactive data analysis and retrieval systems date back to 1972 (Scheuch and Mochmann 1972: 154f). With respect to transnational data infrastructure, the Council of European Social Science Data Archives (CESSDA) is studying the feasibility of Grid enabling. This investigation examines current developments and applications in Grid technologies in order to find efficient and sustainable ways for the implementation of a cyberinfrastructure for the social sciences and humanities and to identify the issues for implementing Grid technology.

Instead of an enthusiastic uptake of Grid technologies, a number of initiatives have followed a bottom-up approach in collaborative systems development, for example, creating access to virtually distributed databases using the World Wide Web in a more sophisticated way. These new trends in the use of WWW technology to enhance collaboration as well as information and data sharing are referred to as Web 2.0 technologies. They are still based on the so far known World Wide Web specifications. Results of these developments are possibly less perfect than those designed for Grid applications, but they are facilitated by cooperative approaches within the science community and they take usually much less time to implement.

3. Social research infrastructure, e-infrastructure, cyberinfrastructure

The social sciences have a long record of infrastructure development in terms of service institutions, databases, data laboratories, and researcher networks in the field of international comparative research (Scheuch 2003). Thus, it was no surprise that the social scientists pointed to the need to distinguish the preexisting infrastructure from the emerging IT-based infrastructure (Sere-

6 http://www.d-grid.de/
7 http://www.textgrid.de/

nate 2003b). The e-infrastructure concept was thus proposed in 2003 to coin a term for the development of the next generation of transnational Information and Communication Technology (ICT) research infrastructure in Europe:

> "e-Infrastructure refers to this new research environment in which all researchers – whether working in the context of their home institutions or in national or multinational scientific initiatives – have shared access to unique or distributed scientific facilities (including data, instruments, computing and communications), regardless of their type and location in the world" (European Commission IST 2005).

At the same time, the National Science Foundation Blue-Ribbon Advisory Panel identified similar objectives for what they called "Cyberinfrastructure":

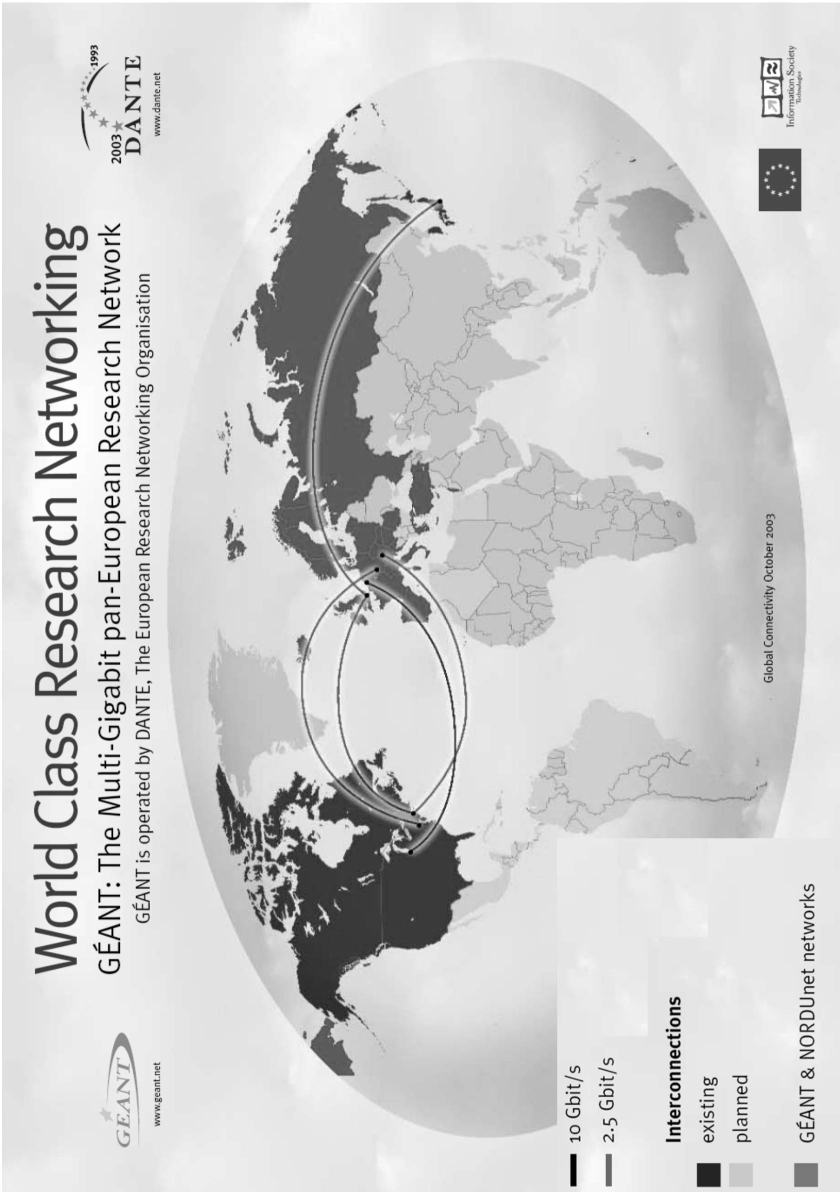
> "We envision an environment in which raw data and recent results are easily shared, not just within a research group or institution but also between scientific disciplines and locations. There is an exciting opportunity to share insights, software, and knowledge, to reduce wasteful re-creation and repetition. Key applications and software that are used to analyze and simulate phenomena in one field can be utilized broadly. This will only take place if all share standards and underlying technical infrastructures" (Atkins et al. 2003: 12).

Cyberinfrastructure is defined in relation to already known infrastructures:

> "Although good infrastructure is often taken for granted and noticed only when it stops functioning, it is among the most complex and expensive things that society creates. The newer term cyber-infrastructure refers to infrastructure based upon distributed computer, information and communication technology. If infrastructure is required for an industrial economy, then we could say that cyberinfrastructure is required for a knowledge economy" (Atkins et al. 2003: 5).

In Europe, the provision of network services to research and education is organized at three levels: the Local Area Network to which the end-user is connected, the national infrastructure provided by the National Research and Education Network (NREN), and the pan-European level provided by GÉANT.

GÉANT currently interconnects the national research and education networks (NRENs) from all over Europe, including Russia. In terms of geographical coverage, technology used, and services made available, GÉANT considers itself the number one research network in the world, which attracts requests for interconnection from all over the world. Under the GÉANT2 project it has grown to include more than 100 partners already. This is much more than the social sciences need so far, but it gains in importance when we think about the potential for International Data Federations to support continuous global comparative and transdisciplinary research. While the technical backbone of the network is in place, many application tools, standards, and content with rich metadata have to be developed in order to make full use of these technologies.


World Class Research Networking
GÉANT: The Multi-Gigabit pan-European Research Network
GÉANT is operated by DANTE, The European Research Networking Organisation
GÉANT
www.geant.net
2003 DANTE 1993
www.dante.net
10 Gbit/s
2.5 Gbit/s
Interconnections
existing
planned
GÉANT & NORDUnet networks
Global Connectivity October 2003
Information Society Technologies

4. Data infrastructure needs of the social sciences

(Major results of the SERENATE project and the AVROSS study)

Exciting visions of the future potential of new technologies like to travel with appealing descriptions of how it is actually implemented in working environments. Closer examination, however, frequently shows that, in practice, services that are needed by end-users on a continuous basis are often far from satisfactory. The economic potential to implement new technologies, the level of expertise in different societies that is available to support these technologies and to adjust them to the specific needs of their user communities, as well as data management and methodological skills vary from country to country.

Needs, challenges, and obstacles in relation to these new technologies have been analyzed by the Study into European Research and Education Networking as Targeted by *e*Europe (SERENATE). Security features were highlighted by a large number of the respondents who deal with sensitive data or even medical images. Another critical element is mobile access to network services – including both home access for researchers, particularly for non-laboratory based research such as humanities and social sciences, and access when abroad. As a consequence of these usage patterns, the deployment of "Authentication, Authorization and Accounting" (AAA) services across the various networks was stipulated to provide the necessary controls on access. The report from the final workshop also noted that access to a rich variety of data from many sources is possible and identified the potential for software to support collaborative working, the sharing of databases, and data integration at many levels. Finally, the networks offer the means to include the "future generation of scientists in schools" (Serenate 2003b: 14).

In the spirit of e-Science approaches to systematically examining options and challenges for enhancing scientific research, SERENATE includes some tough observations on contextual requirements into its findings:

"We have learned that many people – national and European politicians, ministries and agencies in the national governments, the European Commission, telecoms vendors, equipment vendors, various service suppliers, local and regional authorities, universities and user communities all have to be mobilized, and to move in the same direction, if we are to make progress. If we do not make plans to maintain and even improve the situation over the next 5–10 years, then the sustained pace of technical, organisational and political change will inevitably lead to rapid decay" (Serenate 2003a).

Analyses based on the Accelerating Transition to Virtual Research Organisation in Social Science (AVROSS) study concluded that efforts by the US and the UK appear to be an exception, since no other European country has adopted an initiative that promotes e-infrastructure uptake by the social sciences or the humanities. At the same time, the European Strategy Forum on Research Infrastructures (ESFRI) has recognized the importance of including these domains of science in the ESFRI Roadmap report. This foundational report identified three long-term strategic goals for Social Science and Humanities (SSH) research infrastructures: comparative data and modeling, data integration and language tools, and coordination (European Strategy Forum on Research Infrastructures 2006). These aims create a potential for researchers in SSH who want to develop or use e-infrastructure.

5. Status quo and best practice examples from the social sciences

For the most part, social scientists do not see a particular need to use the Grid technology for e-Social Science developments, since most of their data and computation needs could be handled by the existing Internet capacities. There are numerous Internet solutions for access to specific collections, even with local AAA procedures. While many of them grant sufficient user support for their constituency, it is rarely possible for them to provide interoperability of databases and metadata (see the report on Metadata in this publication) or world wide networked access. There are, however, a few remarkable examples for transnational data access in virtually distributed databases.

Building on extensive experience in international data transfer, the Council of European Social Science Data Archives (CESSDA) worked towards networked solutions that ideally would allow interested researchers to access the holdings of member archives from any point in the world. This is operational now as the CESSDA Portal, providing seamless access to datasets from currently twelve social science data archives across Europe.[8] Among other things, it includes prominent reference studies from international comparative research, such as the European Social Survey, Eurobarometers, the International Social Survey Programme, and the European Values Studies.[9] The Data Portal builds on the work of the EU-funded MADIERA project.[10] All content is based on the Data Documentation Initiative (DDI) specifications for documenting datasets including relevant metadata.[11] Multilingual functionality is supported by the European Language Social Science Thesaurus (ELSST) and the NESSTAR technology provides functionality to the user for browsing and analyzing data.[12] The software consists of tools

8 http://www.cessda.org/
9 http://www.cessda.org/accessing/catalogue/
10 http://www.madiera.net/
11 http://www.ddialliance.org/org/
12 http://www.nesstar.com/

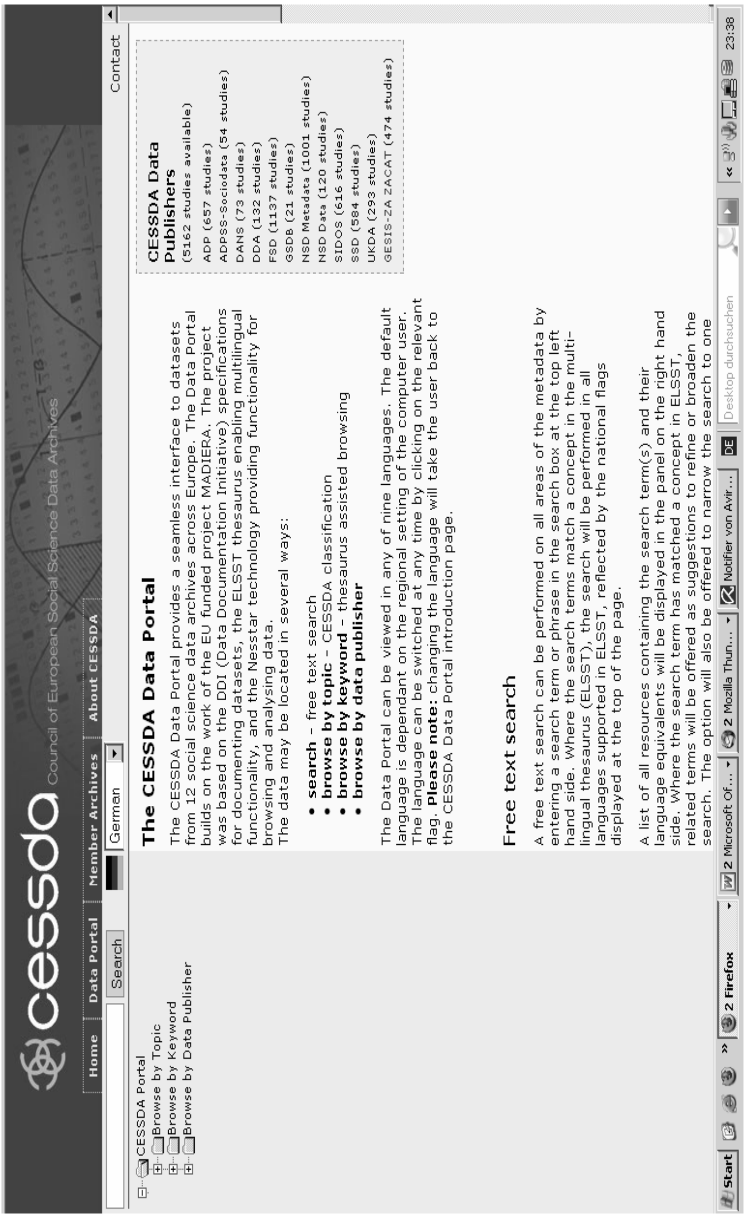
which enable data providers to disseminate their data on the Web. NESSTAR handles survey data and multi-dimensional tables as well as text resources.

A recent user survey conducted by the Institute for Social Research (ISR, Ann Arbor, Michigan) in cooperation with the Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*) under the auspices of the European Science Foundation, with more than 2000 users, shows that there is a high level of satisfaction with these new technologies. These are efficiently supporting simultaneous data access to thousands of studies in a virtual distributed network, frequently including the option to check the measurement instrument, get methodological and technical background information, and then proceed to data analysis in the same session. As a precondition for taking advantage of this functionality on the output side, there are nontrivial investments on the input side. To close the knowledge gap between principle investigators – who designed the study and followed the steps through fieldwork and data management – up to the provision of analysis-ready files, a lot of methodological and technical details covering the research process up to that point have to be communicated to enable further informed analysis.[13]

A frequently discussed area of development is the integration of data, literature, project documentation, and expert databases. One development in this direction is SOWIPORT, which includes among other things references to social science literature and data resources offered by different providers.[14] The Dutch Data Archiving and Networked Services (DANS) have started to store data for long-term preservation and access in the Grid.[15]

There are several other technological developments that have been successfully applied to social science data service for larger international user communities. These include, for example, the Data Service for the European Social Survey (ESS);[16] the ZACAT Data Portal of GESIS,[17] which provides access to most of the continuous international survey programs; the JD-Systems Survey Explorer;[18] or Survey Data and Analysis (SDA), a set of programs for the documentation and web-based analysis of survey data[19] that includes, for example, the German General Social Survey (ALLBUS, *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*) and the American National Election Study (ANES).

13 MetaDater project: http://www.metadater.org/
14 http://www.sowiport.de
15 http://www.dans.knaw.nl/en
16 http://www.europeansocialsurvey.de/)
17 http://zacat.gesis.org
18 http://www.jdcomunicacion.com/ ISSPSpain.asp
19 http://sda.berkeley.edu/


cessda Council of European Social Science Data Archives
Home
Data Portal
Member Archives
About CESSDA
Search
German
Contact
CESSDA Portal
Browse by Topic
Browse by Keyword
Browse by Data Publisher
The CESSDA Data Portal
The CESSDA Data Portal provides a seamless interface to datasets from 12 social science data archives across Europe. The Data Portal builds on the work of the EU funded project MADIERA. The project was based on the DDI (Data Documentation Initiative) specifications for documenting datasets, the ELSST thesaurus enabling multilingual functionality, and the Nesstar technology providing functionality for browsing and analysing data.
The data may be located in several ways:
search - free text search
browse by topic - CESSDA classification
browse by keyword - thesaurus assisted browsing
browse by data publisher
The Data Portal can be viewed in any of nine languages. The default language is dependant on the regional setting of the computer user. The language can be switched at any time by clicking on the relevant flag. Please note: changing the language will take the user back to the CESSDA Data Portal introduction page.
Free text search
A free text search can be performed on all areas of the metadata by entering a search term or phrase in the search box at the top left hand side. Where the search terms match a concept in the multi-lingual thesaurus (ELSST), the search will be performed in all languages supported in ELSST, reflected by the national flags displayed at the top of the page.
A list of all resources containing the search term(s) and their language equivalents will be displayed in the panel on the right hand side. Where the search term has matched a concept in ELSST, related terms will be offered as suggestions to refine or broaden the search. The option will also be offered to narrow the search to one
CESSDA Data Publishers
(5162 studies available)
ADP (657 studies)
ADPSS-Sociodata (54 studies)
DANS (73 studies)
DDA (132 studies)
FSD (1137 studies)
GSDB (21 studies)
NSD Metadata (1001 studies)
NSD Data (120 studies)
SIDOS (616 studies)
SSD (584 studies)
UKDA (293 studies)
GESIS-ZA ZACAT (474 studies)
Start
2 Firefox
2 Microsoft Of...
2 Mozilla Thun...
Notifier von Avir...
DE
Desktop durchsuchen
23:38

A more recent development is the Dataverse Network supported by the Institute for Quantitative Social Science (IQSS) of Harvard University.[20]

"The Dataverse project aimed to solve some of the political and sociological problems of data sharing via technological means, with the result intended to benefit both the scientific community and the sometimes apparently contradictory goals of individual researchers" (King 2007: 1).

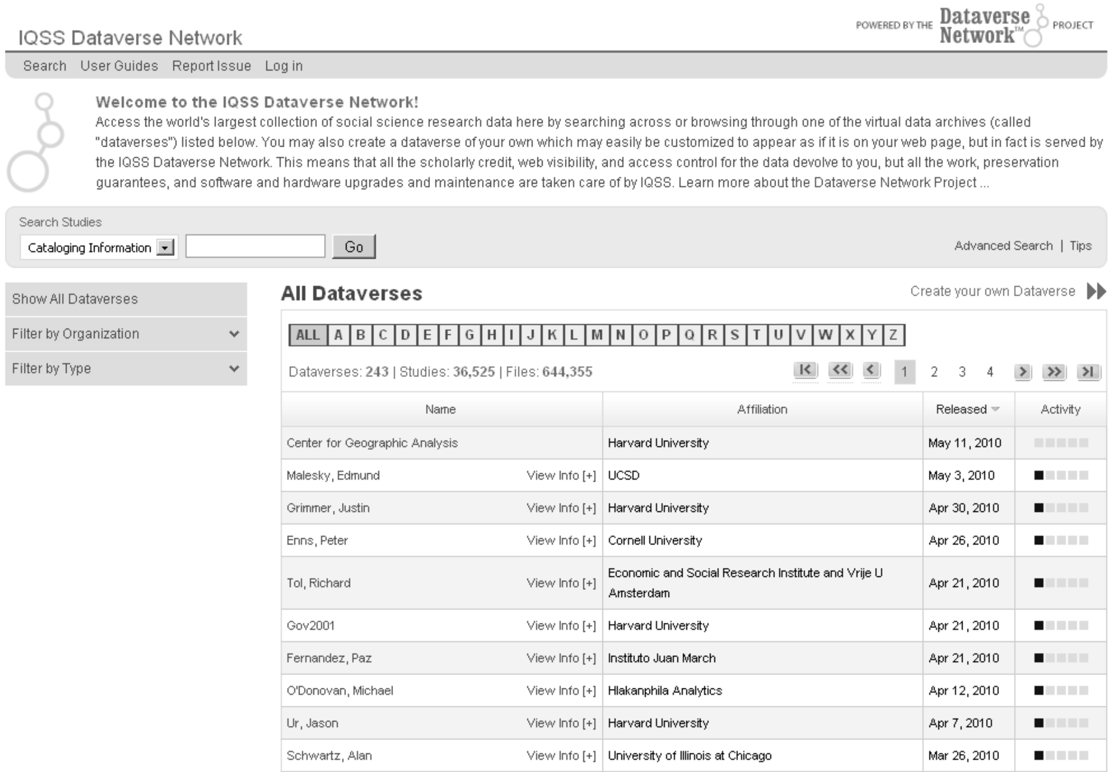
Dataverse provides open source software to host Dataverse networks at larger institutes or to create individual "dataverses" as archives of individual owners that may be just for long-term archiving and analysis, or for access by other users over the Internet. In this way, individually created databases and trusted archives can be networked as the Networks homepage depicts.[21]

As software is only part of the solution, IQSS also provides citation standards for the content to be stored. The digital library services of each dataverse include data archiving, preservation formatting, cataloguing, data citation, searching, conversion, subsetting, online statistical analysis, and dissemination.

6. Conclusions

As we can observe already today, a comprehensive infrastructure based on advanced data communications, computing, and information systems are extremely supportive for conducting high-quality research. They are indispensable for progress, which so far has been unlikely to be achieved in many fields of research. Outstanding examples are the mapping of the human genome and the discovery of new elementary particles, which were facilitated by advanced computational, data storage, and network technologies. Being in touch with widely dispersed research communities, collaborative working and data access in globe-spanning comparative social survey programs that include over 40 countries are already strongly supported by these new technologies. The rapidly growing social science database, including methodologically controlled databases and new kinds of data with related metadata, increasingly leans toward making data linkages across topical domains. This modeling of complex social processes – which may require collaboration in dispersed researcher networks and large-scale data access and computation resources – can be supported more effectively than ever before. One example for creating that kind of research environment is a design study, "Provi-

20 http://dvn.iq.harvard.edu/dvn/
21 http://dvn.iq.harvard.edu/dvn/


IQSS Dataverse Network
POWERED BY THE Dataverse Network PROJECT
Search User Guides Report Issue Log in
Welcome to the IQSS Dataverse Network!
Access the world's largest collection of social science research data here by searching across or browsing through one of the virtual data archives (called "dataverses") listed below. You may also create a dataverse of your own which may easily be customized to appear as if it is on your web page, but in fact is served by the IQSS Dataverse Network. This means that all the scholarly credit, web visibility, and access control for the data devolve to you, but all the work, preservation guarantees, and software and hardware upgrades and maintenance are taken care of by IQSS. Learn more about the Dataverse Network Project ...
Search Studies
Cataloging Information
Go
Advanced Search | Tips
Show All Dataverses
Filter by Organization
Filter by Type
All Dataverses
Create your own Dataverse
ALL A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
Dataverses: 243 | Studies: 36,525 | Files: 644,355
1 2 3 4
Name
Affiliation
Released
Activity
Center for Geographic Analysis
Harvard University
May 11, 2010
Malesky, Edmund
View Info [+]
UCSD
May 3, 2010
Grimmer, Justin
View Info [+]
Harvard University
Apr 30, 2010
Enns, Peter
View Info [+]
Cornell University
Apr 26, 2010
Tol, Richard
View Info [+]
Economic and Social Research Institute and Vrije U Amsterdam
Apr 21, 2010
Gov2001
View Info [+]
Harvard University
Apr 21, 2010
Fernandez, Paz
View Info [+]
Instituto Juan March
Apr 21, 2010
O'Donovan, Michael
View Info [+]
Hlakanphila Analytics
Apr 12, 2010
Ur, Jason
View Info [+]
Harvard University
Apr 7, 2010
Schwartz, Alan
View Info [+]
University of Illinois at Chicago
Mar 26, 2010

ding an Infrastructure for Research on Electoral Democracy in the European Union" (PIREDEU),[22] which brings together all kinds of empirical evidence ranging from survey data to aggregate statistics to party manifestoes on a European level, while the Comparative Study of Electoral Systems is taking a global approach.[23]

The technical backbone and the e-Infrastructure for advanced Grid applications are in place and are currently in use by many international and national science communities. In principle, as well as in practice, there are technological solutions to provide researchers with computational resources on demand; the capability to share complex, heterogeneous and widely distributed data repositories; and the means to enable researchers to collaborate easily and effectively with colleagues around the world. These functionalities, which are available now, have been part of the e-Research Vision that took shape at the beginning of this millennium. This gives an indication of the incredible speed at which these new technologies develop and are adopted in some disciplines.

By and large, the social sciences have so far opted for Web 2.0 solutions. The appeal of Web 2.0 solutions lies in the ease of "ready to use applications." So far, they seem powerful enough to support most data access and analysis needs in domains. This is currently not the case with sensitive microdata from statistical offices and with panel data. Research is underway to integrate disclosure procedures into data access and analysis systems, which pose particular data protection problems. With increasing data availability and research crossing traditional disciplinary boundaries on a global scale, new technologies for large-scale data access and high speed computing may be required.

It is up to each scientific community to assess its specific needs and to decide at what speed it wants to move. Sometimes there are advantages to being a latecomer in adopting new technologies, as many detours may be avoided (Schroeder et al. 2007). Nevertheless, it is obvious that ground laying work needs to be done. A combination of methodological and technical expertise is required to adopt or design and implement the new infrastructures. As has been emphasized in almost all prominent studies quoted, the combination of experts from the social research community working closely with IT specialists is required. Practical experiences from many international projects prove, however, that it is difficult to find the required expertise for limited project lifetimes and that it is even more difficult to keep the additional expertise acquired during the project accessible for further research and development. So, needs assessments, user community studies, and capacity building at the interface of social research methodology and computer science are a prerequisite for viable and sustainable developments.

22 http://www.piredeu.eu/

23 http://www.umich.edu/~cses/

It may be a healthy step to combine future research methodology curricula with modules of what might be called “data science,” which is about data structures, data management, access and interoperability of databases.

The Open Access Initiatives (e.g., the Berlin Declaration 2003) and the OECD declaration on open access to publicly financed data (OECD 2004) certainly support the creation of a culture of data sharing and ease of access to information and data, including metadata. The challenges and development needs in e-Infrastructure are beyond what a normal research institute can afford to invest in order to keep up on its own with the developments and to cover its long-term needs. Forming alliances or multilateral institutional cooperation have been solutions of academic self organization so far. The National Center for e-Social Science in the UK is an example of how to create a competence center designed to serve the social science community in this respect.

Whether future developments will need Grid enabling of social science databases or can be adequately developed using WEB 2.0 support is currently an open question. The challenge here is the seamless integration and interoperability of databases, a requirement that is also stipulated by internationalization and transdisciplinary research.

Progress in e-Infrastructure is also dependent on regulatory frameworks (Hahlen 2010) and data policies (e.g., NERC Data Policy 2002). The best technical solutions may provide some routines and intelligent algorisms to control access to sensitive data. International access, which is technically possible, can be out of question if statistical confidentially or statistics law prohibit outside use. Last but not least, the organizational infrastructure requires sufficient critical mass in terms of expertise, networking capacities, and sustainable resources to efficiently support a research community that wants to “take the lead without catching up.”

7. Recommendations

The present assessment of socio-economic databases does show, once again, that impressive amounts of data are available in many fields of research. It is not surprising, however, that the database as it exists is rather scattered, not well-integrated, and does not lean easily to intranational or international comparative research or even the combination of different sources for analyses with transdisciplinary perspective. Apart from harmonizing data on the measurement level, nontrivial investment is required to get databases organized and to get the metadata in place.

For the most part, social scientists do not see a particular need to use the Grid technology in the development of e-Social Science, since most of their data and computation needs can be handled by the existing Internet capacities. Numerous Internet solutions exist for access to specific collections, even with local AAA procedures in place. While many of them provide sufficient user support for their constituency, the interoperability of databases and metadata (see the report on Metadata in this publication), as well as world wide networked access are rarely possible. There are, however, a few noteworthy examples of transnational data access in virtually distributed databases.

7.1 Data policy and strategic plans for research data management

Some scientific communities have formulated comprehensive *strategic plans* or even published explicit *data policies*. It might provide a good starting point in the social sciences to assess needs in an international context and to identify challenges, drivers, and impediments for the development of a future German e-Infrastructure for the social sciences, which would also provide interfaces to and interoperability with leading international networks.

7.2 Needs assessment and framework conditions

Like other countries, Germany has the technical infrastructure for modern data services in place. Whether there is the *need* and whether the *regulatory framework* conditions will permit the installation of an integrated *German Data Net* has yet to be determined. This could best be done by a *working group* that includes experts on methodological, legal, and technical issues.

7.3 Measurement and metadata standards

Good documentation is a decisive factor that will impact the potential of future data analyses. The Association of German Market Researchers (ADM, *Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute*), the Association of Social Science Institutes (ASI, *Arbeitsgemeinschaft Sozialwissenschaftlicher Institute*), and the Federal Statistical Office agreed on *minimal standards for demographic variables* (*Standarddemographie*) long ago to allow for better comparability of measurements across the three sectors. Likewise, there are standards for metadata that would allow easier identification of and access to data that is related to the concepts central to the respective research questions. It might be advantageous to follow a single metadata standard, but this is not absolutely required. Nevertheless, to follow at least some metadata standard is a precondition for the development of interoperability at a later stage. DDI is being used by several institutes in

Germany already. Working towards wider consensus on adopting metadata standards and agreeing on interfaces is one milestone along the infrastructure highway.

7.4 *Best practice in data management and documentation*

Efficient database management will require the close cooperation of researcher networks and data services**.** Best practices have to be communicated to implement metadata capture already at the point of data collection and to cover the whole life cycle from research design via data collection to publication and reuse.

7.5 *Capacity building*

Training of researchers in best practices of supplying all relevant information from the research process (e.g., the *OAIS model*) and training of data professionals should be oriented toward what could be named "data science" in future curricula. Substantial investment in sound databases needs to be based on the highest level of methodological, data management, and IT expertise. This is hard to find on the labor market in this combination and equally difficult to combine in research teams, simply because there is a serious lack of professionally trained people in this field. Data management, documentation, and access could become one module of "*data science*" in studies of social science research methods. There is a huge market and demand for these skills – such as social and market research, insurance companies, media centers and media archives, data providers, etc.

7.6 *Research funding should also cover data management*

It is not always easy to assess the relevance of data for future needs. Nevertheless, a vast uninspired *omnium gatherum* should be avoided. At least reference studies and data collections that allow comparability over time or space should be properly documented for further use. This is a nontrivial and labor intensive phase in the research process.

Frequently, the data management required to create high-quality databases demands a lot of methodological and technical expertise. This should be acknowledged by funding authorities and evaluation committees, which tend to honor the analyses but not the investment in preparing the data for it. So future funding of data collection should include *a line on data management and documentation.* Likewise *evaluation criteria* should also include whether databases have been created following *methodological and technical best practice*.

7.7 Technical developments

The question of whether current institution-specific data portals, remote access to individual databases, product catalogues in integrated literature, and data portals like SOWIPORT or networked solutions with central data repositories, such as the DRIVER development on global level, or even Data Grid solutions are the needs of the future has to be assessed with a mid-term and a long-term perspective.

7.8 e- Infrastructure competence center for the social sciences

The Open Access Initiatives (e.g., the Berlin Declaration 2003) and the OECD declaration on open access to publicly financed data (OECD 2004) certainly support the creation of a culture of data sharing and ease of access to information and data, including metadata. The challenges and development needs in e-Infrastructure are beyond what a normal research institute can afford to invest in order to keep up on its own with the developments and to cover its long-term needs. The formation of alliances or multilateral institutional cooperation agreements have been solutions of academic self-organization to date. The National Center for e-Social Science in the UK is an example of creating a competence center designed to serve the social science community in this respect.

References:

Atkins, D.E. et al. (2003): Revolutionizing Science and Engineering through Cyber-infrastructure. Report of the National Science Foundation. Blue-Ribbon Advisory Panel on Cyberinfrastructure. January 2003, 5 and 12.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, Max-Planck-Gesellschaft. Berlin 2003.

Berman, F. and Brady, H. (2005): Final Report: NSF SBE-CISE Workshop on Cyber-infrastructure and the Social Sciences.

De Roure, D./Jennings, N.R. and Shadbolt, N.R. (2003): The Semantic Grid: A Future e-Science Infrastructure. In: Berman, F./Fox, G. and Hey, T. (Eds.): Grid Computing – Making the Global Infrastructure a Reality. Chichester.

Engel, B. (2010): Transaction Data: Commercial Transaction Surveys and Test Market Data. [In this publication].

European Strategy Forum on Research Infrastructures (ESFRI) (2006): European Roadmap for Research Infrastructures. Report 2006.

European Commission IST (2005): Building the e-infrastructure. Computer and network infrastructures for research and education in Europe. A pocket guide to the activities of the Unit. ftp://ftp.cordis.europa.eu/pub/ist/docs/rn/leaflet-2005 en.pdf. [Last visited: 02/23/2010].

Foster, I. and Kesselmann, C. (1999): The Grid: Blueprint for a New Computing Infrastructure San Francisco.

Foster, I./Kesselmann, C. and Tuecke, S. (2000): The Anatomy of the Grid. http://www.globus.org/alliance/publications/papers/anatomy.pdf. [Last visited: 02/23/2010].

Gentzsch, Wolfgang (2007): Grid Initiatives: Lessons Learned and Recommendations version 2.0. http://www.ogf.org/UnderstandingGrids/documents/Grid_Initiatives _July_12_2007.pdf

Hahlen, J. (2010): Providing a Permanent Institutional Guarantee for the German Information Infrastructure [In this publication].

Hey, T. [The Corporate Vice President for Technical Computing, Microsoft] (2008): "eScience, Semantic Computing and the Cloud: Towards a Smart Cyber-infrastructure for eScience" auf der 38. Jahrestagung der Gesellschaft für Informatik: Informatik 2008 – Beherrschbare Systeme – dank Informatik, 08.-13.09. Munich.

King, G. (2007): An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. Sociological Methods and Research 32 (2), 173-199.

Lane, J. (2010): Administrative transaction data. [In this publication].

Meulemann, H. (Ed.) (2004): Erwin K. Scheuch. Infrastrukturen für die sozialwissenschaftliche Forschung. Gesammelte Aufsätze. Bonn.

Mochmann, E. (2001): Infrastruktur für die Komparative Sozialforschung in Europa. In: Hasebrink, U. and Matzen, Ch. (Eds.): Forschungsgegenstand Öffentliche Kommunikation. Funktionen, Aufgaben und Strukturen der Medienforschung. Symposien des Hans-Bredow-Instituts, Band 20. Baden-Baden/Hamburg.

Natural Environmental Research Council (NERC) (2002): Data Policy Handbook Version 2.2, 2002.

Neuroth, H./Kerzel, M. and Gentzsch, W. (Eds.) (2007): Die D-Grid Initiative. Göttingen.
Nie, N. (1989): Model vs. data driven science and the role of the ICPSR in the progress of the social sciences. Adress delivered to the 27th Annual Conference of Official Representatives of the ICPSR. Ann Arbor.
OECD (Ed.) (2004): Science, Technology and Innovation for the 21st Century. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29.-30. January, Paris – Final Communique.
Rockwell, R.C. (1999): Data and statistics: empirical bases for the social sciences. In: Kazangicil, A. and Makinson, D. (Eds.): World Social Science Report. Paris.
Scheuch, E.K. (2003): History and visions in the development of data services for the social sciences. International Social Science Journal 177, 385-399.
Scheuch, E.K. and Mochmann, E. (1972): Sozialforschung. In: Merten, P. (Ed.): Angewandte Informatik. Berlin/New York.
Scheuch, E.K. and Stone, P.J. (1964): The General Inquirer Approach to an International Retrieval System for Survey Archives. American Behavioural Scientist 8 (10), 23-28.
Schnell, R. (2009): Biological Variables in Social Surveys. [In this publication].
Schroeder, R. den Besten, M. and Fry, J. (2007): Catching Up or Latecomer Advantage? Lessons from e-Research Strategies in Germany, in the UK and Beyond. Paper presented at the German e-Science Conference 2007. Baden-Baden.
Serenate (2003a): Flyer. http://www.serenate.org/publications/flyer.pdf
Serenate (2003b): Report on Final Workshop results IST-2001-34925. Study into European Research and Education Networking As Targeted by eEurope. SERENATE Deliverable no. D19. [Authors: Vietsch, K. and Martin, J. (Eds.), Cavalli, V./Dyer, J./Robertson, D./Saugstrup, D./Scott, M. and Wood, Sh.].
Taylor, J. (2001): Presentation given at UK e-Science Meeting, London, July.
Tanenbaum, E. and Mochmann, E. (Eds.) (1994): Integrating the European database: Infrastructure services and the need for integration. International Social Science Journal 46/1994, 4 (142), 499-511.

CURRENT STATE OF AND FUTURE DEMANDS IN DIFFERENT FIELDS

I. (NEW) DATA TYPES AND THEIR POTENTIALS

1. Macroeconomic Data

Ullrich Heilemann[*]

* For information on current developments of German official statistics, the author is indebted to *Thomas Luh*, Federal Statistical Office, and Dr. *Wolf-Dietmar Speich*, Statistical Office of the Free State of Saxony. For critical comments the author thanks Prof. Dr. *Roland Döhrn*, RWI, Prof. Dr. *Peter von der Lippe*, Universität Essen-Duisburg and *Klaus Löbbe*, Duisburg.

Contact:

Ullrich Heilemann
Institut für Empirische Wirtschaftsforschung (IEW)
Universität Leipzig
Grimmaische Str. 12
04109 Leipzig
e-mail: heilemann[at]wifa.uni-leipzig.de

Abstract

Germany's macroeconomic statistical infrastructure is well-developed: availability and access (including cost) are reasonable and do not leave much to be desired. Beside old demands of more and better information on stocks and flows, sectoral foreign direct investment, on new technologies, and the service sector, the present crisis will cause new requests with respect to the interaction between the monetary and the real sector. The recent trend of improving actuality of data will continue, although requests from a research perspective reliability, validity, and completeness should not get out of sight. A further strand of improvement might be the production of monthly national account (NA) data by German official statistics to improve short-term analysis and forecasting. Besides, the informational gain could be considerably enhanced by following US practice and publishing the indicator data on which the flash estimates are based. The present crisis may speed up the fulfillment of some of these demands, but given that the financial restrictions of the past decade will continue to apply, this may mean only a shift in priorities. This is even more likely with the new NA system scheduled for implementation in 2014 and it would be surprising if more were to happen than has been planned so far (e.g., the great NA-revision 2011).

1. Introduction

The following look at the current state and the future of macroeconomic data is likely to fail. For one thing, researchers will be disappointed to find that their claims for more and "better" data are not adequately supported; official statistics, while to some degree perhaps sharing this disappointment, may miss suggestions and specific comments on old and new data needs. In a material sense, the situation does not appear lamentable and no case can be made requiring immediate action. In addition, few of the following remarks are new or unique. Indeed, as an empirical macroeconomist, and as a member of various statistical advisory bodies, the present author is impressed by the progress made in numerous areas of research infrastructure that were inconceivable only a decade ago. Within the triad of data, methods, and theory, for an increasing number of areas of the social and behavioral sciences, "data" no longer appear to be the limiting factor (here appetite comes with eating, too) – especially not when also looking at cost, returns, and setting negative priorities. It is true that improvements to the macroeconomic information infrastructure over the last two decades were much smaller than the progress made in microeconomics and many of its sub-categories (for labor economics, e.g., Bender and Möller 2010; Schneider 2010). However, these other areas were only catching up with the state of macroeconomic data, which had experienced a similar jump with the launch of the system of national accounts (NA) in Germany some 50 years ago. Given the breadth of the topic, at least in the context of this publication, the following remarks will be cursory and the references rather general.

2. General remarks

At present, German macroeconomic research appears to be largely content with the existing data supply. Government interventions (e.g., in price statistics) or manipulation of statistics, not exactly uncommon in other Western countries, are more than rare here. More importantly, the general supply differs only slightly in substance or style – consistency, comparability, timeliness, etc. – from that of most other industrial countries. For nearly forty years (1920–1960), the driving forces behind the launch and completion of the present macroeconomic infrastructure (notably the national account system, or NA) had been research institutes, especially the German Institute for Economic Research, (DIW Berlin, *Deutsches Institut für Wirtschaftsforschung*) and some of its offspring. Nevertheless, large parts of the research community within and outside these institutes were all too happy when, in the 1960s, official statistics started to take over most of the business of data production and dissemination. This put an end to some institutes' quasi monopolies on some data but, obviously more important for the institutes, they felt relieved at being released from a never-ending and, in terms of academic reputation, poorly rewarded occupation.

Looking back, there has been a great deal accomplished since the 1970s to broaden and deepen the scope of the NA, for example, by using much more elaborate satellite systems for household production and the environment, and still more are to come (health economy, civil society, etc.). Of course, macroeconomists who rely heavily on the NA do have a number of requests on their agenda (see below). However, neither researchers and research institutes nor the German Social Policy Association (*Verein für Socialpolitik*) have expressed much concern about deficits (or about the state of the information infrastructure in general). For statisticians and the German Statistical Society (*Deutsche Statistische Gesellschaft*), of course, things are different (for a somewhat agnostic view, see Richter 2002). Above all, in recent years there has been enormous energy expended on all sides by the development of new concepts of the System of National Accounts (SNA), price statistics, etc., as well as by the microdata revolution.

Despite all this, the economists' gospel is still: *more and better data*, with "better" meaning "more up-to-date" (i.e., more speedily publicized data). Requests for more reliable, more valid or more compatible data are rarely heard. When one takes methods and theory into consideration, however, the priorities seem less clear. Few economists would agree that the marginal return of a Euro spent for investment in research would yield the most if it were spent on data.

Many recent improvements, notably the speeding-up of publication of NA data can be traced back to international financial markets and policy (in particular on the level of EU). It should have been clear right from the beginning that this might have consequences for data reliability and might increase the amount of revisions necessary. However, at the time, this did not really matter. Of course, it would have helped users of this data to know the actual trade-off between timeliness and "accuracy" – its size, whether it changed over time, what might be done to improve accuracy, which aggregates are the most relevant,[1] whether there are differences from other national statistical systems, and if so, can these differences be linked to particular procedures and models, and what can be done to reduce them. So far, only a few users seem to have asked these questions, and no answers have been given.[2] The same questions might be asked with respect to the new SNA (ESA[3] 95): did the list of trade-offs change, in which direction, etc.? Again, no such questions are being asked.

Requests for more and better data are usually answered by the statistical authorities by pointing at the cost involved, their limited resources, and fixed priorities, all of which are hard to contradict by third parties.[4] In general, the cost-benefit ratios of German official statistics and the approximately 10 Euro spent per capita for statistics appear favorable. However, specific information on the cost, including the burden on respondents, on specific fields (macro- and microeconomic, business cycle-growth analyses and forecasts, etc.) is not available for outsiders (Heilemann 1999). Even more difficult to clarify is the utility of (additional) investment in the various segments of the information infrastructure, most of all from a research perspective. The economic and fiscal savings from precise and timely macro data may be enormous; however, the privilege of setting official statistics' priorities will remain with policy.

Generally, the need to improve consistency, comparability, and timeliness, etc. – factors associated with data "style" – of the available macroeconomic data is more urgent than the need for new data, which is limited to a few areas. This is different from the last ten to fifteen years, which were

1 For the unfulfilled quest for metadata, see Gregory et al. (2010).

2 As an example for such a study (for the UK), see Maitland-Smith (2003), recently for Germany also Kholodilin and Siliverstovs (2009). Leigh and Stehn (2009) rank Germany surprisingly low in a comparison of temporal stability (1965Q1-2004Q4/ 1995Q1-2004Q4) of revisions in the G7 countries. Similary, so does the European Central Bank (2009) in a Euro area comparison of revisions of NA demand aggregates.

3 European System of Accounts.

4 For some elements of the recent discussions regarding the costs of statistics in Germany, see v. d. Lippe (2006) and Schupp et al. (2003).

characterized by rapidly changing needs that made their way onto the agenda of official statistics, of which many items have been settled. What remains open, however, is the degree to which the hugely increased supply of microdata can improve the *empirical* foundation of macrodata (see, for example, Becker et al. 2006), and in turn, thus testing the "macro compatibility" of microdata.

3. Specific demands

A detailed appraisal of current research interests and the resulting demands on the present and future data situations faces a number of challenges.[5] First, with respect to macroeconomic data, there is still a large backlog of "unfilled orders." To mention just a few: data requirements that came with the advent of globalization, such as the need for detailed information on stocks and flows of foreign direct investment coming from a number of sectors; data on new technologies, and the service economy.[6] Second and more fundamentally, researcher data requests are necessarily stimulated by impending problems, as a closer look at the "backlog" of orders demonstrates. Of course, sometimes things also go the other way and, for a number of reasons, new data may stimulate new questions. Present data needs could hardly have been foreseen five years ago. Even harder to anticipate are data needs that may arise in reaction to the present mixture of crises – financial, regulatory, macroeconomic, sector, currency, etc. Economic "theory" will hardly serve as a guide, as some may be hoping: it may march to the beat of the same drummer as empirical research, but its empirical zeal has usually been modest. Despite the availability of more and "better" data, experience tells us that this limited interest in empirical analysis will hardly change in the near future. It is true that growth theories – both old and new – articulate their needs for a better coverage of human capital,[7] but by now these are old requests and part of the "backlog." Third, it should be remembered that the main thrust for improving official statistics are policy needs on the national and, increasingly, on the international level – certainly if monetary or other

5 See also from an US perspective, the Jubilee Volume of The Conference on Research in Income and Wealth (Berndt and Triplett 1990).

6 For details, see, for example, CEIES 2002 and the website of the former 'European Advisory Committee on Statistical Information in the Economic and Social Spheres (CEIES)', now ESAC (European Statistical Advisory Committee) http://forum.europa.eu.int/Public/irc/dsisi/ceies/home, see also Heilemann (2003). For a more (US) research-oriented listing, see http://www.nber.org/CRIW/general. For migration, see also Kahanec and Zimmermann (2010) and Haug (2010), both in this volume.

7 See the various education related papers in this volume.

costs are involved. In the end, all users will benefit from this. The progress of information technology has reduced all these costs (another way is to make better use of administrative data) and will continue to do so, not least because of the government's goal of a "slim state," which will continue to require fiscal prudence. An exception may be the financial sector (including statistics) – crisis is the father of progress.

The ever-increasing interest in data on the service sector was a natural extension of its mounting size. Additional impulses came in the 1990s, when supply factors such as the Information Technology and the demand of the finance industry and of the information/knowledge society shaped the "New Economy." Its direct and indirect links with the rest of the economy intensified, notably with industry, as illustrated by the 1990s productivity miracle in the US. By now, many disputes about the role and scale of technical progress have since been settled, although some of the questions raised – measuring output, hours worked-productivity, prices – still lack convincing answers, particularly in Europe and Germany. National and international statistical bodies have made considerable efforts to overcome some of these difficulties.[8] Germany, for example, employed annual structural surveys in the service sector (activities in transport and communication, real estate, etc.).[9] However, in other parts of the service sector, notably in banking and insurance, such surveys, as well as reliable short-term indicators for the service sector, are still missing.

Other avenues for research (and policy) that were opened by globalization are the causes, forms, and consequences on intra-firm and intra-group trade, FDI proprietorship, trade restrictions, and – strange as this may sound – information on the size, development, forms, and structure of illegal activities (including the shadow economy). While the material and substantial dimensions of these problems are already difficult enough to cover, arguments about "style" pose even greater difficulties given their transnational, all-embracing nature. While by now the problem is recognized, attempts to tackle it have only just started.[10]

Looking more closely at researcher demands, most of them seem to be related to the need for a broader and more fully integrated macroeconomic perspective. Starting with a traditional model of business cycle analysis of the *Keynes/Klein* type as a core model and a general framework of (multipurpose) macroeconomic analysis, since the 1980s a number of subsystems

8 For example, the EU Commission funded the EU Klems project that aims to create a database on measures of economic growth, productivity, employment creation, capital formation and technological change at the industry level for all European Union Member States from 1970 onwards (http://www.euklems.net).

9 It should be noticed that the "great" NA revision in 2011 will picture, among others, the service sector in a more detailed way.

10 For a detailed outline of the problems, user needs, and approaches followed by national and international statistical bodies, see, for example, CEIES (2000).

or satellite systems have been added that interdependently explain demographic developments, human resources and human capital, energy, mobility of capital and people, etc.[11] Clearly, this requires a high compatibility of data, and long time series within or at least compatible with the NA framework.

The needs of business cycle research proper deserve more attention, quite independent from the present crisis (Löbbe 2002). More precisely, while the indicator approach already enjoys a great deal of attention – at least at the level of the Economic and Monetary Union (EMU) of the European Union (Eurostat 2005) – the analytical branch of business cycle analysis seems to be lagging behind. From the point of view of both theoretical and applied analysis, it would be a great step forward if, first, primary data on stocks were freed from the stigma of being residuals; and, second, if inventories were disaggregated – both of which have been demanded for over forty years (Fürst 1967).[12] An even more important leap forward would be an integrated accounting of the distribution of financial and real income and wealth within the NA or compatible with it.[13] This would allow for a detailed examination of the consequences of the functional, as well as personal, distribution of income and wealth as suggested by macroeconomic theory, in particular in mature economies like that of Germany. Whether it will result in an improved explanation (or even more accurate forecasts) of private consumption or private investment remains to be seen. In any case, the information itself would be valuable.

Again, similar ideas have been put forward with respect to a better understanding and modeling of financial motives and financial markets, their actors and institutions (e.g., Eckstein 1983: 77ff). While some flow of funds models have been developed for the German monetary sector, their explanatory power, for a number of reasons, has not been very convincing. However, from a macroeconomic perspective, what is more troublesome is that they have not been linked to the real sector, because the data for the closure of the various channels of transmission – the many forms in which wealth is held – are missing.[14] Of course, things will become even more difficult if we look for a proper inclusion of the international dimension (i.e., globalization and its consequences, not to mention the European Monetary Union). Currently, the first vintage of actual data on international trade in goods is reported about two years after the fact, though preliminary data are not generally

11 While the 1983 version of the DRI model of the US economy (Eckstein 1983) can serve as an early example for such a concept and its implementation, the Dutch CORE model (e.g., CPB 1999) may be seen as an illustration of present demands and possibilities.

12 US official statistics have long since published disaggregated inventory data. German OS acknowledges this need as established in ESA 1995, but because of the high cost has thus far declined to do.

13 For a current synopsis of the aggregated and sectoral non-financial wealth accounting, see, for example, Schmalwasser and Müller (2009).

14 Ibid.

criticized as being particularly deficient. However, neither monetary flows, nor data of (other) assets (including human capital or property rights) are reported with the necessary detail or quality.

A more complete system, linking the flow of funds and asset data from the international economy, would greatly improve our understanding of how the financial sector functions and would make, for example, the current demand for a contagion-related stress test of the financial system more realistic and reliable. Only then will we be able to examine the number and roles of the channels of transmission of various crises and their effects. Again, to analyze such influences on investment, consumption, government, distribution, and the foreign sector[15] requires more information on wealth and income, its composition, and distribution (Hauser 2010), at least as much as possible within the framework of the NA. All of these requests had already been made in the first report of the Council of Economic Experts (1964), and have been repeated many times since (e.g., Hax 1998; Glöckler 2003). In this context, the many discrepancies between financial accounting and NA should also be mentioned. Often, the differences are only the consequence of an incongruent dating of transactions, but this is sufficient to hamper economic analysis and assessment.

While the now easy and nearly cost-free access to official data (journals may soon follow) has been much welcomed by the academic community, equally impressive progress with regard to databases is often overlooked. The timeliness of publication of NA data has been greatly improved and harmonized between EU Member States, developments that may especially benefit forecasters. There is now a continuous quality monitoring process, in particular with respect to revision.[16] However, it would be interesting to know, for example, whether revision needs have been increased by the now shorter publication periods or by the new System of National Accounts (SNA) (ESA 1995). Besides, the informational gain could be considerably enhanced by following US practice and publishing the indicator data on which the flash estimates are based. Forecasters are not the only ones who should benefit from knowing the past and present trade-offs between timeliness and revision practices and needs. It remains to be seen whether the current greater timeliness of the NA data is – from a broader quality perspective – a net gain, not just for policy and the financial markets, but for the academic community as well. Finally, official statistics might also reflect on the handling of chain index-based SNA data by the US and others: the loss of precision when using absolute terms instead of indices is small while computation is greatly eased.

15 For example, how large would Germany's or Japan's net exports be in terms of proprietorship?

16 See Körner and Schmidt (2006). This is a welcome first step but, of course, it could be extended to metadata, once they are reported. See also FN 2.

Even if the previous list is incomplete with respect to both data substance and style, there is still an old and long list of demands made of official statistics.[17] Again, we should realize that considerable returns on investment might come from improving foreign information supply and infrastructures. This holds from both a research perspective – particularly comparative perspectives – as well as from a policy perspective. The support given in the wake of the EU's southern and eastern enlargements were of considerable help, even if, as with any harmonization, we may have to at least temporarily pay for this with a reduction of national standards.[18]

4. How to move forward

Goals and means are dependent on each other, and the quality of data is largely determined by who is collecting and who is processing them. The current crises will shift present priorities in the direction suggested above, even if, so far, there have been no hints that the German government is willing to commit more resources to this purpose, financial or administrative, its own or that of respondents. At present and for the near future, financial resources appear, at best, fixed. Negative priorities will be hard to set, and the potential to increase productivity appears for outsiders to be rather limited, as privatization and outsourcing experiments in other countries during the past decade have shown. To reduce costs, the use of administrative data might be increased, while the use of primary data is reduced – hardly a reason to expect improvements in data quality. Another ambivalent example is the increase in the cut-off limits for enterprises, which has consequences for intermediate consumption, and our picture of the size and the dynamics of the economy, especially in Eastern Germany. At first, this will affect only the structural perspective, but ultimately it will also affect the aggregate level and its dynamics. On the other side, a wider reliance on administrative data may augment the coherence and compatibility of OS data.

Leaving aside the overall comfortable situation for macroeconomics, a way toward further improvement would be to renew researcher interest in data production and their passion for statistics, a source that thus far seems to have been addressed in the discussion of the "information infrastructure"

17 See on this, for example, Richter (2002) and his often very demanding requests.

18 While there is no doubt that in recent years the European Commission (policy!) became increasingly important for national statistics, for a number of reasons not all researchers may be happy with that. For a European policy view on the statistical infrastructure, see Reeh (2010).

only in an indirect way.[19] This is not to overlook the fact that some of the major research institutes are trying to come back to their roots in the creation and improvement of specific research infrastructures.[20] Ignoring policy institutions like central banks and international institutions like the OECD, it is only outside of official statistics that these institutes have enough expertise and motivation to engage in questions of macroeconomic data. More engagement and more reputational reward by the (German) academic community would benefit both their work and the information infrastructure. This is a view supported not only by a look at the US, but also by looking back at German experiences before and after WWII (see above).

5. Conclusions and recommendations

Germany's macroeconomic statistical infrastructure is comparatively well-developed: availability and access (including cost) are reasonable and do not leave much to be desired. The list of proposals for extension and improvements is long and comprehensive, although, once again, this is not that different from the lists drawn up in most other highly developed countries. The present crisis may speed up the fulfillment of some of these demands, but given that the same financial restrictions of the past decade will continue to apply, and the opportunities for additional productivity gains are small, we need a shift in priorities which we did not see so far – despite all of the rhetoric on the statistical needs of the "information society." This is more likely with the new NA system scheduled for implementation in 2014[21] but it would be surprising if more were to happen than has been planned so far (e.g., the great NA-revision 2011).

A new way to diminish this dilemma would be to stimulate, if not a passion at least a stronger interest in questions of macroeconomic data within the academic community. All sides involved would gain much by bringing the academic community closer to this, the forefront of empirical statistical research, making it a closer ally of official statistics, as witnessed in microeconomics over the past twenty years.

19 See KVI (2001): 137ff, 146ff. Improving university education may be one strand, improving research standards another. See also, for example, Richter (2002): 293ff.

20 To mention just one example, the efforts of the DIW Berlin (Cors and Kouzine 2003) to bridge the gap between quarterly data may be cited. For a more complete overview, see KVI (2001): 102ff.

21 In 2003 the Statistical Commission of the United Nations (UNSC) initialized a revision of the SNA 1993 http://unstats.un.org/unsd/sna1993/issues. In the course of this process, ESA 95 will be revised. Different from SNA, this will not be mandatory for EU Member States.

The suggestions made here, if implemented, would broaden and improve the data infrastructure. They would help to find better solutions for our problems, primarily for old problems. Scope, frequency, and timeliness of macroeconomic forecasts will further increase and, with some luck, the amount of revisions will be reduced. In the end, monthly data may trigger a jump of insight in macroeconomic dynamics similar to the one that came with the transition from annual to quarterly data.[22] However, whether the accuracy of rate of growth forecasts of real GDP will increase more than by one or two digits is doubtful. The experience of the last 40 years – not just in Germany – does not support such hopes. However, not to worry: neither do theory nor new methods.

22 The DIW Berlin started reporting quarterly NA data in 1953. OS began publishing complete sets of NA data in 1978. For an exposition of the possible gains of monthly NA-data, as well as the experiments conducted by a number of forecasters to produce this data, see Klein (2009). This exposition also includes principle component analysis.

References:

Becker, R.A./Haltiwanger, J./Jarmin, R.S./Klimek, S.D. and Wilson, D.J. (2006): Micro and macro data integration: the case of capital. In: Jorgenson, D.W./Landefeld, J.S. and Nordhaus, W.D. (Eds.): A new architecture for the U.S. national accounts. Chicago.

Bender, S. and Möller, J. (2010): Data from the Federal Employment Agency. [In this publication].

Berndt, E.R. and Triplett, J.E. (Eds.) (1990): Fifty years of economic measurement – the Jubilee of the Conference on Research in Income and Wealth. (NBER Studies in Income and Wealth, (54). Chicago.

CEIES (Ed.) (2000): Proceedings of the eighth seminar – Foreign direct investment, Venice, Italy, 11 and 12 November 1999. Luxembourg: European Communities. http://forum.europa.eu.int/Public/irc/dsisi/ceies/home.

CEIES (Ed.) (2002): Proceedings of the seventeenth seminar – The new economy – key measurement issues, Rome, Italy, 4 and 5 March 2002. Luxembourg: European Communities. http://forum.europa.eu.int/Public/irc/dsisi/ceies/home.

Cors, A. and Kouzine, V. (2003): An approach for timely estimations of the German GDP. Allgemeines Statistisches Archiv, (87), 201-220.

CPB Netherlands Bureau for Economic Policy Analysis (Ed.) (1999): WorldScan – the core version. The Hague.

Eckstein, O. (1983): The DRI-model of the U.S. economy, New York.

European Central Bank (2009): Revisions to GDP estimates in the Euro Area. ECB Monthly Bulletin, April, 85-90.

Eurostat (Ed.) (2005): Short-term indicators – Priority assessment, quality, revisions and timeliness – 90th DGINS Conference – Paris, 13-14 September 2004. Luxembourg.

Fürst, G. (1967): Reichen die Informationen über den Konjunkturverlauf aus? Der Volkswirt [seit 1973: Die Wirtschaftswoche], Jg. 21, 2853-2856.

Glöckler, W. (2003): Sachverständigenrat und Statistisches Bundesamt. In: Statistisches Bundesamt (Ed.): 40 Jahre Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung 1963 – 2003. Wiesbaden.

Gregory, A./Heuss, P. and Ryssevik, J. (2010): Metadata. [In this publication].

Haug, S. (2010): Migration and Statistics. [In this publication].

Hauser, R. (2010): Household Income, Poverty, and Wealth. [In this publication].

Hax, H. (1998): Anforderungen an die Statistik zur Konjunkturbeobachtung und Konjunkturprognose durch den Sachverständigenrat. Allgemeines Statistisches Archiv, (82), 15-24.

Heilemann, U. (1999): Die amtliche Statistik – Dienstleiter in der Informationsgesellschaft. In: Statistisches Bundesamt (Ed.): 50 Jahre Bundesrepublik Deutschland: Amtliche Statistik – Ein konstitutives Element des demokratischen Staates. Wiesbaden.

Heilemann, U. (2003): Globalisation and statistics – a movable target. Statistical Journal of the United Nations Economic Commission for Europe 20 (2), 83-88.

Kahanec, M. and Zimmermann, K.F. (2010): Migration and Globalization. [In this publication].

Kholodilin, K.A. and Siliverstovs, B. (2009): Do forecasters inform or reassure? Evaluation of German real time data. DIW-Discussion Papers No. 888.
Klein, L.R. (Ed.) (2009): The making of national forcasts. Cheltenham.
Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Ed.) (2001): Wege zu einer besseren informationellen Infrastruktur. Baden-Baden.
Körner, T. and Schmidt, J. (2006): Qualitätsberichte – ein neues Informationsangebot über Methoden, Definitionen und Datenqualitäten der Bundesstatistiken. Wirtschaft und Statistik, 109-117.
Leigh, D. and Stehn, S.J. (2009): Fiscal and monetary policy during downturns: evidence from G7. IMF Working Paper WP/09/50.
Löbbe, K. (2002): Möglichkeiten und Grenzen der Konjunkturanalyse anhand der amtlichen Statistik. Kurzgutachten im Auftrag der Hans-Böckler-Stiftung. Düsseldorf.
Maitland-Smith, F. (2003): Revisions analysis – Quarterly GDP. Presentation at the OECD/IMF Workshop „Assessing + Improving Statistical Quality“, 5-7 Nov. 2003, http://www.oecd.org/dataoecd/26/14/18989859.ppt. [Last visited: 02/19/2010].
Reeh, K. (2010): The European Dimension. [In this publication].
Richter, J. (2002): Kategorien und Grenzen der empirischen Wirtschaftsforschung. Stuttgart.
Schupp, J./Stäglin, R. and Wagner, G.G. (2003): Entbürokratisierung der Statistik durch Flexibilisierung. DIW-Wochenbericht, (24), 396-399.
Schmalwasser, O. and Müller, A. (2009): Gesamtwirtschaftliche und sektorale nichtfinanzielle Vermögensbilanzen. Wirtschaft und Statistik, 137-147.
Schneider, H. (2010): More and Better Data for Labor Market Research. Proposals for Efficient Access to the Currently Unused Potential of Official Statistical Data. [In this publication].
von der Lippe, P. (2006): Ist der Föderalismus in der Statistik noch zeitgemäß? Allgemeines Statistisches Archiv, (90), 341-355.

2. Interdisciplinary Longitudinal Surveys

Jürgen Schupp and Joachim R. Frick

Contact:

Jürgen Schupp
Joachim R. Frick
German Socio-Economic Panel (SOEP)
DIW Berlin
Mohrenstr. 58
10117 Berlin
Germany
e-mail: jschupp[at]diw.de
jfrick[at]diw.de

Abstract

This paper concentrates on the trends in peer-reviewed longitudinal panel studies under scientific direction. Household panel studies have succeeded in broadening their disciplinary scope. Numerous innovations such as questions dealing with psychological concepts, and age-specific topical modules, physical health measures, measures of cognitive capabilities, and behavioral experiments have been incorporated into various panel studies or are soon to be introduced. In the UK, the household panel study Understanding Society comprising 40,000 households was launched in 2009 and recently added an "innovation sample"; in the Netherlands, the new Longitudinal Internet Studies for the Social sciences (LISS) launched in 2006 with over 5,000 households will be used for the testing of innovative measurement methods.

The microdata from household panel studies like the US Panel Study of Income Dynamics (PSID), the British Household Panel Study (BHPS, the predecessor of UK HLS), the Household, Income and Labour Dynamics in Australia (HILDA) Survey, and the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) are in continuously high demand by the research and policy advisory community. More important than "discovering" entirely new survey areas is "tailoring" the details of existing survey content to new, more specific (theoretical) questions, and thus maintaining proven and widely used elements of survey content. In the years to come, "tailoring" survey content will be the real challenge facing surveys that are integrated into the existing research infrastructure like HILDA, LISS, PSID, the Swiss Household Panel (SHP, *Schweizer Haushaltspanel*), SOEP, and the British study Understanding Society.

We argue that, in the future, household panel studies should be designed to take the "margins" of the life course more fully into account. Indeed, household surveys are ideally suited to gather comprehensive data on these life phases. They can be improved, on the one hand, by including specific topics about the fetal phase of life and early childhood of children born into the panel, and on the other hand, by including better information about late life and death. In the middle of the life course, improved questions on income, savings, consumption, and wealth, as well as psychological constructs will play a central role, as will specific "event-triggered" questionnaires on central life occurrences such as marriage, divorce, and entry into and exit from unemployment.

In order to substantially improve the statistical power of long-term longitudinal data, we propose an absolute minimum number of observations of about 500 persons per birth and age cohort. As of now, only the British study Understanding Society will meet this target. A positive side-effect of such an enlargement is a significantly improved potential for analyses of relatively small groups within the population: for example, lone parents or specific immigrant groups. Another positive side-effect would be an improved potential for regional analyses. For example, in Germany, a cohort size of about 500 persons implies a survey sample size of about 20,000 households, which is large enough for analyses in the majority of federal states.

Multidisciplinary panel studies will become even more important if they are accepted as reference datasets for specialized surveys that are independent of the original panel study (e.g., observational studies such as twin studies and laboratory or intervention studies). To enhance this important function, new types of service are needed, including advice on special surveys and possibly also data preparation for special surveys.

Keywords: household panels, multidisciplinary surveys, reference datasets
JEL Classification: A12, C81, C83, C93, C99, H2, H3, H5, I12, I21, I3, J1, J2, J3, J6, J71

1. Introduction

"Longitudinal surveys, which collect information about the same persons over many years, have given the social sciences their Hubble telescope. Both allow the observing researcher to look back in time and record the antecedents of current events and transitions" (Butz and Boyle Torrey 2006: 1899).

If we look back in survey history, social scientists began as early as the 1930s to design a new kind of longitudinal study: the panel survey (Lazarsfeld and Fiske 1938). Panel surveys measure the same variables in the same individuals at two or more points in time. One of the first panel studies was conducted in the US in 1940 in the field of political science (Lazarsfeld et al. 1944). The focus was on the effect of election campaigns, the mass media, and personal communication about politics and causal relationships. Known as the "Erie County Study," Lazarsfeld's study was conducted on a sample of about 600 persons who were surveyed repeatedly over a period of more than six months in seven panel waves. This study remains a model for election studies in political science up to the present day.

In the methodological literature, panel surveys are often described as having a "prospective longitudinal design" (Featherman 1980). In such a design, a group of individuals are interviewed, tracked, and reinterviewed at least once at some future point in time. A "retrospective" panel design, on the other hand, entails collecting data on only one occasion. The longitudinal dimension of such a study is obtained by asking people to recall what things were like at some earlier point in time, as well as at present (de Vaus 2001). This means that it is not strictly necessary to use a longitudinal research design to collect longitudinal data, although there are conceptual distinctions among different types of longitudinal data (Featherman 1980). Here, a crucial question is how reliable retrospective data are as substitutes for direct observations of the past (e.g., concurrent respondent reports in longitudinal panels, independent records, etc.). Such retrospective designs have been used in sociology to collect event history data covering the entire life course. An example of such a study is the German Life History Study (GLHS) (Brückner and Mayer 1998).

In developmental psychology, longitudinal surveys have a clear prospective focus:

> "Longitudinal methodology involves repeated time-ordered observation of an individual or individuals with the goal of identifying processes and causes of intraindividual change and of interindividual patterns of intraindividual change in behavioral development" (Baltes and Nesselroade 1979: 7).

Together with total population designs, which are representative from both a cross-sectional and a longitudinal perspective, longitudinal panel surveys are described as advantageous in several respects:

> "Total population designs and longitudinal panel designs can be used for practically any type of longitudinal analysis, given a sufficient number of cohorts and measurement periods. Other designs are more limited, and their appropriateness must be judged in the context of a particular research problem" (Menard 2002: 33).

High-quality household panel surveys begin, like cross-sectional surveys, with a random sample of a set of households and of the individuals within those households. For decades, the only mode of data collection was through face-to-face, paper-and-pencil interviews. But an increasing variety of other modes of data collection have become common, some reflecting technological advances. For example, mail surveys and web-based surveys are now also being used (e.g., in the Dutch LISS panel). In addition, different modes of assessment are used. In panel surveys, trained interviewers conduct health tests and tests of cognitive ability (e.g., in SHARE[1]). Panel surveys differ from cross-sectional surveys in that they continue to follow sampled individuals at regular intervals, usually once per year (wave). Adhering to the basic "follow-up rules" determining who to contact and interview again, household panel surveys produce data on changes in the demographic, economic, and social conditions of their members and thus attempt to remain representative of the cross-sectional population as well. This is in contrast to individual panel studies covering entire birth cohorts of individuals in the population.[2] These panels represent their cohorts as they age and may gradually decline in representativity for the original age group. The household panel surveys discussed in the following section can be defined as: multiple repeated observations (usually once per calendar year) for age-heterogeneous individuals within their household context and based on a random sample of all (private) households of a country. Their theoretical concept and variables cover a wide range of social and economic issues.

1 Survey of Health, Ageing and Retirement in Europe.

2 Like the longitudinal design of the 1958 National Child Development Study and the 1970 British Cohort Study (BCH) (Schoon 2006) and the Millenium Cohort Study (MCS) or the British birth cohort in 2012 ('Olympic Cohort').

One may ask whether the Hubble telescope (by Butz and Boyle Torrey 2006: 1899) really is the right metaphor for panel studies. After all, since microdata is involved, the comparison with a microscope might seem more appropriate.[3] The answer is twofold. First, panel data have a temporal dimension (as noted by Butz and Boyle Torrey): they do not deliver just a "snapshot," but allow us to actually look back in time (just as telescopes do). Second, panel studies are expensive compared to other studies in economics and the social sciences. Thus, in terms of money, the comparison with a highly sophisticated but expensive device such as the Hubble telescope is much more appropriate than the comparison with more economical microscopes.

This paper gives a summary of current developments in longitudinal household surveys under academic direction. For an overview covering all the various types of panel and cohort studies, see Wissenschaftsrat (2010).

2. Status quo of multidisciplinary household panel studies under academic direction

The success story of large-scale household panels started about 40 years ago, with the US Panel Study of Income Dynamics (PSID) (Brown et al. 1996). Only household panel designs like the PSID, or the designs of the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) and British Household Panel Study (BHPS), represent all individuals and households in the population and contain an endogenous mechanism for representing demographic changes in existing households caused, for example, by new entrants (birth, immigration, regional mobility) as well as drop-outs (death, emigration) reflecting the dynamics of the underlying population.

Household panels start with a representative sample of households and a representative set of individuals residing in those households. If the tracking and following rules used in household panels call for attempted interviews with all household members in the original sample, all individuals born to the original sample members, and any individuals who have moved into those households in the meantime (see Kroh et al. 2008), then this prospective panel design continues to provide a representative cross-sectional picture of the underlying population over the life of the panel. Except for immigration into newly founded households from outside the sampling frame, all demographic events (births, deaths, emigration, and events like divorce and the departure of children from their parents' homes) are covered by a high-

3 Senator Jürgen Zöllner, who is responsible in Berlin's government for education and research, once asked this question.

quality household panel design. Immigration has to be handled through supplemental samples (see Schupp and Wagner 1995).

Due to initial non-response and attrition of panel respondents over the course of time, high-quality response and attrition analyses and carefully designed re-weighting strategies are crucial to achieve representative population estimates in panel studies (Ernst 1989; Rendtel and Harms 2009). Population estimates (indicating representativity) are an important issue, because all longitudinal and cross-sectional results of the household panel survey are in continuously high demand in both the research and policy advisory community (e.g., Wissenschaftsrat 2009: 56).

Today, some of the most widely used long-running household panel studies that seek to provide a representative view of the entire population of a given society include the BHPS, the Household Income and Labour Dynamics in Australia (HILDA) Survey, the Swiss Household Panel (SHP, *Schweizer Haushaltspanel*), and SOEP. These panels differ from the individual longitudinal studies developed by sociologists in both design and scope, using an extended household concept to measure subjective as well as objective variables. They also differ from the longitudinal cohort studies developed by epidemiologists and psychologists.

Over the course of time, household panel studies have expanded in scope – driven by new research questions of their Principal Investigators (PI) and by the demands of their scientific user communities – and now cover a number of new research questions, some dealing empirically with the "utility" of respondents and the parameters of their utility function. These include happiness and satisfaction with life, health, "other preferences" (trust, fairness, and reciprocity), risk, and inequality aversion.

"Biomarkers" are another exciting new area of research providing non-standard measurement of a respondent's "biological and medical status." One such biomarker is "grip strength," which can be used as an indicator of health (Hank et al. 2009).

> "Indeed, biomarkers on social surveys may well reveal more about subjects' predispositions and their ancestry than do their verbal responses on which social scientists have historically depended. Over the past two decades, the theory of evolution has influenced parts of economics and psychology, and to a lesser extent sociology, anthropology, and political science" (Butz and Boyle Torrey 2006: 1899).

In other words, socio-economic panel studies are incorporating an increasing number of concepts from the fields of medicine and psychology. This development has been propelled by the emergence of new research questions, and its pioneers include the Health and Retirement Survey (HRS), the English Longitudinal Study of Ageing (ELSA), and the Survey of Health, Ageing and Retirement in Europe (SHARE)[4]. The latter study provides a new, comprehensive, international view on aging, but does not cover the population under 50 years of age.

The research community unanimously supports the call for more complete data on the individual life course within the household context, and for improved opportunities to analyze intergenerational transmissions of behavior and social structures and thus to disentangle the impacts of "nature" and "nurture." Outside of the social sciences, this kind of analysis is called "behavioral genetics" (e.g., Plomin et al. 2008). And, in fact, household panel data expand the possibilities for doing research along this line because of the variety of different intergenerational relationships captured in the households surveyed.

Another methodological advantage of panel data is the possibility to make causal inferences: natural experiments created through inherent differences between institutions and countries. The international comparability of data is therefore a central objective in the governance of social statistics and longitudinal studies, and this can only be guaranteed through the optimal design of organizational and financial structures.

The Cross-National Equivalent File (CNEF, based at Cornell University in Ithaca, NY, US) provides a common database derived from existing national panels, namely PSID (US), SOEP (Germany), BHPS (UK), the Survey of Labour and Income Dynamics (SLID; Canada), SHP (Switzerland), and HILDA (Australia) (see Frick et al. 2007). And all successful household panel studies under academic direction demonstrate that the real added value of panel studies can be reaped only after ten waves or more.

To put it succinctly, the major household panel studies under academic direction (as mentioned above) stand for *theory-based data collection*, not just for more data and better statistics. And because such household panels are expensive, all of them are part of the national and/or international research infrastructure (Elias 2010).

4 Thus, while SHARE is also a prospective panel study, it is not a fully-fledged household panel, but rather an extended cohort study. The strength of SHARE is its worldwide multi-country coverage (http://www.share-project.org/).

3. Recommendations

3.1 Governance

Two prime examples of good governance in large-scale surveys are the European Social Survey (ESS, a set of repeated cross-sectional surveys run by political scientists) and SHARE (a truly interdisciplinary longitudinal study of economics, sociology, and health). Both surveys provide datasets that form an infrastructure for addressing theory-driven research questions. Unfortunately, initiatives for cross-nationally harmonized household panels, which are more expensive than studies like ESS, are often not research-driven – for example, the European Community Household Panel (ECHP), which provides annual panel data for the period 1994 to 2001. The European Statistics on Income and Living Conditions (EU-SILC), the follow-up survey of ECHP, has a reduced panel component of just four waves focusing on short-term measurement of income and poverty dynamics. EU-SILC will not, however, allow for the kind of in-depth life-course analysis necessary for testing theoretical concepts and hypotheses in the social and behavioral sciences.

We believe that the following list of recommendations can help to ensure good governance of household panel studies under academic direction:

- *Ensure medium-term funding!*
  Household panels – like other prospective longitudinal studies – crucially require stable research questions, survey content, and fieldwork. Annual funding – for example, one-year contracts with fieldwork organizations – cannot guarantee the necessary degree of stability and reliability. Although auditors and accountants may not like medium-term and especially not long-term funding or contracts, medium-term funding (covering at least five years) is the absolute minimum in the case of household panels. And to ensure the quality of the fieldwork and the longitudinal data, ten-year periods of funding and contracting are even better. Other means of quality control than short funding periods must be found to ensure the quality of the panel. In case of panels under academic direction, this is not difficult to achieve because all academic panel studies are under the permanent supervision of advisory boards (and under the "supervision" of users).

- *Get the user community involved!*
  Ongoing panel studies need ideas from their users. However, it is an open question how best to gather user input. Funding agencies are attempting more and more to promote competition. In our opinion – based not least of all on the experience of the British Household Panel Study (BHPS) – the theoretical and methodological standards of major

household panels cannot be raised simply by holding an annual competition among users to suggest an additional "One Minute Question."[5] This was used occasionally by BHPS, but is no longer being used in the new British panel study Understanding Society. While such an approach may produce mainstream add-ons, we feel it is less promising than the approach adopted by the SOEP (which has already undergone pilot testing in recent years): that of focusing on close cooperation with users who are prepared to invest their time, energy, and even resources in pretesting, with the explicit aim of increasing the SOEP's *long-term longitudinal potential*.

- *Oversample subgroups!*
  While gigantic sample sizes of 100,000 households would ensure sufficient sample sizes in the near and more distant future, with high statistical accuracy for all relevant subgroups of the population, they are not realistic in terms of funding. Thus, the oversampling of subgroups is a permanent issue for the governance of household panel studies. The new British panel study Understanding Society, with a sample size of 40,000 households covering all of the British regions, is a good example, because even this large sample cannot cover immigrants in a sufficient manner. As a result, immigrants are over-sampled. In terms of governance and funding, it is a difficult question whether oversampling of special groups should be done with household panels themselves or through related studies (with external funding) that use a fully-fledged household panel as a "reference sample." There is no clear-cut answer to this question. Whether oversampling should take place within a major household survey or by means of related studies must be discussed on a case-by-case basis.

- *Be innovative!*
  The same is true for the use of "innovation samples" to address highly specified, theory-driven research questions that require specific variables and possibly also specific survey methods. Incorporating such aspects into an ongoing longitudinal survey has the advantage that one need not wait for many years before doing longitudinal analysis. A longitudinal innovation sample that is open to new kinds of measurement is of much higher research value than a new cross-sectional innovation sample. The Dutch LISS panel could possibly become a model for future innovation samples.[6]

---

5 This refers to a competition to create special questions, for which a specific amount of time will be allocated in the survey.

6 The governance of this innovative household panel is documented at http://www.lissdata.nl/lissdata/.

- *Push for related studies!*
  A representative, large-scale household panel sample can serve as a control sample for intervention studies that may be carried out using parts of the innovation sample or as related studies (Anger et al. 2009; Siedler et al. 2009).

3.2 *Important areas for substantive enlargements and methodological improvements in the survey programs of household panels*[7]

In order to understand human life and human society better, we need to understand human beings as fundamentally social beings. It is thus important to study the range of networks (and areas) in which humans live. But at the same time, there is increasing evidence that sociality is not only a cultural phenomenon (highlighting the importance of intergenerational networks as mentioned above), but that it is also – to a degree that varies between individuals – "hard-wired" into our genome through epigenetic inheritance (Fehr 2009). International developments suggest the value of more systematically surveying a number of variables on the biological foundations of human life (biological and personality characteristics) in a number of areas, and of studying the networks in which individuals, their families, and their households are embedded.

This systematic approach to measurement is not only the result of theoretical improvements but is also driven largely by new technological opportunities for measurement and analysis (e.g., experiments in the lab and in the field, surveys using the Internet and mobile phones, methods of collecting biomarkers and analyzing the genome). In fact, this new analytical approach currently appears to be driven even more by new technologies than by new theoretical insights. This might seem to contradict textbook reasoning about the primacy of scientific theory over pure measurement possibilities, based on the idea that empirical methods should only be used to test the empirical implications of specific theories. "Measurement without theory" is an old and serious criticism lodged against empirical research and data collection. However, in the history of science, we find numerous examples demonstrating that new measurement methods often precede and indeed pave the way for theoretical reasoning. One prominent example is Galileo's telescope, first used 400 years ago, in the year 1609 in Padua. Although it was invented for

7 Without challenging the importance of the following issues, we do not address here questions of improved data management (e.g., by means of the "long format" and the Data Documentation Initiative (DDI)), data distribution (Rendtel 2009), and improved IT technologies (see, e.g., www.opendatafoundation.org). We also do not discuss the possiblities of "paradata," which deliver information about the fieldwork process (Kreuter and Casas-Cordero 2010). We do believe that paradata are of utmost importance for the analysis and control of fieldwork processes, attrition analyses, and weighting (Schräpler et al. 2010).

practical purposes, it revolutionized not only the measurement of the visible universe, but a lot of theories too. In the future, “new eyes” will show us further “new skies” (Kanipe 2009).

It is self-evident that the ambitious goal of comprehensively measuring human life trajectories could easily overtax respondents and lead to declining and, in particular, selective response rates. For this reason, we propose that for household panels requiring a high cooperation rate among long-term respondents, new survey methodologies should be tested, such as a standardized “multi-method approach” and “matrix sampling.” In matrix sampling, missing values are deliberately created (and later replaced with imputed values) by randomly assigning certain questions that are *not* to be asked to particular subsamples. This reduces the burden of the number of questions to be answered. Though appealing in theory, this method will be challenging to implement successfully in a long-running survey. It may also be worthwhile to use more special proxy questionnaires for the youngest panel “members” who are not able to respond on their own in early childhood, or for those who cannot participate due to temporary absence or bad health.

The comprehensive survey program developed for, and partly realized in, the classic social scientific survey of the “Unemployed of Marienthal” (see Jahoda et al. 1933) appears more promising than ever. Yet since the 1970s, with the growing popularity of standardized survey research, the methodology used in the Marienthal Study has been gradually abandoned. Today, new technologies make more accurate and comprehensive empirical research possible.

Among geneticists, who focus on heritable influences on human behavior, it is broadly accepted that social context is essential for understanding human outcomes. Typically, several different genes and environments play a role in certain outcomes, and it is therefore crucial to study the interactions between the two mechanisms to understand the complexities and dynamics of human behavior. On the other hand, recent work by sociologists and economists provides further evidence that individuals do not respond to societal contextual influences in a unique or socially contingent way. This means that only multidisciplinary collaboration integrating genetic approaches can be expected to produce new insights into this complex relationship (Freese 2009; Guo et al. 2008). The SOEP study has already taken initial steps in this direction, aimed at an interdisciplinary enlargement of the research design (Schupp and Wagner 2010).

- *Better data on the start and end of life*

  Thanks to their longitudinal design, household panel studies are ideally suited to trace the biographies of birth cohorts from the very beginning to the terminal phases of life.

  In an ongoing household panel study, membership does not begin at birth (as is the case in conventional cohort studies) but indeed prior to birth, through the participation of one or both parents in the study. The potential of this unique feature of household panels can be exploited by asking mothers-to-be questions about pregnancy and (very) early childhood. These data allow the economic and social conditions at conception and during pregnancy to be taken into account as aspects defining the individual life course.

Household panels not only provide the opportunity to observe the life course from the very beginning, but also shed light on the terminal phase of life.[8] However, when health declines in later years, respondents often become unwilling or incapable of responding on their own. In these cases, proxy interviews are a useful alternative, yet they remain relatively uncommon. Furthermore, it is often necessary to follow respondents from private households into retirement or nursing homes (Jürges 2010).

- *Consumption and savings*
  Up to now, consumption has generally not been covered well by household surveys. However, in theoretical terms, consumption is an important measure of economic well-being at the individual and the household level. Due to the complexity and respondent burden involved in surveying high-quality data on consumption (levels and patterns), it is widely believed that well-being can be proxied by income. Obviously, this is less than adequate, since income may indeed be much more volatile than consumption, necessitating information on income (a flow measure) as well as on the process of (dis)saving to smooth consumption.

  In order to better understand human behavior in this context, the collection of information on wealth (stock measure), as well as on changes in wealth holdings over time, appears to be especially fruitful for long-running household panel surveys like SOEP (see Frick et al. 2007). Recent advancements in the collection of expenditure data, rather than consumption data, have been made in the Australian HILDA survey, providing clear evidence that income poverty is different from consumption poverty as well as from low wealth (see, e.g., Headey 2008). The 2010 wave of SOEP will, for the first time, include a short assessment of expenditures in the most important domains (housing, nutrition, education, family transfers, and savings).

- *Better measures of competencies*
  In all household panel surveys, human capital has traditionally been operationalized solely by measuring educational attainment as the highest level of schooling or vocational training completed. It seems "natural" to

8 See also Romeu Gordo et al. (2009) about household panels as a resource for research on aging, and Kröger (2008) for a pretest of the SOEP exit questionnaire.

improve household panel survey instruments by collecting better data on the cognitive competencies of respondents using standardized measurement procedures (Schupp et al. 2008). In addition, there should be increased efforts to record what are known as non-cognitive capabilities, that is, competencies that are not necessarily acquired in educational institutions but (to a greater extent) at home during early childhood. The SOEP survey program will be extended in a number of ways in the coming years to cover the area of skills (Grabner and Stern 2010; Uhlig et al. 2009).[9]

- *Health and the biological foundations of social and economic behavior*
  Despite the growing interest in integrating biomarkers into surveys, we are convinced that the collection of biomarkers in household panel studies in an unrestricted manner, and solely to address medical research questions, would not be useful or even practicable. Attempting to move in the direction of medical research would impose too high a burden on respondents (as regards the scope and duration of the survey) and would impede the useful division of labor between different methodological approaches and surveys. Rather, a survey of this kind would be a perfect example of a "related study."

  However, biomarkers that can be used to enhance social and behavioral science analyses, and in some cases consolidate their results considerably, promise to be highly useful (National Research Council 2008). One of the reasons is that longitudinal surveys deliver, through repeated measurement, very reliable pictures of phenotypes (the term used by life scientists to describe organisms as the result of the interaction between genotype and environment). Thus, with longitudinal data produced by social scientists, we are much more likely to identify the biological foundations of human behavior than with converse approaches: for example, if life scientists tried to enrich biobanks with social variables.

- *Other measurement improvements*
  A new technology, and an alternative to item sets, is what are known as factorial designs with vignettes. These questions ask respondents very detailed questions about fictitious situations and decisions. This approach is a kind of quasi-experiment (Sauer et al. 2009).

  New technologies have opened up completely new possibilities for measuring human behavior and biographies in the context of personal networks and local environment. We believe that these new measurement possibilities are especially valuable within prospective panel

9 See the research network "Nicht-kognitive Fähigkeiten: Erwerb und ökonomische Konsequenzen" (Non-Cognitive Skills: Acquisition and Economic Consequences). For more information, see http://www.zew.de.

studies: such new technologies can help to measure behavior between regular panel waves (which are usually conducted once per calendar year) and to measure networks and environments. We briefly mention some of these new opportunities without having the space to discuss them in depth. Mobile phones can be used as devices for sampling between regular panel waves. In fact, this is being done already (Riediger 2010; Riediger et al. 2009). It will be relativeley straightforward to use the same technology to locate respondents who have moved and collecting photos and sound bytes from their everyday life (Mehl et al. 2007). Even monitoring the physical status of respondents over the course of a day (or several days) with systems currently used by physicians to monitor their patients would be possible (Wrzus et al. 2010).

Networks and local environments of respondents can be measured by links to their Facebook accounts (if respondents give permission). And on the basis of respondents' statements, links can be created to organizational data (e.g., on employers or childcare facilities) (Liebig 2010).

In addition, panel studies can gradually be enhanced by carrying out internal surveys of contextual data. At SOEP, we intend to start with specific surveys that gather data on organizational contexts from 2011 onwards. These will include targeted surveys in childcare centers, schools, and at respondents' workplaces. In 2007, we administered such a pre-test and obtained positive results. It showed that respondents are by and large willing to pass on the addresses of their childcare centers, schools, and employers (Schupp et al. 2008). In 2008, the German General Social Survey (ALLBUS, *Allgemeine Bevölkerungsumfrage*) carried out its first survey of this kind[10] at the workplaces of all employed survey respondents; the results will be used to lay the groundwork for similar questions.

Based on the private addresses of respondents, records can be linked to an increasing number of geo-coded databases, providing information on the local weather or availability of local infrastructure, for example (Lakes 2010, Goebel et al. 2010).

Survey data and behavioral experiments also can be combined (Gächter 2010; Naef and Schupp 2009). Online games, for example, can be used to run behavioral experiments (Bell et al. 2009; Castronova and Falk 2010). And for special subsamples, in-depth studies are possible based on approaches of "qualitative social research" (Teddlie and Tashakorri 2003; Laurie and Sullivan 1991).

10 The ALLBUS Organisational survey is being led by Stefan Liebig, who also provided advice to the SOEP when first pilots were being carried out during pretesting (for first results on such an Linked Employer-Employee (LEE) dataset, see Meyermann et al. 2009).

3.3 Developments in sample design and fieldwork

Sufficiently large cohort subsamples allow researchers to analyze the impact of new retirement regulations or measures like the "child-raising allowance" in Germany. To meet our objective of providing statistically reliable information on groups of individuals born in the same year (age cohorts), we consider 500 cases per cohort to be a minimum. With about 500 observations per birth cohort, a researcher can analyze how the new policy instrument works for two very similar birth cohorts: one that is affected by the new law and one that is not. Another example is migration research, which profits similarly from larger samples (Haug 2010; Farwick 2010).

Possible developments in household panel samples are not limited to just enlarging sample size and overall statistical power. The inclusion of special populations (in the case of SOEP, groups like immigrants and high-income households) is another possibility. And not only socio-economic subgroups of the population can be of interest: twins are also candidates for oversampling as a genetically interesting subgroup (Spinath 2010).

In the context of aging societies in Germany and many other Western countries, the coverage of persons in institutions needs to be improved – particularly individuals in (residential) nursing homes. Here the main focus should not be on achieving representative coverage of the institutional population as such, but on covering the life transition from private household to institutional care. This kind of longitudinal data is of high scientific and practical importance for better understanding health changes in old age, intergenerational relations, the relevance of institutional care arrangements for the individual life course, and, last but not least, the process of dying in modern societies. At present, household panels tend to be confronted with non-response when elderly respondents move into (nursing) homes. Here, the difficulties of interviewing persons affected by dementia constitutes a major hurdle; in this special case, the option of having care providers conduct proxy interviews requires further investigation.

International migration and migration dynamics play an increasingly significant role in society. In 2006, more Germans left their native country than ever before, except for the emigration wave of the 1950s. In household panel surveys, respondents moving abroad are no longer included in the sample. At SOEP, groundwork has already been undertaken for surveying abroad. Respondents who have left Germany since 2004 have been contacted, and surveys have been conducted in writing.[11] The hope is that this will make it easier to re-integrate these individuals into the standard sample when and if they return to Germany, since they will never have left the sample completely. Obviously, following internationally mobile indi-

11 "Living outside Germany." See, for first results, Schupp et al. (2008).

viduals will require very sophisticated fieldwork. However, in light of the harmonization of household panel surveys within the European Union, we expect increasing research interest in following mobile EU citizens across national borders to gain a better understanding of the motives and the consequences of mobility.

4. Concluding remarks

Datasets generated from multidisciplinary panel surveys are usually extremely rich in analytical potential. At the same time

> "(t)he richness of panel data is of value only to the extent that the dataset is analyzed, and analyzed in a timely manner. Running a panel survey is like being on a treadmill: the operations of questionnaire design, data collection, processing and analysis have to be undertaken repeatedly for each successive wave. There is a real danger that the survey team will become overwhelmed by this process with the result that the data are not fully analyzed. To avoid this danger, adequate staffing is needed and a well-integrated organization needs to be established" (Kalton and Citro 1993: 212).

Multidisciplinary household panel surveys need an institutionalized organizational setting, and they are outstanding examples of the research infrastructure that is vital for the social and behavioral sciences. Aside from the group of principal investigators running these kinds of panel studies, they also crucially require a multidisciplinary user group active in analyzing the data and publishing results. An exchange of experiences between data producers and data users is also important. Data producers can work to lower the burdens on users – for example, the challenges of learning to work with complex data structures – by providing new technologies of data distribution, documentation, and training courses. On the other side, feedback from data users on their experiences with the data can act as the scientific foundation for improving multidisciplinary household surveys. A future prospect will be the establishment of a European network of household panels under *academic direction*, with the HLS in the UK and SOEP providing key longitudinal data on the European level (Elias 2010).

References:

Anger, S. et al. (2009): Developing SOEPsurvey and SOEPservice: The (Near) Future of the German Socio-Economic Panel Study (SOEP). SOEPpaper No. 155.

Baltes, P.B. and Nesselroade, J.R. (1979): History and Rationale of Longitudinal Research. In: Nesselroade, J.R. and Baltes, P.B. (Eds.): Longitudinal Research in the Study of Behavior and Development. New York

Bearman, P. (2008): Introduction (Exploring Genetics and Social Structure). American Journal of Sociology 114 (S1), v-x.

Bell, M.W./Castronova, E. and Wagner, G.G. (2009): Surveying the Virtual World: A Large Scale Survey in Second Life Using the Virtual Data Collection Interface (VDCI). DIW Data Documentation No. 44.

Borghans, L./Duckworth, A./Heckman, J.J. and ter Weel, B. (2008): The Economics and Psychology of Personality Traits. The Journal of Human Resources 43 (4), 972-1059.

Brown, Ch./Duncan, G.J. and Stafford, F.P. (1996): Data Watch: The Panel Study of Income Dynamics. Journal of Economic Perspectives 10 (2): 155-168.

Brückner, E. and Mayer, K.U. (1998): Collecting Life History Data – Experiences From the German Life History Study. In: Giele, J.Z. and Elder, G.H. Jr. (Eds.): Methods of Life Course Research. Thousand Oaks.

Butz, W.P. and Boyle Torrey, B. (2006): Some Frontiers in Social Science. Science 312, 1898-1900.

Castronova, E. and Falk, M. (2010): Virtual Worlds as Petri Dishes for the Social and Behavioral Sciences. [In this publication].

de Vaus, D.A. (2001): Research Design in Social Research. London.

Diewald, M. and Mayer, K.U. (2008): The Sociology of the Life Course and Life Span Psychology: Integrated Paradigm or Complementing Pathways? DIW Discussion Papers No. 772.

Elias, P. (2010): Providing Data on the European Level. [In this publication].

Ermisch, J./Gambetta, D./Laurie, H./Siedler, Th. and Uhrig, S.C.N. (2007): Measuring People's Trust. ISER Working Papers No. 2007-32.

Ernst, L. (1989): Weighting Issues for Longitudinal Household and Family Estimates. In: Kasprzyk, D./Duncan, G./Kalton, G. et al. (Eds.): Panel Surveys. New York.

Farwick, A. (2010): Internal Migration. [In this publication].

Featherman, D.L. (1980): Retrospective Longitudinal Research: Methodological Considerations. Journal of Economics and Business 32 (2), 152-169.

Fehr, E. (2009): On the Economics and Biology of Trust. Journal of the European Economic Association 7 (2-3), 235-266.

Freese, J. and Shostak, S. (2009): Genetics and Social Inquiry. Annual Review of Sociology 35, 107-128.

Frick, J.R. and Grabka, M.M. (2005): Item-non-response on income questions in panel surveys: incidence, imputation and the impact on inequality and mobility. Allgemeines Statistisches Archiv 89 (1), 49-60

Frick, J.R. and Grabka, M.M. (2007): Item Non-Response and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective. SOEPpaper No. 49.

Frick, J.R./Goebel, J./Schechtman, E./Wagner, G.G. and Yitzhaki, Sh. (2006): Using Analysis of Gini (ANoGi) for detecting whether two sub-samples represent the same universe: The German Socio-Economic Panel Study (SOEP) Experience. Sociological Methods & Research 34 (4), 427-468

Frick, J.R./Jenkins, S.P./Lillard, D.R./Lipps, O. and Wooden, M. (2007): The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. Schmollers Jahrbuch 127 (4), 627-654.

Frick, J.R./Grabka, M.M. and Sierminska, E. (2007): Representative wealth data for Germany: The impact of methodological decisions around imputation and the choice of the aggregation unit. SOEPpaper No. 3.

Frick, J.R./Grabka, M.M. and Groh-Samberg, O. (2007): Estimates of Imputed Rent and Analysis of their Distributional Impact in Germany. (Research project "Accurate Income Measurement for the Assessment of Public Policies" (AIM-AP), funded by European Commission, 6th Framework Programme, 2006-2009, Contract Nr. CIT5-CT-2005-028412).

Gächter, S. (2010): Experimental Economics. [In this publication].

Geyer, S./Norozi, K./Buchhorn, R. and Wessel, A. (2008): Chances of employment in a population of women and men after surgery of congenital heart disease: Gender-specific comparisons between patients and the general population. SOEPpaper No. 91.

Goebel, J./Wagner, G.G. und Wurm, M. (2010): Exploring the Linkage of Spatial Indicators from Remote Sensing Data with Data from the Socio-Economic Panel (SOEP). SOEPpaper [forthcoming].

Guo, G. (2008): Introduction to the special issue on society and genetics. Sociological Methods & Research 37 (2), 159-163.

Guo, G./Tong, Y. and Cai, T. (2008): Gene by Social Context Interactions for Number of Sexual Partners among White Male Youths: Genetics-Informed Sociology. American Journal of Sociology 114 (Supplement), S36-S66.

Grabner, R.H. and Stern, E. (2010): Measuring Cognitive Ability. [In this publication].

Hank, K./Jürges, H./Schupp, J. and Wagner, G.G. (2009): Isometrische Greifkraft und sozialgerontologische Forschung: Ergebnisse und Analysepotentiale des SHARE und SOEP. Zeitschrift für Gerontologie und Geriatrie 42, 117-126.

Haug, S. (2010): Migration and Statistics. [In this publication].

Headey, B. (2008): Poverty Is Low Consumption and Low Wealth, Not Just Low Income. Social Indicators Research 89, 23-39.

Jahoda, M./Lazarsfeld, P.F. and Zeisel, H. (1933): Marienthal: The Sociography of an Unemployed Community – Translated by the authors with John Reginait and Thomas Elsaesser. Chicago 1971.

Jürges, H. (2010): Provision for Old Age: National and International Survey Data to Support Research and Policy on Aging. [In this publication].

Kämper, Eckhard and Manfred Nießen. (2010): The Role of the German Research Foundation. [In this publication].

Kalton, G. and Citro, C.F. (1993): Panel Survey: Adding the Forth Dimension. Survey Methodology 19 (2), 205-215.

Kanipe, J. (2009): New Eyes, New Skies. Nature 457 (7225), 18.

Kreuter, F. and Casas-Cordero, C. (2010): Paradata. [In this publication].

Kroeger, K. (2008): “Exit-Interviews” in SOEP und SHARE. DIW Data Documentation No. 40.
Kroh, M./Pischner, R./Spieß, M. and Wagner, G.G. (2008): On the Treatment of Non-Original Sample Members in the German Household Panel Study (SOEP). Methoden – Daten – Analysen 2 (2), 179-198.
Krupp, H.-J. (2008): The German Socio-Economic Panel: How It All Began. SOEPpaper No. 75.
Kruppe, Th./Müller, E./Wichert, L. and Wilke, R.A. (2008): On the Definition of Unemployment and its Implementation in Register Data – The Case of Germany. Schmollers Jahrbuch 128 (3), 461-488.
Laurie, H. and Sullivan, O. (1991): Combining Qualitative and Quantitative Data in the Longitudinal Study of Household Allocations. ISER Working Paper 1991-07.
Lakes, T. (2010): Geodata. [In this publication].
Lazarsfeld, P. and Fiske, M. (1938): The “Panel” as a New Tool for Measuring Opinion. Public Opinion Quarterly 2 (4), 596-612.
Lazarsfeld, P.F./Berelson, B. and Gaudet, H. (1944): The People´s Choice – How the Voters makes up his Mind in a Presidential Campaign. New York/London.
Liebig, S. (2010): Interdisciplinary Longitudinal Surveys. Linking Individual Data to Organizational Data in Life Course Analysis. [In this publication].
Mayer, K.U. (2009): New Directions in Life Course Research. Annual Review of Sociology 35, 413-433.
Mehl, M.R. et al. (2007): Are Women Really More Talkative Than Men? Science 317, 82.
Menard, S. (2002): Longitudinal Research. Second Edition Thousand Oaks.
Merz, J. (2010): Time Use and Time Budgets. [In this publication].
Meyermann, A./Elsner, J./Schupp, J. and Liebig, S. (2009): Pilotstudie einer surveybasierten Verknüpfung von Personen- und Betriebsdaten – Durchführung sowie Generierung einer Betriebsstudie als nachgelagerte Organisationserhebung zum SOEP-Pretest 2006. DIW Research Notes No. 31.
Naef, M. and Schupp, J. (2009): Measuring Trust: Experiments and Surveys in Contrast and Combination. SOEP Papers on Multidisciplinary Panel Data Research at DIW Berlin No. 167.
National Research Council (2008): Biosocial Surveys. Committee on Advances in Collecting and Utilizing Biological Indicators and Genetic Information in Social Science Surveys. In: Weinstein, M./Vaupel, J.W. and Wachter, K.W. (Eds.): Committee on Population, Division of Behavioral and Social Sciences and Education. Washington.
Plomin, R./DeFries, J.C./McClearn, G.E. and McGuffin, P. (2008): Behavioral Genetics, Fifth Edition. New York.
Rasner, A./Himmelreicher, R.K./Grabka, M.M. and Frick, J.R. (2007): Best of Both Worlds: Preparatory Steps in Matching Survey Data with Administrative Pension Records; The Case of the German Socio-Economic Panel and the Scientific Use File Completed Insurance Biographies 2004. SOEPpaper No. 70.
Rendtel, U. (2010): Teaching and Statistical Training. [In this publication].
Rendtel, U. and Harms, Th. (2009): Weighting and Calibration for Household Surveys. In: Lynn. P. (Ed.): Methodology of Longitudinal Surveys. Chichester.
Riediger, M. (2010): Experience Sampling. [In this publication].

Riediger, M./Schmiedek, F./Wagner, G.G. and Lindenberger, U. (2009): Seeking Pleasure and Seeking Pain: Differences in Prohedonic and Contra-Hedonic Motivation From Adolescence to Old Age. Psychological Science 20 (12), 1529-1535.

Rodgers, J.L./Bard, D.E./Johnson, A./D'Onofrio, B. and Miller, W.B. (2008): The Cross-Generational Mother–Daughter–Aunt–Niece Design: Establishing Validity of the MDAN Design with NLSY Fertility Variables. Behavioral Genetics 38, 567-578.

Romeu Gordo, L./Motel-Klingebiel, A./Wurm, S. (2009): SOEP as a Source for Research on Ageing – Issues, Measures and Possibilities for Improvement. RatSWD Working Paper No. 83.

Sauer, C./Auspurg, K./Hinz, Th./Liebig, S. and Schupp, J. (2009): Die Bewertung von Erwerbseinkommen: methodische und inhaltliche Analysen zu einer Vignettenstudie im Rahmen des SOEP-Pretest 2008. DIW Data Documentation No. 44.

Schäfer, Ch./Schräpler, J.-P./Müller, K.-R. and Wagner, G.G. (2005): Automatic Identification of Faked and Fraudulent Interviews in Surveys by Two Different Methods. Schmollers Jahrbuch 125 (1), 183-193.

Schoon, I. (2006): Risk and Resilience. Cambridge.

Schräpler, J.-P./Schupp, J. and Wagner, G.G. (2010): Individual and Neighborhood Determinants os Survey Nonresponse – An Analysis Based on a New Subsample of the German Socio-Economic Panel (SOEP), Microgeographic Characteristics and Survey-Based Interviewer Characteristic. SOEPpaper [forthcoming].

Schupp, J. and Wagner, G.G. (1995): Die Zuwanderer-Stichprobe des Sozio-oekonomischen Panels (SOEP). Vierteljahrshefte zur Wirtschaftsforschung 64 (1), 16-25.

Schupp, J. and Wagner, G.G. (2010): Zum „Warum“ und „Wie“ der Erhebung von (genetischen) ‚Biomarkern' in sozialwissenschaftlichen Surveys. SOEPpapers on Multidisciplinary Panel Data Research No. 260.

Schupp, J./Siegel, N.A./Erlinghagen, M./Stegmann, T. and Wagner, G.G. (2008): Leben außerhalb Deutschlands – Eine Machbarkeitsstudie zur Realisierung von Auslandsbefragungen auf Basis des Sozio-oekonomischen Panels (SOEP). SOEPpapers on Multidisciplinary Panel Data Research No. 120.

Schupp, J./Spieß, C.K. and Wagner, G.G. (2008): Die verhaltenswissenschaftliche Weiterentwicklung des Erhebungsprogramms des SOEP. Vierteljahrshefte zur Wirtschaftsforschung 77 (3), 63-76. [English Version is forthcoming as a SOEPpaper in 2010].

Schupp, J. (2009): 25 Jahre Sozio-oekonomisches Panel – Ein Infrastrukturprojekt der empirischen Sozial- und Wirtschaftsforschung in Deutschland. Zeitschrift für Soziologie 38 (5), 350-357.

Siedler, Th./Schupp, J./Spieß, C.K. and Wagner, G.G. (2009): SOEP as „Reference Data“. Schmollers Jahrbuch 129 (2), 367-374.

Siegrist, J. (2001): Zukünftige Aufgaben der Sozial-Epidemiologie. In: Mielck, A. and Bloomfield, K. (Eds.): Sozial-Epidemiologie. Weinheim/Munich.

Spinath, F.M. (2010): Genetically Sensitive Sample Designs. [In this publication].

Teddlie, Ch. and Tashakorri, A. (2003): Major Isues and Controversies in the Use of Mixed Mothods in the Social and Behavioral Sciences. In: Tashakorri, A. and

Teddlie, Ch. (Eds.): Handbook of Mixed Methods in the Social and Behavioral Research. Thousand Oakes/London/New Delhi.
Uhlig, J./Solga, H. and Schupp, J. (2009): Bildungsungleichheiten und blockierte Lernpotenziale: Welche Bedeutung hat die Persönlichkeitsstruktur für diesen Zusammenhang? Zeitschrift für Soziologie 38 (5), 418-440.
Wagner, G.G./Frick, J.R. and Schupp, J. (2006): Enhancing the Power of Household Panel Studies: The Case of the German Socio-Economic Panel Study (SOEP). Schmollers Jahrbuch 127 (1), 139-169
Wissenschaftsrat (2003): Strategische Forschungsförderung – Empfehlungen zu Kommunikation, Kooperation und Wettbewerb im Wissenschaftssystem. http://www.wissenschaftsrat.de/texte/5654-03.pdf.
Wissenschaftsrat (2009): Science-Policy Statement on the Status and Future Development of the Socio-Economic Panel (SOEP). http://www.wissenschaftsrat.de/texte/9503-09.pdf.
Wissenschaftsrat (2010): Overview of International Household Panel Studies. RatSWD Working Paper 135.
Wrzus, C. et al. (2010): Do Associations Between “Sleeping Well” and “Feeling Well” Become Weaker with Age? Results from the SOEP and an Ambulatory Monitoring Study. Berlin. [Manuscript in preparation].

3. Geodata

Tobia Lakes

Contact:

Tobia Lakes
Humboldt Universität zu Berlin
Geographisches Institut
Unter den Linden 6
10099 Berlin
Germany
e-mail: Tobia.Lakes[at]geo.hu-berlin.de

Abstract

Empirical data can be characterized by its precise location in space and time. An estimated 80 percent of all data contain such spatio-temporal references and are termed geodata. This paper starts with the question: how does it benefit the socio-economic sciences to use geodata and the spatial dimension respectively? In the following report, a multidimensional approach is taken to outline the current situation of geodata and the use of spatial techniques in Germany. The ever-growing volume and variety of available geodata is given particular emphasis. Data security is another issue of great importance when using geodata. Furthermore, the present developments in price and user concepts, accessibility, technical standards, and institutionalization are addressed. A number of challenges facing the field of geodata are identified including open access to geodata, data security issues, and standardization. The main challenge, however, seems to be cooperation and exchange between the rather segregated fields of geoinformation and the information infrastructure. Furthermore, the German Census in 2011 is identified as a major challenge for the acquisition and management of geodata. Geodata and the use of spatial techniques are a field that is rapidly developing due to technological developments as well as due to a recent surge in public interest. The benefits they hold for socio-economic research should be exploited in the future.

Keywords: geodata, geoinformation, Web GIS, geodata infrastructure, spatial techniques

1. Introduction

Many of the foremost research issues to emerge in recent years – climate change and its impact on human life, megacity development, disparities between the rich and the poor, environmental justice, and security – have one element in common: they benefit from empirical study and therefore rely critically on empirical data (IPCC 2007; UN Habitat 2008; EC 2008). Empirical data about households, the sources and targets of migration, meteorological data, the accessibility of education, and the range of environmental pollution are examples of where empirical data are needed, data that can be characterized by a precise location in space and time. An estimated 80 percent of all data contain such spatio-temporal references and are termed geodata. The use of geodata and spatially explicit techniques is well-established in geography or spatial planning as well as in specific subdisciplines such as social geography or economical geography (Longley et al. 2005). However, until recently, the benefit of using geodata and geoinformatics techniques to develop spatially explicit approaches has rarely been exploited in the socio-economic sciences and policy-related research (Goodchild and

Janelle 2004). In addition to regional data approaches (i.e., the report by Grözinger and Matiaske in this publication), the explicit linkage of data to a location has become an area of growing interest, for example, in the context of the next German Census in 2011.

What are the added benefits for the socio-economic sciences in using geodata and the spatial dimension respectively? First of all, geodata is data like every other dataset, hence *spatial data can provide additional information* and therefore should be valued and included in empirical research. In Germany, a large pool of geodata already exists that is continuously being enlarged – something that is described in detail below – and is waiting to be exploited by new users. Second, geodata can add fundamental advantages by *allowing for visualizations in the form of maps and database search algorithms based on location*. Third, the spatial information *makes it possible to integrate various datasets via the spatial location* and examine possible interrelationships between datasets. In a recent study, for example, the life satisfaction approach is used to evaluate air quality: individual-level panel and high-resolution SO2 data are combined to identify the effect of SO2 concentration on life satisfaction, housing rents, and the total willingness-to-pay for improvements in air quality (Lüchinger 2009). Directly georeferenced data are also of particular interest as a means of creating comparability to repeatedly collected data based on modified statistical units. The final and perhaps most important benefit is that *spatial analyses* enable the inclusion of the context via concepts of proximity, range, containment, overlap, adjacency, or connectedness. The visualization and statistical analyses of these properties is one way of detecting patterns, anomalies, outliers, and sometimes even causation, and thus to generate new insights. Of course, underlying processes cannot be detected, but they can sometimes be approximated. In a recent study, for example, the factors influencing choices about tertiary education among recently graduated students was modeled: the distance between the students' households and the universities turns out to be very significant (Spiess and Wrohlich 2008). These spatially explicit analyses can be extended to *spatio-temporal modeling approaches* that aim at modeling spatial processes in time, including probable future developments such as land-use change at the interface between the human and environmental systems (Lakes and Müller 2008).

The access to and analysis of geodata on national, European, and global scales are necessary in order to undertake the type of cross-disciplinary research required for developing policy-relevant strategies. Such data therefore can be regarded as not only beneficial, but essential. In Germany, "geoinformation" is now regarded as one of the most important crosssectional technologies of this century and a policy field with an outstanding future (Bundesregierung 2008).

2. Status quo: Geodata and spatial techniques

An outline of the research that would potentially benefit from geodata shows not only that the available geodatasets are of interest, but also that there are techniques available for handling and exploiting the spatial dimension of geodata. This paper describes the data and techniques that exist in Germany in a national and international context. The discussion takes a multi-dimensional approach, addressing data availability, factors influencing data availability (accessibility, technical standards, price and user concepts, data security, and institutionalization), and spatial techniques.

2.1 Present situation of geodata

The amount and the variety of available geodata in Germany is continuously expanding. In terms of content, geodata can be divided into spatial base data and spatial thematic data that are acquired and provided by official or private sources. The *spatial base data* contain general topographical and property information and hence offer the basis for most research studies.

Excursus:
Geodata can either contain a **direct spatial reference** or an **indirect spatial reference**. In the case of a **direct spatial reference** – such as the geodata used most frequently in Germany, the Gauss-Krüger and ETRS 89 systems – the information about the location is defined by two- or three-dimensional coordinates within a coordinate reference system. Data that contain an **indirect spatial reference** include systems closer to everyday human experience, such as administrative areas, postal addresses, or place names. In order to digitally process the complexity of real world objects, they must be generalized and simplified. Discrete objects and continuous fields are two approaches to modelling space that each correspond to a data model – the vector or raster model respectively. Points (e.g., trees, cities), lines (e.g., roads, rivers) and areas or polygons (e.g., city-parcels, administrative boundaries) are examples of the **vector model. Raster data** consists of cells within a rectangular grid, such as the remote sensing data of airborne or satellite systems.

The acquisition and management of spatial base datasets is predominately the task of public organizations, and are accessible at the Federal Agency for Cartography and Geodesy (BKG, *Bundesamt für Kartographie und Geodäsie*) and at the survey administrations of the Länder and municipalities, in keeping with the federal system in Germany. The two most important Germany-wide standardized spatial base datasets are:

- **The Authoritative Topographic-Cartographic Information System,** (ATKIS, *Amtliches Topographisch-Kartographisches Informationssystem*) that includes digital landscape models, digital terrain models, digital topographical maps, digital orthophotos, digital street names, geographic names, and administrative boundaries.

- **The Authoritative Real Estate Cadastre Information System** (ALKIS, *Amtliches Liegenschaftskataster Informationssystem*) that contains the Real Estate Map, the Real Estate Book, and the Official House Coordinates.

Within these standardized systems, objects are classified according to a specific hierarchical object catalogue and numbering system, such as 2000 – Residential and Infrastructural Areas (*Siedlungsflächen*), with subclasses including 2100 – More Developed Areas (*baulich geprägte Flächen*), 2111 – Areas with Residential Structures (*Wohnbauflächen*), and 2121 – Sport Facilities (*Sportanlage*). In recent years, *spatial base data from official sources have been increasingly replaced* by new methods of data provision. On the one hand, the *geodata acquired and provided by a worldwide user community via the Internet* is of growing importance. The OpenStreetMap Project is an example of this Wikipedia-style open information source that can be used and updated by anyone in a collaborative way. On the other hand, the influence of *private data providers* within the geodata market is also beyond question. Up-to-date road networks data (e.g., Navteq, Tele-Atlas), household address data, aerial photos, and satellite data are increasingly provided by private companies. While aerial photos are still predominately produced by German companies (e.g., Hansa Luftbild), the market for satellite data is a global one, as seen, for example, in the way that data from satellites being distributed worldwide are useful for local studies. Over the past few years, too, *remote sensing data has captured user interest*. In part initiated by the arrival of new internet-based technologies such as GoogleEarth, it has become obvious that aerial photos as well as satellite data constitute a good data source even at first sight, not to mention with the possibilities opened up by sophisticated remote sensing data for analyses in social science research (Rindfuss and Stern 1998; Goodchild and Janelle 2004). The variety of remote sensing datasets available is growing, each offering specific advantages depending on the objective and context of the study. One can choose, for example, between very high spatial resolution (Quickbird) or very high temporal resolution (Rapid Eye) versus satellite data covering very large areas (Landsat TM). A few companies in Germany have specialized in providing remote sensing data, such as GAF or EuroMap. Remote sensing datasets are also available at the German Remote Sensing Data Center (DFD, *Deutsches Fernerkundungsdatenzentrum*) of the German Aerospace Center (DLR, *Deutsches Zentrum für Luft- und Raumfahrt*).

In addition to the topographical and property information of geospatial base data, the focus in research and application has predominately been on *spatial thematic data*. This can cover a wide variety of fields, including environmental data, employment data, or business data depending on the specific research objective. On the side of government data, these are collected and used *at the federal level, by the Länder, and by municipalities*. While some federal agencies are experienced in working with a spatially explicit approach – such as the German Federal Ministry for the Environment, Nature Conservation, and Nuclear Safety or the Federal Office for Building and Regional Planning – others traditionally provide data either without or with only very aggregated spatial references (e.g., the Federal Employment Agency or the Federal Statistical Office). Particularly important official sources of spatial thematic data for researchers resides at the communal level, in areas such as planning, forestry, environment, statistics, and the police. Of particular interest in this regard is the German Census in 2011, which will provide macro-census information that can be precisely linked to location for further analysis. In addition to government data, a large amount of geospatial thematic data is collected by the research sector itself. Furthermore, both non-profit organizations and commercial data providers hold and provide a significant amount of spatial thematic data. It is particularly in the area of commerce that data needs are not sufficiently covered by public data provision, including branch specific information, as well as data on building, communication, and lifestyle and socio-demographic, market, and consumer data (Fornefeld et al. 2003). Another important source of data comes from the field of *geomarketing*, with companies such as Pitney Bowes Inc. that offer worldwide services in direct marketing and postal services based on a geographic information system (GIS), MapInfo.

2.2 The present situation and factors influencing data availability

The most decisive challenge confronting the current use of spatial base data and spatial thematic data in Germany and elsewhere is the accessibility of a large amount of available geodata, which is distributed in several places and acquired and provided by different sources. The problems inherent in this situation are well recognized by the scientific, business, administration, and political communities in the field of geoinformation. A number of measures have been taken to enhance accessibility. First, *geodata infrastructures and geodatabases have been established on different levels within government agencies and other institutions*. The aim of these geodata infrastructures (GDI) is to improve the accessibility and use of available geodata. Geodata infrastructure projects are very often connected to the management of geodatabases and internet-based geoportals for user-friendly data provision. A Germany-wide national geodatabase has been set up to hold all the geodata needed for legal purposes, government activities, economic development, and research. Its purpose is to provide access to data from different sectors of the federal government, the *Länder*, and municipalities via standardized web services (as a first step, federal geodata is now accessible through the website: www.geoportal.bund.de). A second significant instrument for enabling access to spatial thematic data is the German Environmental Information Portal, PortalU, which allows users to search for environmental information from 120 public agencies and organizations via thematic, spatial, and temporal criteria.

One important issue to emerge recently is the need to create *price and user concepts of geodata that will promote transparent and market-oriented development* without putting the ownership or responsibility for the data into question. The basic approach that the federal government has taken is to charge fees for the use of public geodata based on the cost of data provision (Bundesregierung 2008). The primary building blocks of this policy of data access include the introduction of eGovernment procedures, (e.g., ePayment), legal guidelines such as the Geodata Access Act (*Geodatenzugangsgesetz*) and the Environmental Information Act (*Umweltinformationsgesetz*), and the fees structure established by the Working Committee of the Surveying Authorities of the States of the Federal Republic of Germany (AdV, *Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland*).

The technical difficulties arising from varying specifications and formats have been an ongoing challenge. This has been dealt with by developing standards for geoinformation technology that also need to fit eGovernment strategies. The system of independent access to geodata of different levels requires the definition and adoption of standards based on European (CEN) and international standards (ISO, Open Geospatial Consortium). In 2007, the GDI-DE (Spatial Data Infrastructure Germany) introduced an architectural concept for geodata infrastructure, which contains information on functionality, services, and technology for developing the future infrastructure in Germany.

Data security is of ongoing importance for all types of data. Aside from being generally regulated by the foundational Freedom of Information and Reuse of Public Sector Information Act (*Informationsfreiheits- und Informationsweiterverwendungsgesetz),* geodata presents a specific case for which the issue of personal rights is particularly sensitive. Up to now there has been no consistent approach developed for finding a balance between the release and non-disclosure of geodata. In general it depends on the extent to which the personal right of the persons concerned are invaded (Karg and Weichert 2007). Google's recent activity photographing street panoramas for use in 3-D city models available online has provoked new discussions about data security. Specifically relevant to geodata is the Environmental Information Act (*Umweltinformationsgesetz*), which is the national manifestation of European guidelines on public access to information about the environment. Remote sensing data represent a particular type of data with many advantages due to the area-wide, comparable, and up-to-date information they provide on multiple aspects of the earth's surface. However, this data may also pose critical data security risks, as addressed in the recent Satellite Data Security Act (*Satellitendatensicherheitsgesetz*). The issue of data security is of great importance, but it is very complex as it pertains to the provision of social and economic data and, therefore, cannot be fully explored in this paper (see the report by Schaar on data protection and by Metschke on record linkage in this publication).

The measures that have been taken to assure the accessibility and efficient use of geodata have been strengthened by *major achievements in institutionalizing cooperation* between different levels and types of public administration in Germany, as well as within the economic and research sectors that use geodata. To name only the most important: the Interdepartmental Committee for Geoinformation (IMAG, *Interministerieller Ausschuss für Geoinformationswesen*), the Steering Committee of Spatial Data Infrastructure Germany (*Lenkungsgremium GDI-DE*), the Commission for Geoinformation Business (*GIW-Kommission*), the AdV,[1] and the "Deutschland Online" initiative, in coordination with the Working Group of State Secretaries Responsible for eGovernment in the Federation and the *Länder* (*Arbeitskreis der Staatssekretäre für eGovernment des Bundes und der Länder*).

Not only in Germany but also internationally, the cross-border exchange of geodata is of growing importance. The *international interoperability of geodata and geoinformation* has been particularly strengthened by the European INSPIRE initiative, which has developed a set of basic guidelines for interoperability in terms of geodata management and provision as well as for the development of a European geodata infrastructure.

Furthermore, three major *innovations for newly available data* are of importance in the international context: Galileo, the European satellite navigation system, will provide the basis for the future referencing of geodata, the localization and positioning of objects. In 2013, Galileo is expected to offer positioning data which will be of interest for multiple user groups. A central platform for the future usage of Galileo has been set up with the "Forum for Satellite Navigation" by the Federal Ministry of Transport, Building, and Urban Affairs. A second initiative is the Global Monitoring for Environment and Security, which is supposed to integrate terrestrial, satellite, airborne, maritime, and other data sources for environmental policy, climate measures, and sustainable development, as well as for humanitarian, development, and security relevant issues. It is a joint initiative of the European Commission and the European Space Agency. Third, the Group on Earth Observation (GEO) should be mentioned, which was initiated in 2005 to build a "Global Earth Observation System of Systems" (GEOSS) that offers

1 http://www.adv-online.de.

better access to earth observation information. A central access point (GEO Portal) as well as a catalogue service (GEO Clearinghouse) is envisaged.

2.3 Present situation in spatial techniques

Geodata can be used like any other dataset in a statistical software application to extract the thematic information it holds. However, in addition to the techniques mentioned above for geodata access and distribution, specific spatial extensions are needed to exploit the spatio-temporal dimension of geodata. The specific type of professional software that offers the required spatial techniques is called GIS (geographical information system). It is a computer system used for capturing, management, analyzing, and displaying Geodata. GIS includes hardware, software, networks, standards, and protocols for data handling and analysis (Longley et al. 2005).

Apart from proprietary software, *Open Source GIS and databases* increasingly provide an interesting alternative (e.g., PostGIS, PostgreSQL, GRASS). In addition, *spatial extensions for frequently used database systems* are being employed, such as Oracle Spatial and new releases of SQL Servers. Whether a professional GIS is needed or whether basic tools suffice depends to a large extent on user interests and requirements. Of particular interest are *Web Services*, which offer basic spatial services without the need for an installed GIS software on the user's PC. While basic functions such as map visualization of decentralized servers via Web Services are well-established, more sophisticated techniques are still in development and need further research. Finally, *freely available Internet tools* are a growing sector, including sponsored user-community portals, such as Picasa – which offers a service to place photos in Google Earth – or portals financed by advertisements, such as Map24.de, which offers navigation data and services. These go along with *navigation and mobile services* that have reached operational application level. Accompanied by the development of GPS sensors in mobile phones and widely spread mobile phone cameras, these open up new opportunities for location-based services as well as for research.

3. Future developments

Looking ahead from the current situation there are manifold developments on the horizon concerning geodata and spatial techniques. Only a few examples will be addressed here (see also Bundesregierung 2008).

The *amount of available datasets will continue to grow*, and the variety of thematic, spatial, and temporal characteristics will increase. The development of *new data acquisition* technologies in particular will contribute to the growing amount of data. These include the more frequent use of positioning systems and new remote sensing technologies, to name only the German development of Terra-SAR-X, RapidEye, and EnMAP. Geodata will increasingly be acquired by both public and private data providers. Hence, new forms of *public-private partnerships and cooperation* for data acquisition, including collaborative web-based initiatives, need special attention. A project of major significance in terms of georeferenced data acquisition and provision within Germany is the next *German Census in 2011*.

The already initiated *development of internet-based access points for geodata, or geoportals will continue*, whether they are government or business portals. The overall objective of building up a national geo-database with the goal of establishing a demand-oriented geodata supply will be a major task for the future. The Geodata Center of the Federal Agency for Cartography and Geodesy (*Geodatenzentrum of the Bundesamt für Kartographie und Geodäsie*) envisages a further extension of the www.geoportal.bund.de, with the current access to data from the Federal State expanding to include data from the *Länder* and municipalities as a geoportal for Germany. In addition to the development of geoportals for official data, business geoportals will also grow in number (MICUS 2008).

Another recent trend that will continue is the creation of portals that are not limited to data or metadata, but that include *Web Services*, enabling direct access to data and thematic map visualizations via the Internet without requiring specialized software. In keeping with this, the principle of decentralized data within specific organizations and centralized data provision for the user will continue. With the growing importance of the Internet, coordinated efforts with *eGovernment*, such as ePayment, will be of interest for geodata. Technical standards need further attention not only within the field of geoinformation but also beyond specialized science and as part of the eGovernment concept developed by the Working Group of the GDI-DE.

In addition to new spatial Internet-based technologies, spatial extensions of widely used database systems support the trend, *"GIS Goes Mainstream."* Hence, the user community is expected to grow constantly, spurred on by free and open source products. Furthermore, *new spatial techniques* in professional GIS software offer the potential to integrate different datasets or to support spatio-temporal modeling.

Cost and usage concepts will continue to be an important issue for public sources of data – a centralized tool for assessing geodata fees at all levels of government as well as for private data providers will be a long-term objective. According to the federal government (Bundesregierung 2008), the cost for data acquired by tax money will be limited to the actual cost of provision, which will enlarge the geodata user group.

Discussions about *data security and the need to legally and consistently define data access rights* will continue to increase, for example in the context of the 2011 Census. A consultancy rating of the most frequently used data and a categorization based on data security relevance will be undertaken (Bundesregierung 2008). A draft of geodata access legislation (*Geodaten-zugangsgesetz*) is under development, aiming at the free provision of geodata and geodata services by the federal government and the European Union providing there is no further business usage of the data.

In the near future, *Germany will also have to address the requirements of international developments* (INSPIRE,[2] GMES,[3] GEOSS[4]) and take concrete actions to fulfill them. The INSPIRE guideline has to be transferred into German law by May 15, 2009. The spatial datasets proposed in the annex of the INSPIRE program must be implemented by 2019 in all levels of public administration.

4. European and international challenges

In comparison with Germany, there are similar and yet diverging tendencies that prevail in international contexts (Fornefeld et al. 2003). The strategy of developing geodata infrastructures as a way of optimizing access to geodata from public sources through interministerial organization is an ongoing task in European countries and beyond, for example in the United States. While in the US this goal has reached a well-established level, in most other countries it remains in a development phase. Since internationalization occurs both in the field of private data acquisition as well as in the provision of spatial techniques, the global market is converging. In addition, an increasing number of international guidelines in Europe and across European borders require the comparable use of thematic geodata, such as the European Union Water Framework Directive. International initiatives, such as the INSPIRE guideline, GMES, and GEOSS will thoroughly change the handling of geodata in Germany and Europe. The resulting potential for geodata usage in research and business must be exploited as much as possible. These cross-border developments are leading to the increasing importance of international exchange of geodata beyond the infrastructure of specific government ministries.

2 Infrastructure for Spacial Information in Europe.
3 Global Monitoring for Environment and Security.
4 Global Earth Observation System of Systems.

5. Conclusions and recommendations

The collection of geodata and the use of spatial techniques comprise a rapidly developing field due to developments in data technology and methodology, as well as the new level of political attention it has attracted. This makes it difficult to keep track of current developments, although it is more important than ever to regularly analyze the situation and develop recommendations. After all, it needs to be stressed that geodata is data like any other dataset and the artificial separation between geodata and meteorological, juridical, and demographic data for example is no longer adequate, considering that 80 percent of all information has a spatial reference. It is only the combination of information that offers multiple benefits. Hence, most of the issues addressed by the other reports in this publication are also relevant to geodata. The following section will present selected recommendations concerning geodata, the factors influencing geodata accessibility, and spatial techniques.

5.1 Geodata and factors influencing geodata

The amount and variety of geodata is constantly growing. Hence, the main challenge is to *provide access to geodata* in such a way that they can be combined with other forms of data to provide information for research and public policy (Bundesregierung 2008). A *geodata infrastructure based on geoportals* is very significant, but new sources of internet-based and private data provision must also be considered (MICUS 2003; Bundesregierung 2008).

In terms of the data, a *reliable update of official sources of spatial base data* is lacking. Although a five-year rhythm may be what is envisioned, in reality it is often less frequent. Furthermore, the provision of historical data is also of relevance to longitudinal studies, in the best case, comparable data. This may be a task for public agencies since it is not covered by private data providers. Data gaps in *area-wide coverage of spatial base data* in Germany (not to mention Europe or even beyond) need to be closed, for example, in the very different quality of urban and rural topographical data. In addition, research requires comparable data; hence, *object catalogues for spatial base data and spatial thematic data should be developed in greater detail.* Since linking geodata has been identified as a major task, conversion codes between different datasets should be available. Spatial reference is one key to possible data integration; therefore, *data should be equipped with a spatial reference as far as possible*. While indirect reference via postal codes or election districts might be more feasible, the spatial outline and position can change. Therefore, direct spatial reference seems to present a better solution since it leaves data security issues either untouched or at least manageable (see Schaar in this publication) and only then can the spatial benefits of thematic data such as official statistics, Microcensus, and particularly the 2011 Census be fully exploited. *More thinking must be done about the 2011 Census* in order to enable the linkage and integration of census data with further datasets based on explicit georeferencing via the personal address, and at the same time preventing the extraction of individual-level information via techniques such as data aggregation on the grid level or the thematic-object based level.

User rights, particularly for reuse and further use of data, as well as *regulations for fees and price models* in Germany, Europe, and internationally need to be pursued.

Consistent and up-to-date technical standards continue to be an important subject.

With the growing amount of data available, and the enhanced combination of data from different sources, *quality measures for geodatasets* must be developed. Users miss reliable measures of available datasets for data from both official sources and privately offered data. Imperfect data is better than no data; however, it is essential to be able to estimate the possible limits of explanatory power.

5.2 Spatial techniques

On one side, sophisticated spatial analysis and integration of geodata with additional data within interdisciplinary projects open up new research opportunities and need to be exploited. On the other side, mapping techniques for non-professionals offering user-oriented techniques for their specific tasks are a challenge and require an overview of existing software and tools for non-professional spatial analysis supported by best-case studies.

New techniques such as the freely available GoogleMaps, Picasa, or GoogleEarth open up a wide field for *internet-based data access and tools* that need to be exploited. However, issues around the *quality and reliability of publicly available and free tools* need to be investigated.

5.3 Politics

The main challenge in this area seems to be the *need for exchange between the rather segregated fields of geoinformation and the information infrastructure*, such as the German Data Forum (RatSWD). Parallel developments in terms of geodata infrastructures, geodata portals, and geoinformation management should be integrated into a national approach for the overall information infrastructure. A *round table on geodata and regional data together* *with the German Data Forum (RatSWD)* should be established to bring together the different methods and initiatives taken by data providers and researchers.

The primary issues to be addressed include user and price concepts, data security, and technical standards development to further enhance data exchange. Also, exchange is needed between the public, private, and research sectors in the field of geoinformation in order to get new impulses for and from research.

An awareness of the great potential of geodata and the use of spatial techniques is the prerequisite for their successful use in transdisciplinary, if not interdisciplinary, socio-economic and policy-related research. *Joint research projects* along with a presence in journals and media should be initiated to exploit the potential of integrating geodata in integrated analyses. Integrative modules across departmental (and thematic data) boundaries within universities may be one possibility. *International exchange* should include successful initiatives of geodata usage in the context of scientific infrastructures such as the Center for Spatially Integrated Social Science (US)[5]; SEDAC, the Socioeconomic Data and Applications Center (US)[6]; or the Center for Geoinformation (Ireland)[7].

5 http://www.csiss.org/
6 http://sedac.ciesin.columbia.edu/
7 http://ncg.nuim.ie/

References:

Bundesregierung (2008): Zweiter Bericht der Bundesregierung über die Fortschritte zur Entwicklung der verschiedenen Felder des Geoinformationswesens im nationalen, europäischen und internationalen Kontext. Drucksache 16/10080.

EU Commission (EC) (2008): Global Monitoring for Environment and Security (GMES): we care for a safer planet. Communication from the Commission to the European Parliament, Council, EESC and CoR. OM(2008)748.

Fornefeld, M. et al. (2003): Der Markt für Geoinformationen: Potenziale für Beschäftigung, Innovation und Wertschöpfung. Studie im Auftrag des Bundesministerium für Wirtschaft und Arbeit. Düsseldorf.

Goodchild, M.F. and Janelle, D.G. (2004): Thinking spatially in the Social Science. In: Goodchild, M.F. and Janelle, D.G. (Eds.): Spatially Integrated Social Science. New York.

INSPIRE (2007): Richtlinie des Europäischen Parlamentes und des Rates zur Schaffung einer Geodateninfrastruktur in der Europäischen Gemeinschaft vom 14. März 2007 (2007/2/EG) (ABl. L 108/1 v. 25. April 2007).

IPCC (2007): Impacts, Adaptation & Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). Cambridge.

Karg, M. and Weichert, T. (2007): Datenschutz und Geoinformation. Studie im Auftrag des Bundesministeriums für Wirtschaft und Technologie. Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein. Hannover.

Lakes, T./Müller, D. and Krüger, C (2009): Cropland change in southern Romania – a comparison of logistic regressions and artificial neural networks. Landscape Ecology 24 (9), 1195-1206.

Longley, P. et al. (2005): Geographic Information Systems and Science. New York.

Lüchinger, S. (2009): Valuing Air Quality Using the Life Satisfaction Approach. Economic Journal 119 (536), 482-515.

MICUS (2003): Der Markt für Geoinformationen: Potenziale für Beschäftigung, Innovation und Wertschöpfung. Studie im Auftrag des Bundesministerium für Wirtschaft und Technologie.

Rindfuss, R.R. and Stern, P.C. (1998): Linking Remote Sensing and Social Science: The Need and the Challenges. In: Liverman, D. et al. (Eds.): People and Pixels: Linking Remote Sensing and Social Science. Washington.

Spiess, K. and Wrohlich, K. (2008): Does distance determine who attends a University in Germany? SOEPpapers on Multidisciplinary Panel Data Research 118.

UN-HABITAT (2008): The State of the World's Cities 2008/9. United Nations. [Published 2009].

4. Regional Data

Gerd Grözinger and Wenzel Matiaske

Contact:

Wenzel Matiaske
Helmut-Schmidt-University /
University of the Federal Armed Forces Hamburg
Faculty for Economics and Social Sciences
Holstenhofweg 85
22043 Hamburg
Germany
e-mail: matiaske[at]hsu-hh.de

Abstract

The spatial dimension is an increasingly important aspect of research in the social sciences, as a new wave of recent publications shows. In this work, intra-national comparisons have proven to be as fruitful as the more common international analyses, and regional characteristics are shown to have considerable influence on individual behavior. This movement has been fostered by methodological advances, e.g., in multilevel techniques. Germany has a good basic infrastructure for spatial analysis providing easy access to official and semi-official data. In addition, both scientific researchers and commercial marketing firms are active in collecting valuable information, in some cases on a very detailed local level – even down to just a handful of households. However, there is ample room for improvement: huge existing datasets (e.g., PISA-E) are not open for spatial analysis purposes; in many cases sufficient regional information is not available (e.g., on criminal behavior); and systematic oversampling in sparsely inhabited areas to allow additional regional analysis is relatively uncommon.

1. Research questions

Regional analyses of social behavior have a long tradition in the social sciences. In sociology, Durkheim's famous book on suicide was one of the earliest works addressing the impact of regional characteristics – religiosity, urbanization, and social control – on individual anomic behavior (Durkheim 1952). The basic idea of modeling regional characteristics as independent variables influencing social behavior has been taken up repeatedly since then. But early sociology is also known for studies that concentrate on the regional context, embedding social relationships in a group or community (*Gemeinschaft*). Whyte's well-known case study of "Street Corner Society" in Boston's Little Italy brings the spatial dimension into the tradition of the Chicago School (Whyte 1943). Economic theory contains both approaches to regional analysis as well – the use of regional features as independent variables affecting individual behavior, and their use as dependent variables defining social contexts. In one of the first such economic studies, Marshall emphasized the importance of regional characteristics in shaping industrial districts and their role as a core determinant of economic development (Marshall 1898). Regional aspects have also long been discussed from a business management point of view as a problem of site selection: von Thünen's concentric model of land use may be read as an early precursor of industrial location theory (Thünen 1826).

Both strands of theory still profoundly influence the debate. Their impact has been magnified by theoretical and methodological developments. On the theoretical side, recent work has attempted to more clearly distinguish between macro- and micro-levels of social behavior (Alexander et al. 1987). In the words of contemporary rational choice theory, the context of action on the macro-level of social systems – regions in our context – constrain the "logic of the situation." Regional conditions on the macro-level influence how individual or corporate actors choose goal-oriented actions on the micro- (or meso-) level. The logic of aggregation also leads back from the micro- to the macro-level of the social system. There, it shows emergent effects that are not always collective goods created by the "invisible hand," but may also include situations of collective damage (Coleman 1990).

These theoretical developments correspond with methodological progress. Hierarchical regression models – fixed and random effect models (the terminology differs between sociology and economic methodology) – have had a particularly important impact. These models take the hierarchical structure of the analysis explicitly into account: behavior or attitudes are not only explained by individual properties (micro-level) but also by regional circumstances (macro-level) (Snidjders and Bosker 1993). Examining the different degrees of freedom on the various hierarchical levels increases the reliability of the test statistic. These models often include cross-level interactions. Depending on the subject of analysis, different estimators are available (Blien 2005). However, there is a danger of overextending such analyses and thereby falling victim to the "ecological fallacy" problem. To model the macro-constraints of the logic of the situation, individual data and structural (regional) data must either directly mirror each other or be linked in another way.

Whereas this group of multilevel models is predestined to analyze the macro-micro link, there is no standard model available to describe the micro-macro link. In many cases one can use a microeconomic model of market exchange to analyze the logic of aggregation, typically to study price or power effects (Braun 2008). But the assumption of more or less perfect markets does not always hold true. Therefore, a multiplicity of methods like game theory models, Markov models, and simulation studies are employed. Currently social network analysis is being used more and more in the multilevel context (Wasserman and Faust 1994). Furthermore, multivariate techniques developed or modified for ecological analysis, e.g., restricted or detrended correspondence analysis and other eigenvalue techniques or multi-dimensional scaling, seem to be extremely useful in the case of regional data (Leyer and Wesche 2005).

In addition to its pure scientific interest, the analysis of regional data has always been of interest to policy-makers. After World War II, the collection of German regional data experienced its first renaissance in the late 1960s and early 1970s (at least in West Germany). This was connected with the new public interest in planning policy (Schäfers 1973). Scientific organizations responded to the rising demand with increased professionalization, and as a result many current research activities date back to this decade. The section of the German society of sociology (DGS, *Deutsche Gesellschaft für Soziologie*) on urban and regional sociology was officially established in 1975, the same year that another user group with regional interests was founded – the planners' association (*Informationskreis für Raumplanung)* – with now over 1,500 members. In 1976, the German Economic Association (*Verein für Socialpolitik)* followed with the establishment of a commission on regional theory and regional policy.

With the deepening and enlargement of the European Union, new themes and issues have arisen. Instruments like the Cohesion Fund, the Social Fund, and the Regional Development Fund all need regular data for implementation and evaluation of measures. International comparisons have been facilitated by common definitions of regional units: in 2003, a framework on the definitions of NUTS (Nomenclature des unités territoriales statistiques) was legally enacted in the EU based on past cooperation and experiences among the national statistical offices (Brunner 2008).

Interest in the regional dimension increased further with German unification. Given the strong and persistent differences between East and West, the social sciences began to seek explanations of different development paths (e.g., Bertram et al. 2000). Public interest has increased as well, leading to numerous activities. A huge German national atlas project has been launched in which in twelve volumes with CDs offer a comprehensive view of life in the German regions (*Leibniz-Institut für Länderkunde*). This has been conducted mostly on the level of spatial planning regions (ROR, *Raumordnungsregionen*). Also on the ROR level an online survey was conducted and has served as a basis for many comparisons in the media (Faßbender and Kluge 2006).

The labor market is of key importance for policy making. In Germany, the labor market is characterized by extensive regional disparities, especially in terms of the extent of employment and unemployment, but also in terms of income levels. The Institute for Employment Research (IAB, *Institut für Arbeitsmarkt- und Berufsforschung*) collects and analyses labor market data – employment statistics, unemployment statistics, the IAB Establishment Panel – on different levels (Blien et al. 2001). IAB contains its own research department on regional labor markets and also coordinates a regional research network among the former state employment offices (Eckey et al. 2007).

In specialized spatial and regional research, economic research, and current business administration research – that is, in the development of regional clusters – the region is understood as an independent object of research. However, in behaviorally oriented research fields, the macro-level – i.e., aggregate data on the social environment – is linked with micro-level data on behavior, attitudes, and preferences (see Grözinger and Matiaske

2005; Grözinger et al. 2008 for a summary of current studies). These fields usually use micro and macro data derived from different sources. Below we will highlight research facilities providing such data and discuss characteristic aspects of spatial data and problems of bringing it together with individual data. The potential capacity of datasets containing small-scale coordinates is huge, especially by fusion of data. Matchable datasets are not only from public or scientific sources, but also – especially in commercial research – primarily from other sources.

It is primarily private enterprises that have an interest in regional economic or marketing policies. For such decisions, they frequently make use of databases provided by private research facilities and business consultancies. The Society for Consumer Research (GfK, *Gesellschaft für Konsumforschung*) in Nuremberg is one of the biggest European providers of geomarketing data and support analysis, planning, and evaluation of locations in Germany and abroad. Their regional data based on point-of-sale surveys and socio-demographic and sector-specific data are of interest not only for practical purposes but also for general research. GfK's indicators for purchasing power can be analyzed at all regional levels down to individual street sections (Lochschmidt 2005). Similar data are provided by other companies; Microm, for example, calculates "social milieus" from such data, which are used by the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel)* to complement the survey data (Kueppers 2005).

2. Status quo: Databases and access

For research in the tradition described above, where data are needed for planning purposes, a good basic regional data infrastructure is provided by official sources. This is partly done by the Federal Statistical Office, often in cooperation with the Statistical Offices of the German *Länder*, and a special federal research unit, the Federal Office for Building and Regional Planning (BBR, *Bundesamt für Bauwesen und Raumordnung*). The BBR publishes widely-used regular reports on the structure of regional differences in Germany (2005) and forecasts for future development (2004).

Data from the Federal Statistical Office and the BBR can be usually found on the following hierarchically ordered levels (numbers show the respective amount of entities):

- States (*Bundesländer*): 16
- Regional Planning Units (ROR, *Raumordnungsregionen*): 97
- Cities and Counties (SG, *Städte* und *Gemeinden*): 439.

Three data compilations should be highlighted. All are convenient for scientific use since they are available on CD/DVD; both come without user restrictions, are more or less reasonably priced (approx. €75) and regularly updated. In addition, there are linked websites where the variables are defined and maps provided,[1] or where data updates can even be downloaded.[2]

- INKAR[3] (Bundesamt für Bauwesen und Raumordnung 2007) with approximately 800 indicators
- Statistik Regional (Statistische Ämter des Bundes und der Länder 2008b) with approximately 1,100 indicators
- Statistik Lokal (Statistische Ämter des Bundes und der Länder 2008a) with more than 300 indicators.

In many cases, these datasets fulfill the interest of social researchers in regional background information. Where appropriate, differentiation along the lines of gender and migration is often included. In the case of unemployment, INKAR provides the female unemployment rate, the absolute number, the percentage, and the trend. For foreigners, rate, percentage, and trends are given.

Regional information can often be broken down further into an even more detailed grid. Some of the German states are rather large in population und therefore consist of different administrative areas (*Regierungsbezirke*). Many, especially bigger cities have information broken down on boroughs (*Stadtteile*/*Bezirke*). And on the most detailed level, every municipality provides a land registry (*Kataster*). Whereas such data can only be obtained from the regional or local administrations, detailed general information about the approx. 12,000 municipalities (*Gemeinden*) is conveniently available on a special DVD:

- Statistik Lokal (Statistische Ämter des Bundes und der Länder 2007a).

However, it must be mentioned that the statistical units used are defined either following political traditions or for planning purposes, which are also based on political boundaries. For scientific questions, one therefore has to deal with huge variations in both the population and area, which can make analysis rather difficult. In many contests, the number of inhabitants – an important piece of information – ranges from:

- On the state level, the minimum is 0.7 million (Bremen); the maximum 18 million (North Rhine-Westphalia).

1 http://www.raumbeobachtung.de
2 http://www.regionalstatistik.de
3 Indikatoren und Karten zur Raum- und Stadtentwicklung.

- On the ROR level, the minimum is less than 300,000; the maximum is Berlin with over 3 million.
- On the SG level, the minimum is barely over 50,000; the maximum again Berlin with over 3 million.

Besides these official statistical entities, there are other principles of classification, mostly used by scientific or marketing institutions for sampling, such as:

- ZIP codes (*Postleitzahlen*)
- Electoral districts (*Wahlbezirke*)
- Telephone area codes (*Telefonvorwahlen*)
- Labor market regions (*Arbeitsmarktregionen*)
- License plates (*Autokennzeichen*)
- Households (*Haushalte*).

Some of them can also be (dis)aggregated according to the needs of the user. For example, the ZIP code has five digits and is hierarchically ordered. It can therefore be used in its entirety or just the first or first two, three, or four digits.

Households are the smallest unit of information sampled by marketing institutions. Although not set out in law, it is generally understood that to meet German privacy protection mandates all local statistical information has to be based on at least five households (Mietzner 2005). It is permitted to combine information on such clusters, however. On this basis, information collected using consumer marketing techniques provides a wealth of data that can be assembled to describe a certain area according to sociological criteria.

Whereas both of the lists above rely on the principle of physical proximity, it is also possible to classify regional entities by common properties. Frequently used principles in the social sciences are:

- Number of inhabitants
- Income levels
- Types of urbanization.

The latter category can be differentiated according to the needs and the levels of regional aggregation. The BBR, for example, offers a classification of three general regional types of settlements, seven types on the ROR level, and nine on the SG level.

The SOEP deserves special mention. It is by far the most widely-used dataset for social science questions in Germany. Registered users with appropriate data safety measures can obtain access to a version on the ROR level. On the SOEP premises in Berlin, one can even work with a version on the state level.[4]

4 http://www.diw.de/english/soep/29012.html

Basically, every special dataset that contains information on the sampling point is a potential source for aggregation to some regional level. For example, one can estimate the regional religious distribution (not available from official statistical sources) on the basis of a survey (Dülmer 2005). But the regionalized sample size must exceed a critical number to provide reliable estimators (Bliese 2000).

Finally, not all data is available on the appropriate regional level, as one may reasonably expect. Three examples are found in areas that are the subject of widespread public debate: (1) the Criminal Statistic is not published regularly or in comprehensive form (Bundeskriminalamt 2008), (2) the PISA-E study, which in Germany refers to the national supplement to the international PISA study, is not provided for secondary analysis below the state level, (3) the outcome of the IQ tests of young men in connection with the military draft system is also seen as private property although it can be successfully linked to regional variables (Ebenrett et al. 2003).

3. Future developments and challenges

The historical dimension of regional characteristics is frequently underestimated, often exceeding the periods of official data. A recent study on the impact of social capital analyzed regional crime rates using historical data on household, population, occupation, etc. as instrumental variables, from 1795 to 1970 (Akcomak and Weel 2008). The Netherlands Volkstellingen Archive (Dutch census) provides this data and more.[5] It would be an improvement if Germany's historical regional data from different sources – church and land registers, historical reports, etc. – were also properly edited and made available for quantitative analysis.

Looking over the border leads to another area for future research improvement. The European NUTS classification has been available for several years, which facilitates comparative research. However, this classification system is more appropriate for planning purposes than for social research. On the European level, a future challenge will therefore be the development of a more detailed classification system, based on the needs of social scientists.

Generally, there is a trade-off between a highly detailed classification system and data privacy. In particular, providing household data for geomarketing may have the negative side-effect of discriminating against the inhabitants of certain areas ("scoring"). In the long run, the effect may not only lead to intra-regional migration and a self-perpetuating vicious cycle of

5 http://www.volkstellingen.nl/en/

discrimination; it may also increase public distrust in data collection and endanger the legitimacy of social science research. Furthermore, problems may arise in the reliability of measured datasets when the data from different reference levels are brought together or methods of data fusion are applied (Zimmermann 2005).

4. Conclusions and recommendations

The following list contains the most important measures to improve the infrastructure on regional data in Germany. From an organizational point of view, the most relevant are:

- In addition to its publications, the BBR has a huge amount of unpublished data on different regional levels on file. They should provide at least a regularly updated list of these data with proper descriptions and a well-defined policy on data access for scientific purposes.
- The GESIS Data Archive, where many of the German survey data are stored, should be granted extra funding to classify all surveys according to their appropriateness for regional analysis.
- Future surveys aimed at being nationally representative should be sampled in such a way that detailed regional analysis is also possible at least on the ROR level. Due to the different population levels, this would need some systematic oversampling in sparsely inhabited areas.
- The five-household entity – currently not formalized – could be used as a basis for any detailed data structuring. Notwithstanding the aforementioned danger of illegitimate use of such information, it would be useful if marketing firms would cooperate to work out a single list of blocks that then could be used universally. Alternatively, the eight-household grid of the Microcensus – which is due to be renewed for the 2011 Census – could be used for this purpose.
- A concordance should then be provided in which the different levels and principles could be easily transferred upward (e.g., a particular ROR, ZIP code, etc. consists of certain numbers of blocks).
- Finally, the wide range of research interests in regional and geographical information from scientific, administrative, and commercial users and data producers leads to the recommendation of a roundtable where common interests could be defined. The German Data Forum (RatSWD) should initiate such a group.

References:

Akcomak, I.S. and Weel, B.T. (2008): The Impact of Social Capital on Crime: Evidence from the Netherlands. Maastricht.

Alexander, J.C./Giesen, G./Münch, R. and Smelser, N.S. (Eds.) (1987): The Micro-macro Link. Berkely.

Bertram, H./Nauck, B./Klein, T. (Eds.) (2000): Solidarität, Lebensformen und regionale Entwicklung. Opladen.

Blien, U. (2005): Die Mehrebenenanalyse regionaler Fragestellungen. In: Grözinger, G. and Matiaske, W. (Eds.): Deutschland regional. Sozialwissenschaftliche Daten im Forschungsverbund. Munich.

Blien, U./Haas, A./Hirschenauer, F./Meierhofer, E./Tassinopoulos, A./Vollkommer, D. and Wolf, K. (2001): Regionale Arbeitsmarktforschung im IAB. Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 1, 45-73.

Bliese, P.D. (2000): Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In: Klein, K.J. and Kozlowski, S.W. (Eds.): Multilevel Theory, Research, and Methods in Organizations. San Francisco.

Braun, N. (2008): Sozialkapital aus Sicht der Rational Choice Soziologie. In: Matiaske, W. and Grözinger, G. (Eds.): Sozialkapital: eine (un)bequeme Kategorie [Jahrbuch Ökonomie und Gesellschaft, Band 20]. Marburg.

Brunner, C. (2008): European Datasets: Regional and Urban Statistics. In: Grözinger, G./Matiaske, W. and Spieß, C.K. (Eds.): Europe and its Regions. The Usage of European Regionalized Social Science Data. Cambridge.

Bundesinstitut für Bau-, Stadt- und Raumforschung (2004): Raumordnungsprognose 2020. Informationen zur Raumentwicklung 3/4.

Bundesinstitut für Bau-, Stadt- und Raumforschung (2005): Raumordnungsbericht 2005. Berichte 21.

Bundesinstitut für Bau-, Stadt- und Raumforschung (2007): INKAR – Indikatoren und Karten zur Raum- und Stadtentwicklung. Bonn.

Bundeskriminalamt (2008): Polizeiliche Kriminalstatistik 2007. Wiesbaden.

Coleman, J.S. (1990): Foundations of Social Theory. Cambridge.

Dülmer, H. (2005): Die Schätzung von kleinräumigen Kontextinformationen aus Umfragedaten. In: Grözinger, G. and Matiaske, W. (Eds.): Deutschland regional. Sozialwissenschaftliche Daten im Forschungsverbund. Munich.

Durkheim, E. (1952): Suicide. A Study in Sociology. London.

Ebenrett, H.J./Hansen, D. and Puzicha, K.J. (2003): Verlust von Humankapital in Regionen mit hoher Arbeitslosigkeit. Aus Politik und Zeitgeschichte B 6-7, 25-31.

Eckey, H.-F./Schwengler, B. and Türck, M. (2007): Vergleich von deutschen Arbeitsmarktregionen. Nuremberg.

Faßbender, H. and Kluge, J. (2006): Perspektive Deutschland. Was die Deutschen wirklich wollen. Berlin.

Grözinger, G. and Matiaske, W. (Eds.) (2005): Deutschland regional. Sozialwissenschaftliche Daten im Forschungsverbund. Munich.

Grözinger, G./Matiaske, W. and Spiess, K.C. (Eds.) (2008): Europe and its Regions: The Usage of European Regionalized Social Science Data. Cambridge.

Kueppers, R. (2005): MOSAIC von microm. In: Grözinger, G. and Matiaske, W. (Eds.): Deutschland regional: Sozialwissenschaftliche Daten im Forschungsverbund. Munich/Mering.

Leibniz-Institut für Länderkunde (Eds.). (Div.): Der Nationalatlas Bundesrepublik Deutschland. Heidelberg.

Leyer, I. and Wesche, K. (2005): Ordinationsmethoden zur Analyse ökologischer Daten. In: Grözinger, G. and Matiaske, W. (Eds.): Deutschland regional: Sozialwissenschaftliche Daten im Forschungsverbund. Munich/Mering.

Lochschmidt, B. (2005): Wissen gesucht? Wissen gefunden: GfK Regionalforschung. In: Grözinger, G. and Matiaske, W. (Eds.): Deutschland regional: Sozialwissenschaftliche Daten im Forschungsverbund. Munich/Mering.

Marshall, A. (1898): The Principles of Economics. London.

Mietzner, L. (2005): Anwendungsfelder für mikrogeographische Daten im Marketing In: Sokol, B. (Ed.): Living by Numbers. Leben zwischen Statistik und Wirklichkeit. Düsseldorf.

Schäfers, B. (Ed.) (1973): Gesellschaftliche Planung. Stuttgart.

Snidjders, T. and Bosker, R. (1993): Multilevel analysis. An introduction to basic and advanced multilevel modelling. London.

Statistische Ämter des Bundes und der Länder (2008a): Statistik lokal. Gemeindedaten für ganz Deutschland. Düsseldorf.

Statistische Ämter des Bundes und der Länder (2008b): Statistik regional. Daten für die Kreise und kreisfreien Städte Deutschlands. Düsseldorf.

Thünen, J.H. v. (1826): Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie. Hamburg.

Wasserman, S. and Faust, K. (1994): Social network analysis: Methods and Applications. Cambridge.

Whyte, W.F. (1943): Street corner society: the social structure of an Italian slum. Chicago.

Zimmermann, E.J. (2005): Möglichkeiten und Grenzen der Datenfusion. In: Grözinger, G. and Matiaske, W. (Eds.): Deutschland regional. Sozialwissenschaftliche Daten im Forschungsverbund. Munich.

5. Genetically Sensitive Sample Designs

Frank M. Spinath

Contact:

Frank M. Spinath
Differentielle Psychologie und psychologische Diagnostik
Universität des Saarlandes
66123 Saarbrücken
Germany
e-mail: f.spinath[at]mx.uni-saarland.de

Abstract

Understanding the sources of individual differences beyond social and economic effects has become a research area of growing interest in psychology, sociology, and economics. A quantitative genetic research design provides the necessary tools for this type of analysis. For a state-of-the-art approach, multigroup data is required. Household panel studies, such as Understanding Society in the UK or the SOEP in Germany, combined with an oversampling of twins, provide a powerful starting point since data from a reasonably large number of non-twin relatives is readily available. In addition to advances in our understanding of genetic and environmental influences on key variables in the social sciences, quantitative genetic analyses of target variables can guide molecular genetic research in the field of employment, earnings, health, and satisfaction, as combined twin and sibling or parent data can help overcome serious caveats in molecular genetic research.

Keywords: genetics, twins, psychology, sociology, economics, heritability, environment, multigroup design, BHPS, SOEP
JEL Classification: B40, B49, C51, C83

1. Motivation (research questions)

The present report argues that household panel studies that were initiated for the analysis of household income offer a unique opportunity to study the importance of genetic and environmental influences on variation across individuals in key areas of social, economic, and psychological research. It should be noted that, from a genetic point of view, the "environment" includes all influences other than inheritance – a much broader use of the term than is usual in the behavioral sciences. By this definition, environment includes, for example, prenatal events and biological events such as nutrition and illness, not just family socialization factors. Similarly, in this paper, the term environment encompasses a wide variety of biological, natural, social, and economic environments.

Research questions like the origin of earnings variation, life satisfaction, health, and their interrelation with psychological variables such as personality can be addressed. By disentangling the interplay of genes and environmental factors (social scientists may call those effects "socio-economic"), the analyses of genetically informative samples make it possible to derive more accurate estimates of social and economic effects on behavior than social and economic studies, which ignore the influence of genes. A recent Special Issue on "Society and Genetics" in the journal *Sociological Methods & Research* illustrates the growing interdisciplinary readiness to stop treating the differences across individuals at birth as a black box (Guo 2008). In a similar vein, Diewald (2008) argues that genetically sensitive research designs can be of immense value to sociological research in providing evidence to test sociological hypotheses against competing explanations. As a result, more sophisticated methodological approaches in the social sciences should become best-practice, acknowledging and involving genetically informative samples.

Since the inherent design of household panels includes participants of varying genetic and environmental similarity (biological full siblings, biological half-siblings, parent-child dyads, and to a smaller extent adoptive children, twins, and triplets), such panel studies are an ideal – and up to now underutilized – starting point for state-of-the-art quantitative genetic analyses. This report illustrates how household panel studies enriched with an oversampling of twin participants can even address dynamic gene-environment interplay.

This report focuses on the quantitative genetic approach. Molecular genetic research strategies (e.g., genetic association and candidate gene studies) constitute a different methodological approach that is not addressed here (for an outlook on possible combinations of both methods, see section 5 below). Due to the fact that genetically sensitive sample designs are a relatively new topic in the discussion of the research infrastructure and future needs in social and economic research, this report also provides a basic theoretical and methodological background to the understanding of quantitative genetic analyses.

The benefit of utilizing genetically informative data is not limited to research of a predominantly psychological nature, and the number of studies on the etiology of key variables in economic and social research is growing. For example, twin data indicates that basic political attitudes like liberalism and conservatism are likely to be heritable (Hatemi et al. 2007). In two further independent twin studies, Fowler, Baker, and Dawes (2008) showed that voter turnout and political participation have very high heritabilities.

In a recent multigroup analysis, Björklund, Jäntti, and Solon (2005) studied the influences of nature (genes) and nurture (socio-economic characteristics) on earnings variation using observed sibling correlations in earnings for nine types of sibling pairs: monozygotic (MZ) twins reared together, monozygotic twins reared apart, dizygotic (DZ) twins reared together, dizygotic twins reared apart, non-twin full siblings reared together, non-twin full siblings reared apart, half-siblings reared together, half-siblings reared apart, and adoptive siblings. On the basis of this variety of sibling types in the analyses, the authors were able to estimate models that involved less restrictive assumptions and provided opportunities to examine the sensitivity of their results to variation in modeling assumptions; namely, the introduction of nonzero GE correlation, of estimates for the genetic relatedness of DZ twins, non-twin full siblings, half-siblings, and adoptive siblings, and varying sibling correlation in environmental influences. The results turned out to be sensitive to flexibility in modeling the variation across types of sibling pairs in the similarity of their environments. Even the smallest estimate of the genetic component of earnings variation, however, suggested that it accounts for about 20 percent of earnings inequality among men and more than 10 percent among women. The largest environmental influence was of the non-shared variety, which is in line with the results of many quantitative studies on personality. In the present study, even among MZ twin brothers, an estimated 64 percent of their earnings variation was explained by neither genetic nor shared environmental resemblance.

The latter study is also a good example of how quantitative genetic methods can be used to target key research topics in labor economics, that is, understanding the sources of earnings inequality and accounting for the rise in earnings inequality that has occurred in most developed countries over the last quarter-century (Katz and Autor 1999). Inequality research focusing on the role of family and community origins ties in particularly well with the quantitative genetic understanding of shared and nonshared environmental factors. The basic idea is that if family and community origins account for a large portion of earnings inequality, siblings will show a strong similarity in earnings; if family and community background hardly matters at all, siblings will show little more resemblance than would randomly selected unrelated individuals.

2. Theoretical and methodological background

Results from classical twin studies have made a remarkable contribution to one of the most dramatic developments in psychology during the past few decades: the increased recognition of the important contribution of genetic factors to virtually every psychological trait (Plomin et al. 2008). However, enriching classical twin studies by data from additional dyads (non-twin siblings, parents-children, etc) can improve behavioral genetic analyses for the following reasons.

The classical twin design compares the phenotypic resemblances of identical or MZ and fraternal or DZ twins. MZ twins derive from the splitting of one fertilized zygote and therefore inherit identical genetic material. DZ twins are first-degree relatives because they develop from separately fertilized eggs and are 50 percent genetically identical on average. It follows that a greater within-pair similarity in MZ compared to DZ twins suggests that genetic variance influences the trait under study.

To disentangle and to quantify the contributions that genes and the environment make to human complex traits, data are required either from relatives who are genetically related but who grow up in unrelated environments ("twin adoption design"), or from relatives who grow up in similar environments but are of differing genetic relatedness ("twin design"). Most twin studies that have been conducted over the past 80 years are of the latter type. Only two major studies of the former type have been conducted, one in Minnesota (Bouchard et al. 1990) and one in Sweden (Pedersen et al. 1992). These studies have found, for example, that monozygotic twins reared apart from early in life are almost as similar in terms of general cognitive ability as are monozygotic twins reared together, a result suggesting strong genetic influence and little environmental influence caused by growing up together in the same family. These influences are typically called *shared environment influences* because they refer to environmental factors contributing to the resemblance between individuals who grow up together. *Nonshared environmental influences*, on the other hand, refer to environmental factors that make individuals who grow up together different from one another.

One reason why a predominant number of twin studies have utilized the twin design instead of the twin adoption design is that twins typically grow up together, thus it is much easier to find a large number of participants for the classic twin study. In humans, about 1 in 85 live births are twins. The numbers of identical and same-sex fraternal twins are approximately equal. That is, of all twin pairs, about one-third are identical twins, one-third are same-sex fraternal twins, and one-third are opposite-sex fraternal twins. The rate of twinning differs across countries, increases with maternal age, and may even be inherited in some families. Greater numbers of fraternal twins are the result of the increased use of fertility drugs and in vitro fertilization, whereas the rate of identical twinning is not affected by these factors.

Comparing the phenotypic resemblance of MZ and DZ twins for a trait or measure under study offers a rough estimate of the extent to which genetic variance is associated with phenotypic variation of that trait. If MZ twins resemble each other to a greater extent than do DZ twins, the heritability (h^2) of the trait can be estimated by doubling the difference between MZ and DZ correlations, that is, $h^2 = 2(r_{MZ} - r_{DZ})$ (Falconer 1960). *Heritability* is defined as the proportion of phenotypic differences among individuals that can be attributed to genetic differences in a particular population. It should be noted that for a meaningful interpretation of twin correlations in the described manner, a number of assumptions have to be met: the absence of assortative mating for the trait in question, the absence of G(enotype) × E(nvironment) correlation and interaction, and the viability of the Equal Environments Assumption. A more detailed discussion of these assumptions as well as the effects of variation attributable to chorionicity differences is available elsewhere (Spinath 2005), so a short introduction should suffice here:

Assortative mating describes nonrandom mating that results in similarity between spouses and increases correlations and the genetic similarity for first-degree relatives if the trait under study shows genetic influence. Assortative mating can be inferred from spouse correlations which are comparably low for some psychological traits (e.g., personality), yet are substantial for others (e.g., intelligence), with average spouse correlations of about 40 (Jensen 1998). In twin studies, assortative mating results in underestimates of heritability because it raises the DZ correlation but does not affect the MZ correlation. If assortative mating is not taken into account, its effects are attributed to the shared environment.

Gene-Environment (GE) correlation describes the phenomenon that genetic propensities can be correlated with individual differences in experiences. Three types of GE correlations are distinguished: passive, evocative, and active. Previous research indicates that genetic factors often contribute substantially to measures of the environment, especially the family environment (Plomin 1994). In the classic twin study, however, GE correlation is assumed to be zero because it is essentially an analysis of main effects.

Gene-Environment (G × E) interaction is often conceptualized as the genetic control of sensitivity to the environment. Heritability that is conditional on environmental exposure can indicate the presence of a G × E interaction. The classic twin study does not address G × E interaction and the classic twin model assumes the equality of pre- and postnatal environmental influences within the two types of twins.

Finally, the classic twin model assumes the equality of pre- and postnatal environmental influences within the two types of twins. In other words, the *Equal Environments Assumption (EEA)* assumes that environmentally caused similarity is roughly the same for both types of twins reared in the same family. Violations of the EEA in the sense that MZ twins experience more similar environments than DZ twins would inflate estimates of genetic influences.

3. Methodological advances and new research questions

The comparison of correlations between MZ versus DZ twins can be regarded as a reasonable first step in our understanding of the etiology of particular traits. To model genetic and environmental effects as the contribution of unmeasured (latent) variables to phenotypic differences, Structural Equation Modeling (SEM) is required. Analyzing univariate data from MZ and DZ twins by means of SEM offers numerous advances over the mere use of correlations, including an overall statistical fit of the model, tests of parsimonious submodels, and maximum likelihood confidence intervals for each latent influence included in the model.

The true strength of SEM, however, lies in its application to multivariate and multigroup data. During the last decade powerful models and programs to efficiently run these models have been developed (Neale et al. 2003). Extended twin designs and the simultaneous analysis of correlated traits are among the most important developments that go beyond the classic twin study (Plomin et al. 2008).

Multigroup designs using a wider variety of sibling types bring more power to bear on quantitative genetic analyses (e.g., Coventry and Keller 2005). For example, it is useful to include non-twin siblings in twin studies to test whether twins differ statistically from singletons, and whether fraternal twins are more similar than non-twin siblings.

Multigroup designs also enable the application of more general (i.e., less restrictive) models, such as relaxation of the EEA or the introduction of GE correlation, as well as to examine the sensitivity of results to variations in modeling assumptions. Furthermore, results from multigroup analyses are less prone to systematic method bias and sampling error.

4. Status quo: Databases and access

More than 5,000 papers on twins were published during the five years from 2001 to 2006, and more than 500 of these involve behavior (Plomin et al. 2008). The value of the twin method explains why most developed countries have twin registers (Bartels 2007).

About a decade ago, Boomsma (1998) published the first paper in a series aimed at giving an overview of existing twin registers worldwide. A short description of 16 registries in nine European countries was presented. At the time, these registries had access to over 350,000 pairs providing a resource for genetic-epidemiological research. In the years 2002 and 2006, special issues of the scientific journal *Twin Research and Human Genetics* documented further progress in this field. Currently, worldwide registers of extensive twin data are being established and combined with data from additional family members, offering completely new perspectives in a refined behavioral genetic research (Boomsma et al. 2002).

However, datasets required for multigroup analyses are typically not readily available, especially in countries without official twin or extensive population registers such as Germany. Even in Sweden, home of one of the most extensive twin registers in the world, samples for multigroup data have to be matched from different sources (Björklund et al. 2005). In the study described in the introduction, data on non-twin siblings came from random samples of the Swedish population drawn by Statistics Sweden whereas the twin sample came from the Swedish Twin Registry (Medlund et al. 1977).

The situation in Germany is even more complicated because a central twin register is not available. The Bielefeld Longitudinal Study of Adult Twins (BiLSAT; Spinath et al. 2002), the first large-scale twin study in Germany, was initiated in 1993. Twins were recruited through newspaper and media announcements as well as twin organizations. A telephone hotline was installed and twins who expressed interest in the BiLSAT were informed about the aims of the study and the approximate time required to complete the questionnaire sets. Names, addresses, date of birth, and self-reported zygosity of twin pairs who decided to participate were entered into the database. Within six months, approximately 1,500 twin pairs were enrolled in the BiLSAT and questionnaire data was collected for approximately 75 percent of the initial sample. The twins' age varied between 14 and 80 years (M = 32, SD = 13 years) and the sample was heterogeneous with regard to education and employment status. As is typically observed with voluntary twin samples, females participated more frequently than males and MZ twins participated more frequently than DZ twins.

In two more recent twin studies (Spinath and Wolf 2006), a different recruitment procedure aimed at reducing self-selective sampling was applied: through individual inquiries at registrations offices in two German federal states (North Rhine-Westphalia and Thuringia), contact information on persons with the same birth name, the same birthday, and also the same birthplace was gathered. These requests resulted in 36,574 addresses of potential twin pairs – adult twins as well as parents of twins. From this list, people in the relevant age-groups for the planned projects (birth cohorts 1995–1998 and 1955–1970) were selected. After matching the provided addresses with data found in public telephone directories, 1,014 adult twins and 715 families with children twins were contacted by phone in 2005. An additional 3,832 households were contacted by mail. First contact by phone turned out to be more efficient, because almost two-thirds of all personally contacted twins agreed to participate as compared to only 26 percent (children sample) and 10 percent (adult sample) participations when first contact was made by mail. The total number of false positive contacts (people born on the same day and with the same surname who claimed not to be twins) was relatively small, yielding 2.4 percent for the children sample and 4.3 percent for the adult sample and rendering the chosen way of recruitment feasible.

5. Future developments

Interdisciplinary efforts to collect data of relevance to psychologists, sociologists, and economists alike, using genetically sensitive designs are highly desirable since the challenges of recruiting a multigroup sample can be met with greater ease in a collaborative effort combining household panel study data and data from traditional twin samples.

Studies such as the British Household Panel Study (BHPS) and the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*), representative longitudinal studies of private households providing information on all household members and covering a range of topics including employment, earnings, health, and satisfaction indicators, are ideal for many reasons:

First of all, household panels naturally include biological full-siblings, biological half-siblings, parent-child dyads, and to a smaller extent adoptive children, twins, and triplets.

Second, an explorative analysis showed that with nearly 11,000 households and more than 20,000 persons sampled in the SOEP, data from a reasonably large number of non-twin relatives is readily available. In the SOEP data collected in 2007, for example, it was possible to identify 2,209 individuals from 983 families who have at least one sibling as well as 179 adopted children. With 47 individuals in twin or triplet pairs from 20 families, the number of twins who are already enrolled in SOEP is not large enough for a multigroup analysis. However, the recruitment of twins who participate in the assessment of SOEP variables and who could ultimately be enrolled in the regular longitudinal assessment offers a unique opportunity to enrich an already powerful dataset to allow for quantitative genetic analyses.

Studying the families of identical twins, for example, has come to be known as the families-of-twins method (D'Onofrio et al. 2003). When identical twins become adults and have their own children, interesting family relationships emerge. For example, in families of male identical twins, nephews are as related genetically to their twin uncle as they are to their own father. Furthermore, the cousins are as closely related to one another as half siblings are. Studying twins and their family members is a powerful method in differentiating and quantifying environmental and genetic processes underlying associations between family-level risk factors and child adjustment to environmental stimuli. In addition to refined modeling opportunities for estimating genetic and environmental influences on target variables in such samples, repeated measurements provide the opportunity to address genetic and environmental influences to stability and change over time as well as covariance among variables of interest. To summarize: in principle, household panel studies which trace individuals with their families and households for decades are ideal databases for such studies. However, up to now the number of twins assessed in such studies is too small.

Finally, twin and multigroup samples are valuable for determining behavioral areas in which molecular genetic research efforts and candidate gene studies are more likely to be fruitful. As an example, Fowler and Dawes (2008) recently reported that a polymorphism of the MAOA gene significantly increases the likelihood of voting. Additional household information as well as twin and parent data combined (also known as the Nuclear Twin Family Design, NTFD), allow for a separation of environmental factors shared only between siblings (S) and familial environmental factors passed from parents to offspring (F).

Two possible ways to establish an oversampling of twins (i.e., to arrive at a sufficiently large number of twin participants) in Germany have already been outlined above. These possibilities can be combined with a third recruitment strategy: the screening of people by survey research. In cooperation with TNS Infratest, a feasibility check was carried out in which a random sample was contacted via telephone.[1] As part of a larger interview, respondents were asked whether they happened to be a member of a twin pair. If this was the case, a second question addressed the willingness to be contacted and informed about a twin research project. A total of 17,529 interviews yielded 312 members of twin pairs (1.8 percent). From this sample, 149 individuals (48 percent) agreed to be contacted by phone or mail. The twins' age varied between 14 and 75 years (M = 43, SD = 16 years). In contrast to the voluntary twin sample in BiLSAT mentioned above, male and female twins agreed to be contacted with equal frequency.

The fact that twin and non-twin sibling pairs need to be matched in a pairwise fashion requires the introduction of suitable pointer variables into the dataset. Quantitative genetic analyses also require zygosity information for same-sex twin pairs. The best way to determine twin zygosity is by means of DNA markers (polymorphisms in DNA itself). If a pair of twins differs for any DNA marker, they must be fraternal because identical twins are identical genetically. If a reasonable number of markers are examined and no differences are found, it can be concluded that the twin pair is identical. Physical similarity on highly heritable traits such as eye color, hair color, or hair texture, as well as reports about twin confusion are also often used for zygosity determination. If twins are highly similar for a number of physical traits, they are likely to be identical. Using physical similarity to determine twin zygosity typically yields accuracy of more than 90 percent when compared to genotyping data from DNA markers (e.g., Chen et al. 1999).

1 This study is supported by a BMBF grant (Grant Number 01UW0706).

6. Conclusions and recommendations

Understanding the sources of individual differences – compared to social and economic effects – has become a research area of growing interest in psychology, sociology, and economics. A quantitative genetic research design provides the necessary tools for this type of analysis. For a state-of-the-art approach, multigroup data is required. Household panel studies, such as the SOEP in Germany or BHPS in UK,[2] combined with an oversampling of twins, provide a powerful starting point since data from a reasonably large number of non-twin relatives is readily available.

Quantitative genetic analyses of target variables can guide molecular genetic research in the field of employment, earnings, health, and satisfaction, and combined twin and sibling or parent data can help overcome serious caveats in molecular genetic research.

The implementation of a pilot assessment of key socio-economic variables in a special sample of MZ and DZ twins that is comparable to BHPS or SOEP is highly recommended. Initial data collection in the twin sample including zygosity diagnosis can be realized online to minimize attrition. A total of approximately 400 twin pairs of each group of twins (that is, MZ, same-sex DZ, and opposite-sex DZ twins) enrolled in such a pilot assessment can provide a meaningful basis for the development of a more refined strategic plan, such as the integration of a twin cohort into the regular interview-based assessment in the British panel study Understanding Society and SOEP.

2 Where the new panel "Understanding Society" with a larger number of households will provide even better research opportunities.

References:

Bartels, M. (2007): An update on longitudinal twin and family studies. Twin Research and Human Genetics 10, 3-12.

Björklund, A./Jäntti, M. and Solon, G. (2005): Influences of nature and nurture on earnings variation: a report on a study of various sibling types in Sweden. In: Bowles, S./Gintis, H. and Osborne, M. (Eds.), Unequal chances: family background and economic success. New York.

Boomsma, D.A. (1998): Twin registers in Europe: an overview. Twin Research 1, 34-51.

Boomsma, D./Busjahn, A. and Peltonen, L. (2002): Classical twin studies and beyond. Nature Reviews Genetics 3, 872-882.

Bouchard Jr, T.J./Lykken, D.T./McGue, M./Segal, N.L. and Tellegen, A. (1990): Sources of human psychological differences: the Minnesota study of twins reared apart. Science 250, 223-228.

Chen, W.J./Chang, H.W./Lin, C.C.H./Chang, C./Chiu, Y.N. and Soong, W.T. (1999): Diagnosis of zygosity by questionnaire and polymarker polymerase chain reaction in young twins. Behavior Genetics 29, 115-123.

Coventry, W.L. and Keller, M. C. (2005): Estimating the extent of parameter bias in the classical twin design: A comparison of parameter estimates from Extended Twin-Family and classical twin designs. Twin Research and Human Genetics 8, 214-223.

D'Onofrio, B.M./Turkheimer, E.N./Eaves, L.J./Corey, L.A./Berg, K./Solaas, M.H. and Emery, R.E. (2003): The role of the children of twins design in elucidating causal relations between parent characteristics and child outcomes. Journal of Child Psychology and Psychiatry and Allied Disciplines 44, 1130-1144.

Diewald, M. (2008): Zwillings- und Adoptivkinder-Stichproben für soziologische Analysen? Zum Ertrag verhaltensgenetischer Ansätze für sozialwissenschaftliche Fragestellungen und Erklärungen [Using twin and adoptive child samples for sociological analysis? On the value of behavioral genetics approaches for social scientific issues and interpretations]. DIW Research Notes No. 27.

Falconer, D.S. (1960): Introduction to quantitative genetics. New York.

Fowler, J.H./Baker, L.A. and Dawes, C.T. (2008): The genetic basis of political participation. American Political Science Review 102, 233-248.

Fowler, J.H. and Dawes, C.T. (2008): Two genes predict voter turnout. Journal of Politics 70, 579-594.

Guo, G. (2008): Introduction to the special issue on society and genetics. Sociological Methods and Research 37, 159-163.

Hatemi, P.K./Medland, S.E./Morley, K.I./Heath, A.C. and Martin. N.G. (2007): The genetics of voting: an Australian twin study. Behavior Genetics 37, 435-448.

Jensen, A.R. (1998): The g factor: the science of mental ability. London.

Katz, L.F. and Autor, D.H. (1999). Changes in the wage structure and earnings inequality. In: Ashenfelter, O.A. and Card, D. (Eds.): Handbook of Labor Economics (Vol. 3). Amsterdam.

Medlund, P./Cederlof, R./Floderus-Myrhed, B./Friberg, L. and Sorensen, S. (1977): A new Swedish twin registry. Acta Medica Scandinavica Supplementum 60, 1-11.

Neale, M.C./Boker, S.M./Xie, G. and Maes, H. (2003): Mx: Statistical Modeling (6th Edition). VCU Box 900126, Department of Psychiatry, Richmond, 23298.
Pedersen, N.L./McClearn, G.E./Plomin, R. and Nesselroade, J.R. (1992): Effects of early rearing environment on twin similarity in the last half of the life span. British Journal of Developmental Psychology 10, 255-267.
Plomin, R. (1994): Genetics and experience: the Interplay between nature and nurture. Thousand Oaks.
Plomin, R./DeFries, J.C./McClearn, G.E. and McGuffin, P. (2008): Behavioural genetics (5th Edition). New York.
Spinath, F.M. (2005): Twin Designs. In: Everitt, B.S. and Howell, D.C. (Eds.): Encyclopedia of Statistics in Behavioral Science (Vol. 4). Chichester.
Spinath, F.M./Angleitner, A./Borkenau, P./Riemann, R. and Wolf, H. (2002): German Observational Study of Adult Twins: a multimodal investigation of personality, temperament and cognitive ability. Twin Research 5, 372-375.
Spinath, F.M. and Wolf, H. (2006): CoSMoS and TwinPaW: initial report on two new German twin studies. Twin Research and Human Genetics 9, 787-790.

6. Biological Variables in Social Surveys

Rainer Schnell[*]

* For helpful comments on an earlier draft I am indebted to Günther Heller and Johannes Kopp.

Contact:

Rainer Schnell
University of Duisburg-Essen
Institute for Sociology
Lotharstr. 65
47057 Duisburg
Germany
e-mail: rainer.schnell[at]uni-due.de

Abstract

Social scientists have long virtually ignored the biological constraints of human behavior. Yet if the prediction of behavior is considered essential to a social science, neglecting any variable that might influence human behavior is unacceptable. This paper provides examples of important biological variables and describes their measurement in social surveys.

1. Introduction

Social surveys today are collecting increasing amounts of data on biological variables that might influence social behavior. I will refer to such variables in the following as "biologically relevant variables" or "biological variables" for short. These include biometric features (e.g., fingerprints), biomarkers (e.g., cortisone levels), biomaterial (e.g., hair), and measures of anthropometric variables (e.g., body-mass index, or BMI).

Historical background. Social scientists have long virtually ignored the biological constraints of human behavior.[1] This historical development culminated in the qualitative conception of sociology as a "text science" dealing solely with how social actors understand and interpret one another. For this kind of sociology, the goal of social science is not to develop predictive models of social behavior but to reconstruct meaning. As such, quantitative and qualitative sociology do not differ methodologically but in their scientific objectives. If the prediction of behavior is considered essential for a social science, it cannot afford to neglect any variable that might influence human behavior. This paper will cite examples of important biological variables and describe their measurement in social surveys.

Biosocial surveys. The combination of questionnaire data and biological variables measured in a random sample of a population is increasingly denoted as a "biosocial survey." Such surveys have the advantage of every large sample: population parameters can be estimated even for small subgroups of a population. In general, this is impossible with the small sample sizes common in biopsychology, biology, and medicine. Furthermore, in many cases, samples in these disciplines are not random samples of a population but convenience samples of self-selected populations. Finally, most medical surveys are restricted to health variables, thereby lacking biographical data and those dependent variables of most interest for social scientists: employment history, mating behavior, value systems, and fertility. On the other hand, biological variables are usually not measured in social science

1 Steven Pinker (2002) has discussed this at length in "The blank slate".

surveys. Even studies on divorce seldom measure the obviously relevant and time-varying variables like body and face symmetries, BMI, fertility indicators, testosterone levels, etc.

Using biological variables affecting social behavior as independent variables together with sociologically relevant dependent variables in large-scale surveys will allow more detailed examination of longstanding sociological problems. More technically: the goal of including biological variables in social science population surveys is to reduce unexplained variance and the amount of misspecification in social science models.[2]

2. Increase of studies with biosocial variables in core social-science journals

Sociobiological hypotheses and biosocial surveys are still considered exotic by many social scientists, and prominent sociobiologists are often regarded with some suspicion. This will change very slowly. Two books published by the National Academy Press are of particular importance for this process. The first was the book *Cells and Surveys*, edited by Finch et al. (2001), with the rhetorical subtitle "Should Biological Measurements be included in Social Science Research?". The follow-up volume, *Biosocial Surveys*, was edited by Weinstein et al. (2008).

A review by Freese et al. (2003), appearing in the *Annual Review of Sociology*, was the beginning of a series of publications on biosocial variables in core journals of the social sciences. The *American Political Science Review* published an article on the genetic transmission of political orientations in 2005 (Alford et al. 2005), followed in 2008 by an article on genetic variations in political participation (Fowler et al. 2008). *Social Forces* published an issue in September 2006 with the editorial "The Linking of Sociology and Biology" (Guo 2006), containing four articles on biosocial variables. *Sociological Methods Research* had a "Special Issue on Society and Genetics" in 2008. Even the *American Journal of Sociology* released a special issue in 2008 on "Exploring Genetics and Social Structure" (Volume 114, Supplement 2008). Parallel to these publications, the steering groups of the large-scale panel studies in the social sciences published recommen-

2 The self-restriction of model builders on likelihood-ratios and Wald-statistics as inferior substitutes for model testing and residual diagnostics keep them forgetting about the small explanatory power of social sciences models. Even for simple problems like voting, fertility decisions or divorce, the proportional reduction of error of the model compared with the marginal distribution is rarely larger than 10 percent. After 40 years of multivariate research this is quite shameful.

dations for the inclusion of biosocial indicators in surveys (Lillard and Wagner 2006; Kumari et al. 2006).

To sum up, biosocial problems, hypotheses, and studies can now be found even in the core social science journals – at least the American ones. The technical and statistical level of these publications is still not up to the standards of the medical literature, but given sociology's longstanding neglect of biology, this was to be expected.

3. Biosocial data for social sciences applications

There are many examples of sociological problems in which biological variables set constraints for human behavior. Among them are genetic factors, variables on mating behavior, and perinatal variables. Only a few examples will be given; a complete and systematic review is still missing in scientific literature.

Genetic factors. For many traditional social science problems, empirical evidence of genetic effects has been found. Examples are suicide (Voracek and Loibl 2007), aggressive behavior (Craig and Halton 2009), and "antisocial behavior" in general (Moffitt 2005).

The list of dependent variables of social science interest for which genetic effects or gene-environment interactions have been reported is growing daily: from the frequency of life events (Bemmels et al. 2008) to economic decision making (Zhong et al. 2009) and the preference for coffee (Vink et al. 2009). Particularly interesting are genetic variations that correlate with numerous dependent behavioral variables. Another politically relevant topic in this context is attention-deficit hyperactivity disorder (ADHD): there are a considerable number of candidate genes for ADHD.[3] At the Bremen Institute for Prevention Research and Social Medicine (BIPS, *Bremer Institut für Präventionsforschung und Sozialmedizin*), the new study "German Population Based Long Term Follow Up of ADHD" was launched in July 2009. This study will track treated and non-treated children displaying ADHD over 12 years. Variables of interest are medical aspects, like symptoms of ADHD and other psychiatric diseases, as well as accidents, drug abuse, school achievement, juvenile crime, professional careers, and indicators of life quality.

Mating and marital stability. A surprising amount of research in German sociology over the last 15 years had been done on divorce. Even more surprising is the almost complete lack of biological variables associated with mating behavior in this literature. Even obvious factors, which might be

3 see Gizer et al. (2009).

varying with time, like differential attractiveness of the partners, have seldom been considered.[4] Despite the fact that many of the possibly relevant measurements (for example: BMI, facial and body symmetry, waist-to-hip ratio, fertility indicators) could have been measured easily and inexpensively,[5] these variables have been included in almost no study to date. Other variables associated with mating behavior, such as odor (Ebberfeld 2005), are much more difficult to measure within a survey context, but still not impossible. Due to technical problems and circadian effects, interpersonal and intrapersonal hormonal differences are even harder to measure within surveys. Nevertheless examples do exist in the sociological literature (for testosterone levels, see Booth et al. 2006).

Perinatal variables. Different perinatal variables have been associated with human behavior in later life. An important example is the level of intrauterine testosterone (see Manning 2002). The clinical quality of births is often accessed with the so called Apgar Score; furthermore, birth weight and size of the newborn are considered as predictors of many mortality events. There are studies on long-term effects, for example, of birth weight on cognitive development (Goosby and Cheadle 2009). Even effects of birth order have been studied, for example, with regard to school achievements (Booth and Kee 2009) and homosexuality (Blanchard 2008).

4. Biorelevant data in medical surveys

Medical surveys measure numerous variables on health status. To clarify the discussion, we should distinguish between medical surveys and examination surveys. Examination surveys usually ask medical survey respondents to visit an examination center. Due to the required technical equipment for techniques like sonography, CT, radiology, MRI, EEG and ECG, mobile examination centers have sometimes been used. These high-tech exams are hardly the most practicable measures for use in social surveys. Measurements that can be conducted by medically untrained interviewers in respondent households are of prime interest. These include respondent weight and height, waist-to-hip ratio, and blood pressure. Even more interesting for social scientists are measurements of a more general state of health, for example, grip strength with a dynamometer or a simple pulmonary function test ("peak flow meter").[6] A simple but useful test of limited mobility that is occasionally

4 see Hill and Kopp (2006).
5 see Zebrowitz (1997), Rhodes and Zebrowitz (2002), and Swami and Furnham (2008).
6 see ibid.

used in surveys of the elderly is how long it takes the respondent to pick up a pencil from the floor.

5. Bio-materials in the true meaning of the word

Blood. Perhaps the most versatile bio-material usable in surveys is blood. Many analyses can be done with venous blood. Unfortunately, to draw blood, German law requires the presence of an MD. The collection of blood samples thus faces practical restrictions. Even preparing blood samples for transport to a laboratory is an unusual task for non-medical fieldworkers and requires special training. Finally, the long-term storage of blood samples requires significant technical effort and costs. Taking blood using the "finger prick" method, where a drop of blood from a fingertip is dried on a small piece of paper (dried blood spot, DBS) is much easier. The analytical options are restricted compared to those of venous blood, but sampling, transport, and storage of the samples is considerably simpler. So far – with the exception of pure medical surveys – little is known about the general willingness to participate in blood samples and the long-term storage of the samples.

Saliva. Collecting saliva is the easiest way to obtain material for DNA analysis. Saliva may be used for other tests such as the level of cortisol (as a stress indicator or in the context of aggressive behavior; see Yu and Shi 2009) and cotinin (as an indicator of nicotine exposure; see Shahab et al. 2008). Saliva is usually collected from the mouth using a cotton swab. Today, a number of analyses are even possible on material collected with chewing gum. This method is non-invasive and has the potential to become widely accepted to collect such data in random samples of the population.

Hair. Hair and fingernails can be collected without any problems even under survey conditions. These materials can be used for the analysis of absorbed contaminants ("biomonitoring") and consumed drugs.[7]

Urine. McCadden et al. (2005) report on a random sample of 5105 men and women (aged 16-44), who were asked for a urine sample after a CAPI interview. Of these, 3628 (71 percent) agreed, and 3608 samples were collected successfully. The samples are used to screen for "chlamydia trachomatis," a sexually transmittable bacteria that causes almost no immediate but serious long-term problems in women. Another noteworthy study collected urine in a mail survey of a random sample of 21,000 Dutch men and women (age 15-29), for whom van Bergen et al. (2006) reports a response rate of

7 The book edited by Tobin (2005) gives an overview on the chemical analysis of human hair. For potential usages of other noninvasive bio-materials, see Esteban and Castano (2009).

almost 41 percent. A number of other similar studies are now available; Low et al. (2007) give an overview.

6. Long-term measurements

For studies on specific population such as overweight children or diabetics, long-term measurement instruments are used. These include instruments for recording blood pressure, heart rate, and intensity of movement (more specifically acceleration, using a device called an accelerometer).[8] Small-sized sensors like SmartPatch and SmartBand allow wireless measurements of heart rate (via WLAN), breathing rate, oxygen saturation of the blood, and temperature for 24 hours, even on infants.[9] Although such instruments are becoming much smaller, more portable, and less onerous, they still affect daily routine. Technical developments open up new perspectives every day, for example, the use of mobile phones with GPS as a substitute for accelerometers, since subjects carry mobile phones anyway. Another example is "intelligent clothing," where sensors in the clothes provide information on temperature, pulse rate, skin resistance, and transpiration (see Solaz et al. 2006).[10] For many cognitive tasks (and of course for diabetics), glucose levels throughout the day are important. A newly developed probe that can be mounted by trained persons in abdominal fat allows continuous recording of glucose levels. The corresponding electronic device is currently carried in a waist bag and barely affects daily activities (Dye et al. 2010).

7. Environmental data

Many health surveys collect samples of environmental materials to determine environmental pollution. These include samples of soil, tap water, and air. In Switzerland there is a nationwide noise map in which the objective magnitude of noise exposure is measured or interpolated (Ingold and Koepfli 2009). Such maps exist in other countries as well, but covering only particular regions.[11] With the consent of the respondents, some studies collect items

8 For accelerometers, see Puyau et al. (2004) and Murphy (2009).
9 www.intelligentclothing.com/wireless.html
10 Another example might be „intelligent shoes," where sensors measure speed or pressure distribution. An early example is the „Adidas Micropacer".
11 EU- directive 2002/49/EG (June 25th 2002) states that communities with a population over 250.000 people are committed to publish regional noise maps, see also www.lärmkarte.de.

of daily practical use, like toothbrushes, washcloths, combs, and vacuum cleaner bags.[12] In at least one older American study, household garbage was collected for response validation without the consent of the respondents (see Rathje 1984).

8. Research needs

Extensive research is needed on the use of biosocial variables in social surveys. This is especially true for problems of respondent cooperation in biosocial surveys.

Cooperation problems. Few studies exist on the willingness of respondents to cooperate in the collection of biological indicators within social surveys. If respondents correctly identify the purpose of a survey as non-medical, this will have strong effects on the perceived cost/benefit ratio of participation. Nearly nothing is known up to now on the resulting biases. Most biological variables in social surveys are measured in panel studies. Repeated participation in a panel may result in a biased remaining sample, but the repeated participation may also increase respondents' trust that their participation will not entail negative consequences. Results based on panels should therefore be treated with care when generalized to standard surveys. Furthermore it has to be taken into consideration that, as a rule, respondents (as well as scholars) react positively to most new methods: cooperation rates are initially high for most data collection modes (in person, by phone, and by the Internet), but deteriorate quickly with the widespread use of these techniques. This also seems plausible for the measurement of biosocial variables in social surveys. For this reason, experimental studies are urgently needed on response rates in the general population depending on organi-

12 The German environmental survey of 1990/1992 collected (for subsamples) respondents hair in order to measure aluminium, barium, plumb, boron, cadmium, calcium, chrome, copper, magnesium, phosphorus, platinum, strontium, thallium, zinc, caesium, palladium, uranium, vanadium as well as nicotine and cotinin. In the environmental survey of 1998 blood and urine samples were taken for "human biomonitoring". In subsamples, tap water was analysed for arsenic, plumbum, boron, cadmium, copper, nickel and zinc. Dustbags content was analysed for PCB, biocides, phtalates and triphosphates. The surprisingly short list of publications based on the survey can be found on the homepage of "Umweltbundesamt" at www.Umweltbundesamt.de/gesundheit/publikationen. More interesting for social scientists may be a volume on environmental justice by another federal agency (Bundesamt für Strahlenschutz et al. 2008).

zation conducting the survey, type of biological indicator, incentives used, and explanations of the survey given to respondents.[13]

Collecting and processing biosocial variables. In medical surveys, medically trained staff members are available for collecting and processing biological materials. Very little is known about whether medically untrained persons who do the fieldwork in social science surveys can be used for collecting biosocial information, ranging from the simple measurement of the BMI to collecting dried blood spots. Recent experiences with the low quality of paradata recorded by interviewers may raise some doubt as to the feasibility of traditionally trained interviewers collecting non-standard data. This doubt is even greater since the results of such fieldwork can hardly be controlled at this stage of research: after all, nothing is known about the data quality that can be expected under such field conditions.[14]

The standard procedure for special survey measurements with high technical demands is the use of few, but highly trained qualified interviewers. Adoption of this procedure for biological variables will result in considerable interviewer effects, since measurement errors are clustered within interviewers. Therefore, intraclass correlations are high. Usually, the effective decrease in sample size due to interviewer effects is computed by multiplication of interviewer workload with the intraclass correlation (Schnell and Kreuter 2005). High intraclass correlations multiplied with high workloads will yield a considerable underestimation of population variance. Therefore, more highly trained interviewers than usual will be needed for biosocial surveys, further increasing the cost of such surveys. Finally, neglect of these kinds of interviewer effects will increase the amount of errors of the first kind (alpha error rate) in biosocial surveys. Therefore, detailed studies of interviewer effects on biosocial variables are needed.

Long-term storage. For research with biological material, long-term storage of the samples is highly desirable. This allows the material to be tested at a later stage using analytical techniques that currently do not exist or on research problems that are still unknown. Long-term storage of biological samples creates considerable technical and logistical problems, however, and these remain unresolved, even for medical research in Germany.[15]

By comparison with other countries in Europe, the situation in Germany is disheartening: due to the large number of federal statistical agencies and the oligarchic structure of German academic medicine, the country still does not even have a mortality register, which would provide fascinating research

13 The comparison of stated cooperation in factorial surveys and actual cooperation in factorial experiments might be interesting in itself: I expect only a small amount of agreement.

14 Exceptions are Kroh (2005) and Jaszczak et al. (2009).

15 On technical requirements for the storage of human tissue see Troyer (2008). Helpful advice on storing other biomaterials can be found in *Cancer Epidemiology, Biomarkers and Prevention*, 15 (9) of September 2006.

opportunities if it were linked to samples on long-term storage in a biobank. The UK Biobank[16] is based on precisely this concept. More than 10 assessment centers will collect biosamples of 500,000 persons (at present between 40 and 69 years) across the entire United Kingdom. The resulting biodata will be combined with environmental and lifestyle data. The corresponding German project (the "Helmholtz cohort") has just completed the stage of identifying institutions willing and able to recruit participants for the study.

Data protection problems. The German Ethics Council (*Deutscher Ethikrat*)[17] published a detailed statement on the ethical problems and legal restrictions of biobanks in 2004. A special problem of biobanks results from the fact that persons could raise objections to the use of their samples for scientific projects that were not foreseen at the time of their consent to sample storage. Scientific progress may require disclosure of biological information to third parties. The German Ethics Council reminded researchers that biological samples may reveal information not only about the person from whom the sample was taken but also about his genetic relatives, perhaps even subgroups of the population or the total population of a country (2004, 109). Finally, the protection of persons unable to consent must be taken into account. The German Ethics Council noted, in conclusion, that collecting, storing, handling, and analysis of biological samples must be carried out in accordance with the protection of the individual. A corresponding legal framework has to be developed at an international level.[18]

The absence of a clear legal framework imposes considerable problems on social scientists seeking approval of biosocial projects from university ethics review boards, and resistance has to be expected, especially from other social scientists. In order to promote this kind of research, we need some successful examples of biosocial surveys – preferably not conducted by social scientists – to overcome institutional resistance. Under the current conditions in Germany, I personally consider cooperation with foreign research groups more promising.

Lack of biosocial theories for biological variables in surveys. A theoretical foundation for the use of biological variables in social surveys is lacking. Sociobiologists have proposed plausible hypotheses on generative behavior, some on hormonal differences, morbidity differences, and deviant behavior, and a few isolated results on trust, justice, risk behavior, and even voting behavior.[19] But by and large, we simply have very few theories on biological constraints of human behavior at present. Filling this research gap

16 www.ukbiobank.ac.uk.

17 Bevore 2008: *Nationaler Ethikrat*.

18 For an European discussion, see the book edited by Hayry et al. (2007).

19 The frequent publications of a small number of cases with surprisingly strong effects underscores the importance of publishing only significant effects after thorough testing. Without independent replications the statistical problems of multiple testing must be kept in mind.

will require far closer cooperation among biologists, psychologists, and social scientists than ever before. Without a corresponding new infrastructure for research, this seems impossible to me.

9. Recommendations

Inclusion of biosocial hypotheses and techniques in graduate studies. Due to the very slow adoption of new techniques in the social sciences in general and the tentative reception of sociobiological considerations in particular, the fastest way to promote biosocial research in the social sciences may be to include sociobiological theories and techniques in graduate studies and summer schools.

- In order to promote this kind of research, expertise is needed in the committees deciding on the topics in large-scale social science projects.
- The Leibniz Institute for the Social Sciences (GESIS, *Leibniz-Institut für Sozialwissenschaften*) should therefore, for the first time, include biologists and behavioral scientists on their committees.
- Since the technical details of collecting, processing, analyzing, and storing biomarkers are unknown outside the scientific fields from which they originate, appropriate training seminars should be included in the list of the standard GESIS summer schools.

Research on the willingness to cooperate. Research is necessary on respondents' willingness to cooperate in the collection of biosocial information and indicators in non-health surveys.

- We urgently need experiments on respondents' willingness to cooperate in the collection of different biomarkers, depending on the explanation given of the purpose of the survey, the organization conducting the survey and different incentives.

Funding opportunities. German research traditions make interdisciplinary research fields like sociobiology quite difficult. None of the traditional academic fields (medicine, biology, psychology, anthropology, the social sciences, etc.) consider human sociobiology a central research topic. Therefore, this seemingly exotic field is competing for research grants under relatively unfavorable conditions.

- To promote biosocial research we will need new tools for granting research proposals.

- An interdisciplinary priority program of the German Research Foundation (*Schwerpunktprogramm* of the DFG, *Deutsche Forschungsgemeinschaft*) in human sociobiology or even better on biosocial surveys would be a first step.
- Due to the resistance from German sociologists and the organizational structure of German university medicine, an EU project on human sociobiology seems more promising to me than an attempt to change German decision-making structures.

References:

Alford, J.R./Funk, C.L. and Hibbing, J.R. (2005): Are political orientations genetically transmitted? American Political Science Review 99 (2), 153-167.

Bemmels, H.R./Burt, S.A./Legrand, L.N./Iacono, W.G. and McGue, M. (2008): The heritability of life events: an adolescent twin and adoption study. Twin Research and Human Genetics 11 (3), 257-265.

Blanchard, R. (2008): Review and theory of handedness, birth order, and homosexuality in men. Laterality 13 (1), 51-70.

Booth, A./Granger, D.A./Mazur, A. and Kivlighan, K.T. (2006): Testosterone and social behavior. Social Forces 85 (1), 167-191.

Booth, A.L. and Kee, H.J. (2009): Birth order matters: the effect of family size and birth order on educational attainment. Journal of Population Economics 22 (2), 367-397.

Bundesamt für Strahlenschutz (BfS) et al. (Ed.) (2008): Umweltgerechtigkeit – Umwelt, Gesundheit und soziale Lage. Nummer 2, Umweltmedizinischer Informationsdienst (UMID). Umweltbundesamt.

Craig, I. and Halton, K. (2009): Genetics of human aggressive behaviour. Human Genetics. [Electronic prepublication on 9. June].

Dye, L./Mansfield, M./Lasikiewicz, N./Mahawish, L./Schnell, R./Talbot, D./Chauhan, H./Croden, F. and Lawton, C. (2010): Correspondence of continuous interstitial glucose measurement against arterialised and capillary glucose following an oral glucose tolerance test in healthy volunteers. British Journal of Nutrition 103, 134-140.

Ebberfeld, I. (2005): Botenstoffe der Liebe: Über das innige Verhältnis von Geruch und Sexualität. Münster.

Esteban, M. and Castano, A. (2009): Non-invasive matrices in human biomonitoring: a review. Environ Int. 35 (2), 438-449.

Finch, C. et al. (Eds.) (2001): Cells and surveys: should biological measures be included in social science research? Washington.

Fowler, J.H./Baker, L.A. and Dawes, C.T. (2008): Genetic variation in political participation. American Political Science Review 102 (2), 233-248.

Freese, J./Li, J.-C.A. and Wade, L.D. (2003): The potential relevances of biology to social inquiry. Annual Review of Sociology 29, 233-256.

Gizer, I./Ficks, C. and Waldman, I. (2009): Candidate gene studies of ADHD: a metaanalytic review. Human Genetics. [Electronic prepublication on 9. June].

Goosby, B.J. and Cheadle, J.E. (2009): Birth weight, math and reading achievement growth: a multilevel between-sibling, between-families approach. Social Forces 87 (3), 1291-1320.

Guo, G. (2006): The linking of sociology and biology. Social Forces 85, 145-149.

Hank, K./Jürges, H. and Schaan, B. (2009): Die Erhebung biometrischer Daten im Survey of Health, Ageing and Retirement in Europe. Methoden – Daten – Analysen 3 (1), 97-108.

Hayry, M./Chadwick, R./Arnason, V. and Arnason, G. (Ed.) (2007): The ethics and governance of human genetic databases. European perspectives. Cambridge.

Hill, P. and Kopp, J. (2006): Familiensoziologie. Wiesbaden, 4th edition.

Ingold, K. and Köpfli, M. (2009): Lärmbelastung in der Schweiz. Ergebnisse des nationalen Lärmmonitorings SonBase. Bern.

Jaszczak, A./Lundeen, K. and Smith, S. (2009): Using nonmedically trained interviewers to collect biomeasures in a national in-home survey. Field Methods 21 (1), 26-48.

Kroh, M. (2005): Intervieweffekte bei der Erhebung des Körpergewichts in Bevölkerungsumfragen. Das Gesundheitswesen 67, 646-655.

Kumari, M./Wadsworth, M./Blake, M./Bynner, J. and Wagner, G.G. (2006): Biomarkers in the proposed UK longitudinal household study. Technischer Bericht, Economic & Social Research Council.

Lillard, D. and Wagner, G.G. (2006): The value added of biomarkers in household panelstudies. DIW Data Documentation 14, 1-12.

Low, N. et al. (2007): Epidemiological, social, diagnostic and economic evaluation of population screening for genital chlamydial infection. Health Technol Assess 11 (8), iii-iv, ix-xii, 1-165.

Manning, J.T. (2002): Digit ratio: a pointer to fertility, behavior, and health. New Brunswick.

McCadden, A./Fenton, K.A./McManus, S./Mercer, C.H./Erens, B./Carder, C./Ridgway, G./Macdowall, W./Nanchahal, K./Byron, C.L./Copas, A./Wellings, K. and Johnson, A.M. (2005): Chlamydia trachomatis testing in the second British national survey of sexual attitudes and lifestyles: respondent uptake and treatment outcomes. Sexually Transmitted Diseases 32 (6), 387-394.

Moffitt, T.E. (2005): Genetic and environmental influences on antisocial behaviors: evidence from behavioral-genetic research. Adv Genet 55, 41-104.

Murphy, S.L. (2009): Review of physical activity measurement using accelerometers in older adults: considerations for research design and conduct. Preventive Medicine 48 (2), 108-114.

Nationaler Ethikrat (Ed.) (2004): Biobanken für die Forschung. Berlin.

Pinker, S. (2002): The blank slate: the modern denial of human nature. London.

Puyau, M.R./Adolph, A.L./Vohra, F.A./Zakeri, I. and Butte, N.F. (2004): Prediction of activity energy expenditure using accelerometers in children. Medicine and science in sports and exercise 36 (9), 1625-1631.

Rathje, W.L. (1984): ”Where’s the beef?”: red meat and reactivity. American Behavioral Scientist 28, 71-91.

Rhodes, G. and Zebrowitz, L. (Ed.) (2002): Facial attractiveness. Evolutionary, cognitive, and social perspectives. Westport.

Rylander-Rudqvist, T./Håkansson, N./Tybring, G. and Wolk, A. (2006): Quality and quantity of saliva DNA obtained from the self-administrated oragene method – a pilot study on the cohort of Swedish men. Cancer Epidemiology, Biomarkers and Prevention 15 (9), 1742-1745.

Schnell, R. and Kreuter, F. (2005): Separating Interviewer and Sampling-Point Effects. Journal of Official Statistics 21 (3), 389-410.

Shahab, L./Hammond, D./O’Connor, R.J./Cummings, K.M./Borland, R./King, B. and McNeill, A. (2008): The reliability and validity of self-reported puffing behavior: evidence from a cross-national study. Nicotine & Tobacco Research 10 (5), 867-874.

Solaz, J./Belda-Lois, J./Garcia, A./Barbera, R./Dura, J.V./Gomez, J.A./Soler, C. and Prat, J. (2006): Intelligent textiles for medical and monitoring applications. In: Mattila, H.R. (Ed.): Intelligent textiles and clothing. Cambridge.

Swami, V. and Furnham, A. (2008): The psychology of physical attraction. London.

Tobin, D.J. (Ed.) (2005): Hair in toxicology. An important bio-monitor. Cambridge.

Troyer, D. (2008): Biorepository standards and protocols for collecting, processing, and storing human tissues. In: Liu, B.C.-S. and Ehrlich, J.R. (Eds.): Tissue proteomics. New York. [Methods in molecular biology, vol. 441].

van Bergen, J./Götz, H./Richardus, J.H./Hoebe, C./Broer, J./Coenen, T. et al. (2006): Prevalence of urogenital Chlamydia trachomatis infections in the Netherlands suggests selective screening approaches. Results from the PILOT CT population study. Drugs Today (Barc), 42 Suppl A, 25-33.

Vink, J.M./Staphorsius, A.S. and Boomsma, D.I. (2009): A genetic analysis of coffee consumption in a sample of Dutch twins. Twin Research in Human Genetics 12 (2), 127-131.

Voracek, M. and Loibl, L.M. (2007): Genetics of suicide: a systematic review of twin studies. Wiener Klinische Wochenschrift 119, 15-16, 463-475.

Weinstein, M. et al. (Eds.) (2008): Biosocial surveys. Washington.

Wong, D.T. (Ed.) (2008): Salivary diagnostics. Ames.

Yu, Y.-Z. and Shi, J.-X. (2009): Relationship between levels of testosterone and cortisol in saliva and aggressive behaviors of adolescents. Biomedical and Environmental Sciences 22 (1), 44-49.

Zebrowitz, L.A. (1997): Reading faces: window to the soul? Boulder.

Zhong, S./Chew, S.H./Set, E./Zhang, J./Xue, H./Sham, P.C./Ebstein, R.P. and Israel, S. (2009): The heritability of attitude toward economic risk. Twin Research in Human Genetics 12 (1), 103-107.

7. Administrative Transaction Data

Julia Lane

Contact:

Julia Lane
e-mail: jlane[at]nsf.gov

The New Astronomy

"All astronomers observe the same sky, but with different techniques, from the ground and from space, each showing different facets of the Universe. The result is a plurality of disciplines (e.g., radio, optical or X-ray astronomy and computational theory), all producing large volumes of digital data. The opportunities for new discoveries are greatest in the comparison and combination of data from different parts of the spectrum, from different telescopes and archives."[1]

1. Introduction

The value of administrative transaction data, such as financial transactions, credit card purchases, telephone calls, and retail store scanning data, to study social behavior, has long been recognized (Engle and Russell 1998). Now new types of transaction data made possible by advances in cyber-technology have the potential to further expand social scientists' research frontier. For example, a person's interests and social networks can be uncovered through their online behavior documented by the major search engines, such as Yahoo! and Google, as "data collection events."[2] Geographic movements can be tracked by cell phones which include GPS location information.[3] Health, work, and learning information can be tracked by the use of administrative data from hospital records, employment records, and education records (Jones and Elias 2006). In sum, the new cyber-enabled ability to collect information from a wide variety of sources, which has transformed many disciplines ranging from astronomy to medical science, can potentially transform research on social behavior.

To be sure, the use of some transaction data for research and statistical purposes is becoming routine.[4] The *Handbook of Survey Research* will include a chapter on linking administrative records to survey data. The United Kingdom's Economic and Social Research Council (ESRC) has established

1 NVO: http://www.us-vo.org/; IVOA: http://www.ivoa.net/.

2 http://bits.blogs.nytimes.com/2008/03/09/how-do-they-track-you-let-us-count-the-ways/?scp=17&sq=privacy%20yahoo!&st=cse accessed Sept 19, 2008.

3 http://www.nytimes.com/2008/06/22/technology/22proto.html?scp=3&sq=gps%20privacy&st=cse accessed Sept 19, 2008.

4 The term "transaction data" is broadly used in this chapter to include administrative records which are "information that is routinely collected by organizations, institutions, companies and other agencies in order that the organization can carry out, monitor, archive or evaluate the function or service it provides" (Calderwood and Lessof 2006: 2). The term as used here also includes the enormous amount of transaction datasets that are becoming available from, for example, credit card records, and stock trading, as well as the location information stored from cellular telephone and the clickstreams derived from online activity.

an Administrative Data Liaison Service to link the producers of administrative data to the academic community. Furthermore, both the OECD and the Conference of European Statisticians are examining ways to use administrative data for the production of official statistics.

The opportunities are immense. The social sciences could be transformed by access to new and complex datasets on human interactions. The impact of social science on policy could also be transformed as a result of new abilities to collect and analyze real-time data. In addition, the funding exists: the United States has invested heavily in cyberinfrastructure[5] and the United Kingdom has established a National Centre for eSocial Science.[6] A good review of European Union activity is provided in a recent report by Barjak et al. (2007).[7]

A number of important issues remain.

- What is the potential for new data (e.g., citation tracking, web-scraping, biomarkers, geospatial information, through radio frequency identification devices (RFIDs) and sensors, web-based social interactions) to be included in the scientific data infrastructure? How can such data be validated, analyzed, matched and disseminated?
- How have new approaches to data dissemination (e.g., protected remote access, combined with organizational, educational, and legal protocols) advanced the potential for using transaction data in scientific research?
- What is the optimal infrastructure to promote the scientific analysis of administrative data – so that research can be generalized and replicated? What can we learn from the study of virtual organizations?

2. Background

The value of administrative data has long been recognized by the research community (Hotz et al. 2000). The study of medical outcomes, for example, has been transformed by the use of administrative records (Skinner and Wennberg 2000). Administrative data vastly expands the potential to examine the employment and earnings outcomes of low-wage workers (Autor 2009). Of course, there are a number of challenges: a detailed discussion of

5 The Office of Cyberinfrastructure was established at the National Science Foundation in 2006.
6 http://www.ncess.ac.uk.
7 http://ww3.unipark.de/uc/avross/.

the issues associated with using administrative data is provided in Lane (2009).

Increasingly, statistical agencies are also using administrative records because of the considerable pressures to keep costs down at the same time as creating new information. Indeed, the Public Policy Program of the Washington Statistical Society, in partnership with the Federal Committee on Statistical Methodology's Subcommittee on the Statistical Uses of Administrative Records, is pleased to have launched a seminar series on "Administrative Data in Support of Policy Relevant Statistics." More concrete examples are provided by the LEHD program in the United States,[8] and the LEED program in New Zealand.[9] Because an infrastructure based on administrative records created a new sample frame for economic dynamics, it has been used in its own right to create new measures of workforce dynamics at detailed geography and industry levels ranging from earnings for incumbent workers, new hires, and separated workers, to the number of quarters of non-employment of separated workers and measures of job retention and stability.

Another reason that the approach has been attractive is that administrative data have a breadth of information that is simply unattainable from other sources. For example, outside of manufacturing industries, the United States Census Bureau's measurement of inputs does not even distinguish between production and supervisory employees. After the implementation of the LEHD program, however, economic entities in all sectors (establishments or enterprises, as appropriate) were used to create detailed summaries of the distribution of observable (demographic) and unobservable characteristics of the workforce in terms of earnings, external earnings potential, and mobility.

Finally, administrative records shed new light on new economic structures. For example, using the LEHD program as an illustrative example, such data can be used to create new ways of classifying firms into particular industries based on worker activities (Benedetto et al. 2007); new ways of identifying the changing structures of firm mergers, acquisitions, and births and deaths, based on worker flows (Benedetto et al. 2007); new approaches to providing place of work and industry coding on demographic surveys such as the American Community Survey (Freedman et al. 2008), more accurate and complete coding of individual outcomes (Abowd and Vilhuber 2005) and new measures of demand side factors on household and individual surveys. Statistics on individual and household income and income mobility now include factors like whether the employer was growing or shrinking, whether the employer was profitable, and what other kinds of employees were also at the employer (Andersson et al. 2005).

8 http://lehd.did.census.gov/led/ [Last visited:10/20/2008].

9 http://www.stats.govt.nz/leed/default.htm.

3. What is the potential for transaction data to inform research?

In 2006, the amount of digital information created, captured, and replicated (worldwide) was 1,288 x 1018 bits. In computer parlance, that's 161 exabytes or 161 billion gigabytes. This is about 3 million times the information in all the books ever written.[10] The sheer magnitude of this information means that this paper can only provide an illustrative, rather than exhaustive review of the types of data that can be collected and used to describe human behavior: here we describe what can be captured using RFID's, web archiving, web-scraping and data mining of electronic communications.

The potential to describe minute-by-minute human interactions with the physical environment became reality with the development of RFID and video technologies. RFID's can be produced for pennies a unit and emit a wireless signal that enables the bearer to be tracked. Businesses now use the technology routinely to track employees (e.g., to ensure that night guards do their assorted tours at the assorted times) and to track their customer behavior (see figure 1). The potential for social science research is clear – ranging from tracking time-use information in a far more granular fashion than from survey data, to the environmental impacts on social behavior, to measuring

Figure 1:

PARIS: Thousands of garments in the sprawling men's department at the Galeria Kaufhof are equipped with tiny wireless chips that can forestall fashion disaster by relaying information from the garment to a dressing-room screen. The garments in the department store, in Essen, Germany, contain radio frequency identification chips, small circuits that communicate by radio waves through portable readers and more than 200 antennas that can not only recommend a brown belt for those tweed slacks but also track garments from the racks, shelves and dressing rooms on the store's third floor. … But the rapid development of RFID technology is also being regarded cautiously by the authorities in the European Union, who are moving quickly to establish privacy guidelines because the chips – and the information being collected – are not always visible. Their goal is to raise awareness among consumers that the data-gathering chips are becoming embedded in their lives – in items like credit cards, public transportation passes, work access badges, borrowed library books and supermarket loyalty cards.

Source: International Herald Tribune 2 March 2008.

10 The Expanding Digital Universe, March 2007, IDC White Paper sponsored by EMC Corporation.

the number and quality of human interactions. In fact, similar technologies are already being used for research purposes to great advantage. For example, Schunn uses video data collected from a recent highly successful case of science and engineering, the Mars Exploration Rover, to study the way in which human interactions contributed to the success of the project. While the project both wildly exceeded engineering requirements for the mission and produced many important scientific discoveries, not all days of the mission were equally successful. Schunn uses the video records to trace the path from the structure of different subgroups (such as having formal roles and diversity of knowledge in the subgroups) to the occurrence of different social processes (such as task conflict, breadth of participation, communication norms, and shared mental models) to the occurrence of different cognitive processes (such as analogy, information search, and evaluation) and finally to outcomes (such as new methods for rover control and new hypotheses regarding the nature of Mars) (2008).

Figure 2:

The Wayback Machine: http://www.archive.org/index.php

Of course, human behavior is increasingly captured through transactions on the internet. For example, most businesses, as well as registering with the tax authority, also create a website. It is now entirely possible to use web-scraping technologies to capture up-to-date information on what businesses are doing, rather than relying on administrative records and survey information. Historical records on businesses can also be created by delving into the repository of webpages on the Wayback Machine (see figure 2 for an example of the webpages for Citibank). This archive takes snapshots of the web every two months and stores them in the manner shown, providing a rich archive of hundreds of billions of web pages. Individual as well as business behavior can be studied using this archive. Indeed, major NSF (National Science Foundation) grants, such as the Cornell Cybertools ward,[11] have funded the study of social and information networks using these very large semi structured datasets.

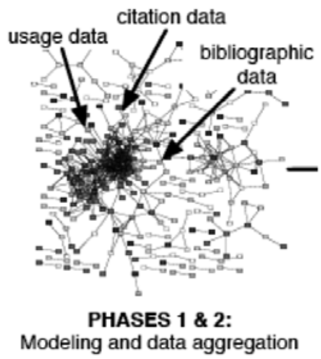
Other ways of collecting information on human behavior from the Web include capturing clickstreams from usage statistics. The MESUR project,[12]

11 Very Large Semi-Structured Datasets for Social Science Research, NSF award 0537606 http://www.infosci.cornell.edu/SIN/cybertools.

12 MESUR: Metrics from Scholarly Usage of Resources http://www.mesur.org/MESUR.html.

for example, has created a semantic model of the ways in which scholars communicate based on creating a set of relational and semantic web databases from over one billion usage events and over ten billion semantic statements. The combination of usage, citation, and bibliographic data (see figure 3) can be used to develop metrics of scholarly impact that go well beyond the standard bibliometric approaches used by academics (Bollen et al. 2007).

Figure 3:

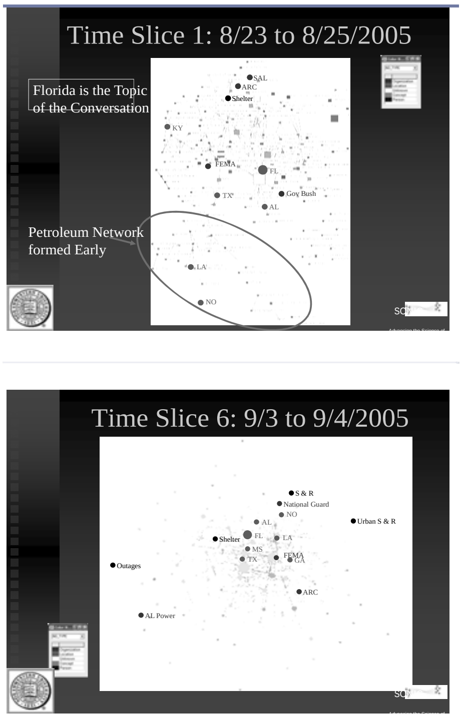


A final illustration of the value of capturing transaction data is evident from the work of Noshir Contractor. He studies a variety of ways in which humans interact with each other, including cell phone and email interactions. In a recent study he examined the emergency response of key agencies and individuals to Hurricane Katrina. The first slide in figure 4 shows the result of analytical work based on the “Data to Knowledge” application at the National Center for Supercomputing Applications (NCSA) at the University of Illinois. This is a rapid, flexible data mining and machine learning system which allows automated processing by creating itineraries that combine processing modules into a workflow. This procedure was first applied to the body of communication between 8/23/2003 and 8/25/2005 (as Katrina was approaching Florida). An examination of the top panel of figure 4 shows the American Red Cross (ARC) on the top. FEMA interactions only exist at FEMA Administration (Middle Left). Florida and Palm Beach have many mentions. At the bottom of the figure, it is clear that Oil and Power groupings are quite important, as is the pocket of National Parks in the middle. The location flags are heavily based in Florida, except for the Petroleum Network. New Orleans is very much on the fringe at the bottom.

The second slice of time that was examined was 3 September 2005 to 4 September 2005 – as the hurricane was hitting New Orleans. As is evident from the pictorial description of the analysis, Mississippi and Louisiana are the most frequently mentioned states. Urban Search and Rescue has joined the network as a key concept. The topic of power has changed to Outages, Alabama Power is still at the margin, and Shelter has moved back to the middle. FEMA and ARC have essentially swapped positions and the National Guard is moving towards the center (Contractor 2008).

This vividly illustrates how new approaches to capturing information could transform social scientists ability to provide information to policy-makers. Imagine a similar exercise being done in the study of financial markets, for example. Real-time data collected from the web analysis of online blogs and newspaper articles could have picked up clusters of concern about Lehman Brothers, Goldman Sachs, and Bear Stearns, and could potentially have described the information cascades that transformed the financial infrastructure in September and October of 2008. Or, in another example, new data could be collected on the innovation processes that generate competitive advantage within firms.[13]

Figure 4:



Of course, together with new data, new analytical techniques need to be developed. Standard regression analysis and tabular presentations are often inadequate representations of the complexity of the underlying data generation function. There are a variety of reasons for this inadequacy. First, the units of analysis are often amorphous – social networks rather than indivi-

13 http://www.conference-board.org/nsf. Carol Corrado "Workshop on developing a new national research data infrastructure for the study of organizations and innovation".

duals, firm ecosystems rather than establishments. Second, the structural relationships are typically highly nonlinear, with multiple feedback loops. Third, theory has not developed sufficiently to describe the underlying structural relationships. Therefore, making sense of the vast amounts of data is a substantive challenge. There has been considerable effort invested in developing new models and tools to address the challenge, however. For example, since a major national priority is understanding the formation and evolution of terrorist networks through the internet and other communication channels, substantial resources have been devoted to the field of visual analytics. Their research agenda aligns very closely with a potential research agenda for social scientists, focusing as it does on the science of analytical reasoning, visual representations and interaction techniques, data representations and transformations, as well as the production, presentation, and dissemination of complex relationships (Thomas and Cook 2005). It is also worth noting that new partnerships are being formed to address the nontrivial computing challenges.[14]

4. The effect of new data dissemination protocols

Both transaction and administrative data are often highly sensitive. The dissemination of such data is, however, critical for a number of reasons. The first is that data only have utility if they are used. Data utility is a function of both the data quality and the number and quality of the data analysts. The second is replicability. It is imperative that scientific analysis be able to be replicated and validated by other researchers. The third is communication. Social behavior is complex and subject to multiple interpretations: the concrete application of scientific concepts must be transparently communicated through shared code and metadata documentation. The fourth is building a collective knowledge base, particularly with new data whose statistical properties are unknown. The fifth is capacity building. Junior researchers, policy-makers, and practitioners need to have the capacity to go beyond examining tables and graphs and develop their understanding of the complex response of humans to rapidly changing social and legal environments. Access to complex microdata provides an essential platform for evidence based decision-making. Finally, access to microdata permits researchers to examine outliers in human and economic behavior – which is often the basis for the most provocative analysis.

A major barrier to the use of administrative data is the difficulty of getting permission to use administrative data for purposes other than which it

14 http://www.nsf.gov/news/news_summ.jsp?cntn_id=111470.

was collected. This is an extremely time-consuming process: since the data are collected to administer programs and not for research purposes. Legal, ethical, and financial issues similarly act to restrict access.

However, new data dissemination protocols are being developed. Remote access approaches use modern computer science technology, together with researcher certification and screening, to replace the burdensome, costly, and slow human intervention associated with buffered remote access (Lane et al. 2008). The Office for National Statistics (ONS) (Ritchie 2005) for example, instituted a full "remote laboratory" service in January 2004. Their approach is to use a thin client service, which means there is no data transfer at the user end. They have also centralized data management operations, which makes it much more efficient to work across different sites. Statistics Denmark (Borchsenius 2005) has found that remote access arrangements are now the dominant mode of access to microdata. Statistics Sweden's system for remote access to microdata (MONA; Söderberg 2005) provides users with secure access to databases at Statistics Sweden from almost any place with internet access. In this manner, Statistics Sweden has increased the accessibility of microdata for external users at the same time that it has increased security precisely because the client's computer functions like an input-output terminal. All application processing is done in the server. Statistics Netherlands (Hundepohl and de Wolf 2005) has gone even further in terms of its remote access. It has begun a pilot project, called the OnSite@Home facility,[15] which makes use of biometric identification – the researcher's fingerprint – to ensure that the researcher who is trying to connect to the facility is indeed the person he or she claims to be.

The NORC (National Opinion Research Center) data enclave has taken the remote access approach one step further. Recognizing that a remote access environment also permits the development of an environment that allows the sharing of information about data in the same fashion as that adopted by the physical and biological sciences, it has created virtual organizations (Foster et al. 2001; Pang 2001). Tools such as the Grid, MySpace, and Second Life have changed how people congregate, collaborate, and communicate: the NORC enclave offers social scientists the same opportunities. Promoting virtual collaboration not only serves the function of ensuring the generalizability and replicability of work that is fundamental to high-quality research, but also promotes a healthy interaction between data collectors, data producers, and data users. In particular, the NORC enclave allows multiple people on a team access to the data, and team members are set up with individual workspaces that are complemented by team workspaces. Each workspace allows the user to save their result sets and related

15 Hundepohl, Anco, and Paul-Peter de Wolf "OnSite@Home: Remote Access at Statistics Netherlands," paper presented at the Joint UNECE/Eurostat work session on statistical data confidentiality (Geneva, Switzerland, 9-11 November 2005).

notes. NORC supports the ongoing collaborative annotation of data analysis and results through wikis and blogs and discussion spaces. There is also a group portal environment that enables the collaborative development of research deliverables such as journal articles. Figure 5 gives a visual idea of the enclave approach.

The social science community could potentially transform its empirical foundations if it adopted such a collaborative framework. It could use remote access to a common dataset to move away from the current practice of individual, or artisan, science, towards the more generally accepted community-based approach adopted by the physical and biological sciences. Such an approach would provide the community with a chance to combine knowledge about data (through metadata documentation), augment the data infrastructure (through adding data), deepen knowledge (through wikis, blogs, and discussion groups) and build a community of practice (through information sharing). Adopting the type of organizational infrastructure made possible by remote access could potentially be as far-reaching as the changes that have taken place in the astronomical sciences, and cited in the opening section. It could lead to the "democratization of science" opening up the potential for junior and senior researchers from large and small institutions to participate in a research field.

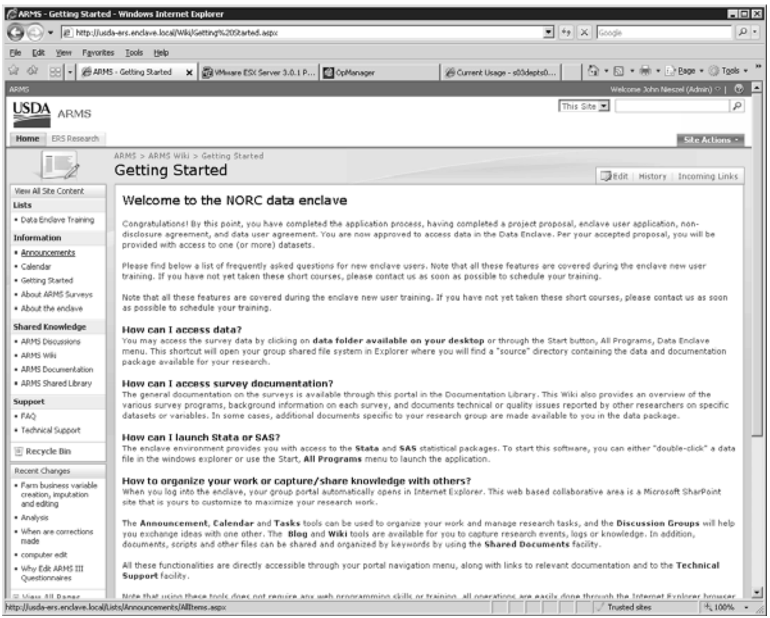
However, it is worth noting that the establishment of a virtual community to advance the development of a data infrastructure is itself a social science challenge. Indeed, the study of virtual organizations is attracting attention in its own right as a way of advancing scientific knowledge and developing scientific communities. As Cummings et al. note:

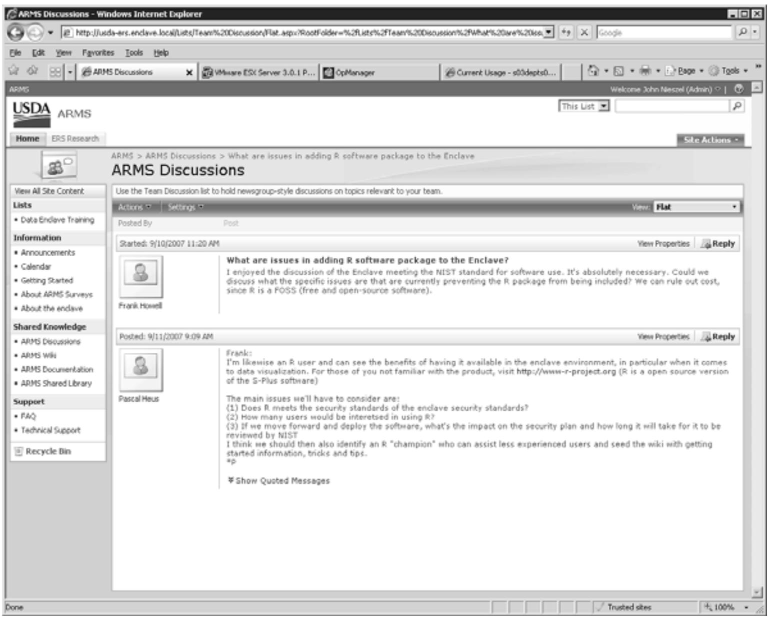
> "A virtual organization (VO) is a group of individuals whose members and resources may be dispersed geographically and institutionally, yet who function as a coherent unit through the use of cyberinfrastructure. A VO is typically enabled by, and provides shared and often real-time access to, centralized or distributed resources, such as community-specific tools, applications, data, and sensors, and experimental operations. A VO may be known as or composed of systems known as collaboratories, e-Science or e-Research, distributed work-groups or virtual teams, virtual environments, and online communities. VOs enable system-level science, facilitate access to resources, enhance problem-solving processes, and are a key to national economic and scientific competitiveness" (2008:1).

It is clearly an open research question for the social science data community to determine how such an organization should be established, how data should be accessed, how privacy should be protected, and whether the data should be shared on a central server or distributed servers. Some approaches can be centralized, like the approach taken by the ESRC in the UK in creating a specific call for a secure data archive,[16] or decentralized, such as

16 http://www.esrc.ac.uk/ESRCInfoCentre/opportunities/current_funding_opportunities/ads_sds.aspx?ComponentId=25870&SourcePageId=5964.

Figure 5:


ARMS - Getting Started - Windows Internet Explorer
Getting Started
Welcome to the NORC data enclave
How can I access data?
How can I access survey documentation?
How can I launch Stata or SAS?
How to organize your work or capture/share knowledge with others?


ARMS Discussions - Windows Internet Explorer
ARMS Discussions
What are issues in adding R software package to the Enclave?

the US National Science Foundation approach that lets the community decide.[17] Certainly both the users and the owners of the data, whether the data be survey, administrative, transaction based, qualitative or derived from the application of cybertools, would need be engaged in the process.

Similarly, it is an open research question as to the appropriate metrics of success, and the best incentives to put in place to achieve success (Cummings and Kiesler 2007). However, a recent solicitation[18], as well as the highlighting of the importance of the topic in NSF's vision statement,[19] suggests that there is substantial opportunity for social science researchers to investigate the research issues.

5. Ethics and privacy issues

A related social science research challenge that the new cyber-technologies pose, as well as potentially help to solve, is the ethical issues raised by the new capacities to collect data on human beings, particularly a focus on the privacy and confidentiality issues raised by collecting data on the interaction of human subjects.

The philosophical issues are well summarized by Madsen (2003). He identifies a "privacy paradox" in confidentiality research – occurring when data managers, in interpreting the right to privacy very narrowly, results in less social benefit, rather than in more. Two factors contribute to this paradox. One is the fear of a panopticon society, in which an all-seeing few monitor the behavior of many, which has been exacerbated since September 11, 2001. The second is a fundamental uncertainty about data ownership – whether data constitute private or public property. It is possible that the tension in the core paradox results from a framework which simply includes rights and responsibilities into the decision-making mix, rather than including social utility. But much more research must be done in this area.

The second set of issues is economic in nature (Lane 2003). Given the clear public good aspects of data collection and dissemination, how can the costs and benefits of the social investment in data be tallied to identify the optimal level of data collection? A partial list of the social benefits would include: improved decision making, avoidance of the moral hazard associated with monopoly government control of information, and improved data quality. A similar list of the social costs would include legal sanctions, the cost of breaches of confidentiality (which might substantially reduce data

17 http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141.
18 www.nsf.gov/pubs/2008/nsf08550/nsf08550.htm.
19 NSF Cyberinfrastructure Vision for 21st Century Discovery, March 2007.

quality), and support costs. Simply refusing to collect and analyze data which could inform public decision making – with tremendous public benefit – may not be a socially optimal decision.

Also of interest is how to convey the quality of such confidentiality measures to the humans who are the subject of study. Social scientists could expand their current interest in confidentiality to develop approaches that ensure the collaboration and engagement of individuals and organizations in providing data to the research community, as well as permit the data to be shared so that empirical analyses can be generalized and replicated.

It is worth noting that there is increasing interest by computer scientists in ways of protecting confidentiality so that sensitive data can be collected and analyzed without revealing individual identities – and so that researchers can generalize and replicate scientific results.[20] This interest includes policies for the anonymization and sanitization of the data, retention and storage protocols, transformation prior to dissemination, and retaining usability.

6. Recommendations

The social science community should act to address these challenges. Some work is already being done, such as the work by Peter Elias on behalf of a number of international agencies to establish the International Data Forum. However, specific, targeted activities could be undertaken to develop a new social science data infrastructure capable of answering new scientific and policy issues.

Recommendation 1: Invest in new methods of collecting transaction data

The community should take advantage of the interest that funding agencies have in funding cyberinfrastructure for the social sciences to collect new data sources. These would include clickstream information, data from web-archives, email transactions, firm administrative records, social interactions in cyberspace (such as Facebook and MySpace), and video data. The social science community should partner with data collectors, such as Google, Yahoo!, Facebook and the business community to create joint value.

Recommendation 2: Invest in new ways of analyzing transaction data

The social science community should recognize that while new units of interest to social scientists can now be studied, such as social networks, there

20 http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5033268&org=CNS.

are a number of analytical challenges. The units of analysis are amorphous and change rapidly over time. The information that is collected is no longer precisely measured: there is a high noise to signal ratio. There are large amounts of heterogeneous data. The social science community should partner with other disciplines to develop new analytical techniques. Computer and behavioral scientists have substantial expertise in creating analytical datasets in this environment; the visual analytics community has experience in making sense of such data.

Recommendation 3: Invest in new ways of disseminating transaction data

In order to develop the scientific basis for studying transaction data, the social science community needs to develop an open and transparent data infrastructure. A scientific dialogue needs to be developed about the establishment of a scientific frame, the integrity of the data, and the validation of results. In other words, social scientists must join the "hard" sciences in ensuring that their work is generalizable and replicable (i.e. scientific). A number of remote access sites are being established by leading data disseminators, such as the NORC data enclave, the UK ESDS (Economic and Social Data Service) and CESSDA (Council of European Social Science Data Archives) that promote the development of virtual organizations around data. These new access modalities offer the social sciences a way of creating virtual organizations that have new ways of collecting, accessing, and analyzing transaction microdata.

Recommendation 4: Invest in new ways of conveying complex information

The social science community should invest in new ways of conveying complex information to the broader policy making and lay communities. Tabular techniques may no longer adequately provide sufficient clarity: further investment in such visualization techniques as maps and graphs is warranted.

References:

Abowd, J. and Vilhuber, L. (2005): The Sensitivity of Economic Statistics to Coding Errors in Personal Identifiers. Journal of Business and Economic Statistics 23 (2), 133-152.

Andersson, F./Lane, J. and McEntarfer, E. (2005): Successful Transitions out of Low-Wage Work for Temporary Assistance for Needy Families (TANF) Recipients: The Role of Employers, Coworkers, and Location. http://aspe.hhs.gov/hsp/low-wage-workers-transitions04/index.htm [Vol.: http://aspe.hhs.gov/hsp/low-wage-workers-transitions04/index.htm]. Washington.

Autor, D. (2009): Labor Market Intermediaries. Chicago.

Barjak, F./Lane, J./Procter, R. and Robinson, S. (2007): Accelerating Transition to Virtual Research Organisation in Social Science (AVROSS). Brussels.

Benedetto, G./Haltiwanger, J./Lane, J. and McKinney, K. (2007): Using Worker Flows in the Analysis of the Firm. Journal of Business and Economic Statistics 25 (3), 299-313.

Bollen, J./Rodriguez, M. and Sompel, H.V.D. (2007): MESUR: usage-based metrics of scholarly impact. Paper presented at the Joint Conference on Digital LIbraries, Vancouver.

Borchsenius, L. (2005): New Developments In The Danish System For Access To Microdata, Joint UNECE/Eurostat work session on statistical data confidentiality. Geneva.

Calderwood, L. and Lessof, C. (2006): Enhancing longitudinal surveys by linking to administrative data. Centre for Longitudinal Studies Working Paper.

Contractor, N. (2008): CI-KNOW: A Tool for Understanding and Enabling the Transformative Power of Cyberinfrastructure in Virtual Communities, Presentation at the National Science Foundation, September 15, 2008.

Cummings, J./Finholt, T./Foster, I./Kesselman, C. and Lawrence, K. (2008): Beyond Being There: A Blueprint for Advancing the Design, Development and Evaluation of Virtual Organizations.

Cummings, J. and Kiesler, S. (2007): Coordination costs and project outcomes in multi-university collaborations. Research Policy 36 (10), 1620-1634.

Engle, R. and Russell, J. (1998): Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. Econometrica 66 (5), 1127-1162.

Foster, I./Kesselman, C. and Tuecke, S. (2001): The Anatomy of the Grid: Enabling Scalable Virtual Organizations. International Journal of Supercomputer Applications 15 (3), 200-222.

Freedman, M./Lane, J. and Roemer, M. (2008): New Approaches to Creating Data for Economic Geographers. Journal of Official Statistics 24 (1), 133-156.

Hotz, J./Goerge, R./Balzekas, J. and Margolin, F. (2000): Administrative Data for Policy-Relevant Research: Assessment of Current Utility and Recommendations for Development. Evanston.

Hundepohl, A. and de Wolf, P.-P. (2005): OnSite@Home: Remote Access at Statistics Netherlands, Joint UNECE/Eurostat work session on statistical data confidentiality. Geneva.

Jones, P. and Elias, P. (2006): Administrative data as research resources: a selected audit. Swindon.

Lane, J. (2003): The Uses of Microdata: Keynote Speech, Conference of European Statisticians. Geneva.

Lane, J. (2009): Administrative and Survey Data. In: Marsden, P. and Wright, J. (Eds.): Handbook of Survey Research. Oxford.

Lane, J./Heus, P. and Mulcahy, T. (2008): Data Confidentiality in a Cyber World: Making Use of Cyberinfrastructure. Transactions on Data Privacy 1 (1), 2-16.

Madsen, P. (2003): The Ethics of Confidentiality: The Tension Between Confidentiality and the Integrity of Data Analysis in Social Science Research, NSF Workshop on Confidentiality Research. Arlington.

Pang, L. (2001): Understanding Virtual Organizations. Information Systems Control Journal 6/2001.

Ritchie, F. (2005): Access to Business Microdata in the United Kingdom, Joint UNECE/Eurostat work session on statistical data confidentiality. Geneva.

Schunn, C. (2008): Integrating Social and Cognitive Elements of Discovery and Innovation. In: N.S. Foundation (Ed.) [Vol. Award number 0830210].

Skinner, J. and Wennberg, J. (2000): Regional Inequality in Medicare Spending: The Key to Medicare Reform? In: Garber, A. (Ed.): Frontiers in Health Economics. Cambridge.

Söderberg, L.-J. (2005): MONA – Microdata On-Line Access At Statistics Sweden, Joint UNECE/Eurostat work session on statistical data confidentiality Geneva.

Thomas, J.J. and Cook, K.A. (2005): Illuminating the Path: The Research and Development Agenda for Visual Analytics. http://nvac.pnl.gov/docs/RD_Agenda_VisualAnalytics.pdf. [Last visited: 03/02/2010].

8. Transaction Data: Commercial Transaction Surveys and Test Market Data

Bernhard Engel

Contact:

Bernhard Engel
German Television ZDF
Media Research Department
ZDF Strasse 1
55100 Mainz
Germany
e-mail: Engel[at]zdf.de

Abstract

Commercial transaction surveys and test market data are important sources for the analysis of consumer behavior in various markets. The advantage of these surveys is that they do not simply rely on the "weak" data of consumers but also on "measured" data (e.g., sales information, marketing information). The key questions for the analysis of commercial transaction surveys and test market data concern the prospective evaluation of market success for launched or relaunched products and services, the influence of marketing and media on product purchases under "real market conditions," and the comparison between the test market and the total market. These data are not yet used by the scientific community. There are three major challenges to getting access to the data. First, the owners of the data (market research institutes and their clients) need to allow data access. Second, the data must be anonymized in various ways (individuals and households, brands and products) without losing relevant information. Furthermore, quality guidelines for commercial transaction surveys and test market data must be developed. The German Data Forum (RatSWD) could get this process underway by initiating a project that included the participation of official statistics, the scientific community, and commercial market research.

Keywords: consumer behavior, test market

1. Introduction

Until recently, commercial transaction surveys have not been a focal point of interest for either the RatSWD or research funding agencies like the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*) or the Federal Ministry of Education and Research (BMBF, *Bundesministerium für Bildung und Forschung*). The term "commercial transaction survey" is not common even in market research. In order to define commercial transaction surveys for this advisory report, we will introduce some of the main characteristics and uses of this type of data infrastructure and also provide examples of what is *not* a commercial transaction survey.

Topics related to this report can be found in other contributions to this publication, including:

- Administrative Transaction Data (Lane)
- Interdisciplinary Longitudinal Surveys. Linking Individual Data to Organizational Data in Life-Course Analysis (Liebig)
- The Availability of Market Research Data and its Potential for Use in Empirical Social and Economic Research (Wiegand)

Also the keywords "access panels" and "(micro-)geographical data" may be useful links to this advisory report.

Commercial transaction surveys are surveys where transactions from business to consumers (B2C) are observed under controlled conditions. The extent of control, the unit of measurement, and the unit of analysis may vary. The typical commercial transaction survey is known as a "test market."[1]

In test markets, there is statistical control of variables not only on the demand side, but also on the supply side of the market. On the supply side, there is information collected about product properties, pricing, marketing activities, etc. On the demand side, consumers give information about their shopping behavior (e.g., shopping baskets, frequency of shopping, preferred package sizes), demographics, preferences, etc. Hence, it is possible to set up a test environment with single or multiple stimulus response models under controlled conditions. Test markets may be representative samples or not. It is necessary, however, for some degree of "functional representativeness" to be established. This term is used in qualitative research and means that all relevant influence factors are covered by the sample. Test markets are normally defined as surveys in a clearly defined area.

There are, however, various other survey and research methods that use the name "test market" that we will not discuss in this paper. For this report, we will *not* discuss test markets such as:

(1) *Surveys in a single store*. These surveys are often very small and are not relevant data sources for a scientific data structure.

(2) *Surveys and test markets for a single client*. The access to customer-specific surveys is difficult and the market research focus of these surveys is not always well documented.

(3) *Virtual Test Markets*. These test markets are statistical models and the database is completely derived from a calculation model, so there are no data at the respondent level.

(4) *Test markets that are fully developed markets* (e.g., Austria as a quasi test market for Germany,[2] or the use of Ireland as a quasi test market for the US through the introduction of special digital TV services).

The following sections of this expert report will concentrate on the specific characteristics of test markets as defined in this introduction. The basic research questions around such test markets concern

- the prospective evaluation of market success for launched or relaunched products and services;

1 The NHS (Nielsen Home Scan single source), a major transaction survey completed at the end of 2005, measured TV viewing behavior (electronic measurement) and consumer behavior (scanning of purchases).

2 See "T-mobile bestätigt UMTS-iPhone: Österreich wird TestMarkt." In: Der Standard. 09.06.2008.

- the influence of marketing and media on product purchases;
- the influence of these under "real market conditions"; and,
- a comparison between the test market and the total market.

Test markets are one instrument among others in the product development process. The industry will never use the results from a test market as the only criterion for decisions. This should be kept in mind when one has access to data from test markets.

There are different stages in the product development cycle where market research helps to optimize the launch to market (e.g., focus groups for concept and packaging tests, standardized procedures to evaluate an adequate market price). The last step before launch to market is very often the launch in a test market.

2. Test markets in Germany: Relevant cases

There are few sources of information about test markets in Germany that are available for general use. For the purposes of this report, there are three examples that provide *descriptions* of test markets, but not data; namely, the GfK Behavior Scan, the TNS Bonsai Deutschland, and IP Test Market Friedrichshafen. The data providers are market research institutes and the results are confidential to their clients.

GfK Behavior Scan (Hassloch)[3]

The GfK Group's test market is the largest in Germany. Focusing on the town of Hassloch (approx. 20,000 inhabitants), mainly fast moving consumer goods (hereafter FMCG) are tested. The sample size is 3,500 households. In approximately 2,500 households, it is possible to change television advertising to include targeted test spots. Between 90 percent and 95 percent of the total expenditures for FMCG is spent in stores within the Hassloch area.

The following overview illustrates the basic structure of the Hassloch test market based on the types of data collected in this project. The project produces extensive data. All purchased products are labeled with the EAN (European Article Number) Code. There are identifiers for the household, the store, the basket of all purchases, and a time stamp. The EAN Code can also be linked to additional product information. On the household level, it is

3 The GfK (*Gesellschaft für Konsumforschung*) Group is one of the largest market research companies in the world. The Group has a staff complement of 10,000+ employees working in 115 operating companies covering more than 100 countries of the world. (Högl and Hertle 2009).

possible to identify exposure to advertisements and sales promotion. Clients do not have access to raw data. The delivery of data is on an aggregated level.

The Hassloch test market is connected with other test possibilities in a larger area called *Vorderpfalz*, or the Anterior Palatinate region. In this area, the sample size is very large. There are, however, no data on a respondent level, but only on an aggregate level (testing television advertisements and store turnover) available for clients.

Normally, the structure of the Hassloch test market is compared with that of Rhineland-Palatinate. Relevant variables include age, sex, housing conditions, household type (single, family with or without children, foreigners) and spending power. According to these variables Hassloch has a similar structure to Rhineland-Palatinate, although the spending power is slightly higher (Index = 104).

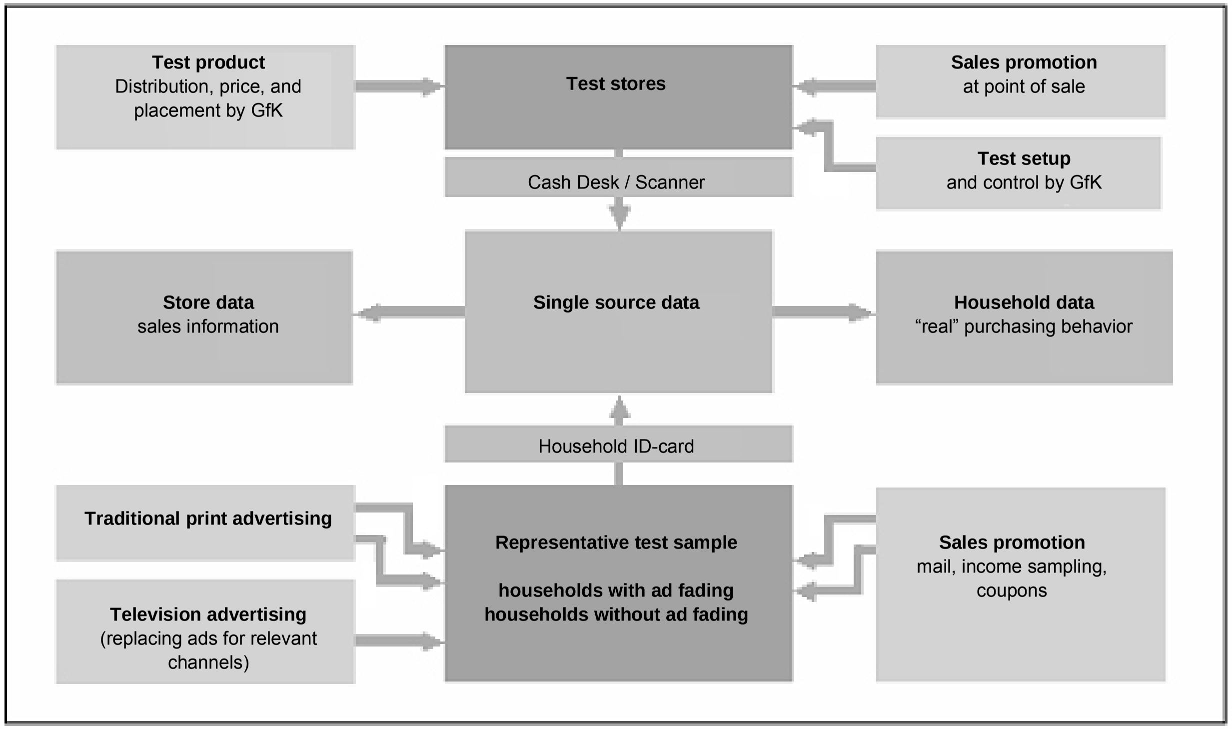
TNS Bonsai Deutschland[4]

The TNS Bonsai Deutschland is another test panel in Germany. The basic objective of this project is to optimize product lifecycles. TNS Bonsai Deutschland has no continuous consumer panel but offers client-specific surveys with data integration from other sources. The test market is located in Bremen.

The specific "unique selling proposition" (hereafter USP) or, said simply, value, of TNS Bonsai Deutschland is the OTC optimizer. TNS Bonsai Deutschland has a cooperation with about 150 pharmacies in Bremen. Bonsai Deutschland continuously generates sales data of pharmacies in the OTC and free choice area (Wawi, *Warenwirtschaft*). In combination with a nationwide pharmacy panel, there are various testing opportunities for the launch and relaunch of products but also for marketing activities at the point of sale. Although these data are very important for the success of products in the OTC sector and for category management in pharmacies, they are probably not as useful more generally as a source of data for the German data infrastructure.

4 TNS Bonsai Deutschland is part of TNS Group. TNS is one of the top five market research companies worldwide. The shareholder of TNS Bonsai Deutschland is TNS Infratest. Generally, only very little information from TNS Bonsai-Deutschland is accessible.

Figure 1: Overview of the GfK Test System in Hassloch/Pfalz



Source: Högl and Hertle 2009, English translation by B. Engel
Arrows indicate possibility of experimental influence

IP Test Market Friedrichshafen[5]

The IP test market in Friedrichshafen was established in August 2007 by Deutsche Telekom and the city of Friedrichshafen. Friedrichshafen was selected because it won the T-City contest for the best ideas on how to use modern broadband networks to improve the quality of everyday life. Consequently, Deutsche Telekom will invest €35 million into the IP infrastructure in Friedrichshafen and will spend another €80 million for the development of new products and services on the broadband network. There are various ongoing projects in this test market in areas such as education and searching, mobility and traffic control, tourism and culture, citizen and state, economy and job, and health and healthcare. The project has a cooperative agreement with Zeppelin University in Friedrichshafen (Deutsche Telekom Institute for Connected Cities – TICC)

The IP test market in Friedrichshafen can also be seen as Deutsche Telekom's contribution to the German IT Summit.[6] Activities are in place to establish continuous evaluation of projects and their acceptance in the market. Testing in the IP test market of Friedrichshafen not only evaluates consumer behavior, but also presents opportunities for testing technology.

The IP test market in Friedrichshafen is probably the most interesting test market for broader research questions. IP technology will change our everyday life in the future with its diverse array of services. In the context of this program, the IP platform allows a continuous tracking of user actions without added burden for the users. However, it should be noted that the IP test market in Friedrichshafen is not a public service. Rather, Deutsche Telekom has set up this test market to improve its competitive position in the IP market.

5 The IP Test market is a project led by Deutsche Telekom for testing and implementing new services based on internet protocol (IP based services).

6 The German IT Summit (*Nationaler IT-Gipfel*) was initiated by German Chancellor Angela Merkel in December 2006 to improve the position of Germany's IT industry. The second summit was held on December 10, 2007. There are several working groups that report to chancellor Merkel. One group, headed by René Obermann, CEO of Deutsche Telekom, works on the "Convergence of Media: The Future of Networks and Services."

3. Conclusions and recommendations

Changes in social and economic life are complex processes. In test markets, this complexity is reduced to the influence of certain measures (e.g., effects of marketing, changes in the quality or the prices of products, influence of the media on consumer behavior). The analysis of test market data can help to develop hypotheses about social and economic change.

Test markets are also used in other European countries and are extensively used in the US. The GfK Group, for example, follows test markets for FMCG that are comparable to Hassloch in Angers and le Mans (France). Insofar as sources allowing more general access to the data do exist, the US is the leader in the methodology and usage of test markets. In the system that has developed there, extensive rankings designate whether a specific MSA (Metropolitan and Micropolitan Statistical Area, according to the definition of the US Census Bureau) is a good consumer test market. There are about 150 named MSAs in the US that can be used as test markets. The determining criteria include not only demographics, but also consumer and media behavior, leisure activities, etc.[7]

3.1 Access to existing test market data

Gaining access to existing test market data in Germany presents three major obstacles. First, market research institutes and clients are the owners of the data. In many cases, more then one client is involved in the project. Data protection and anonymization are necessary not only at a respondent level, but also for other entities associated with the database, such as stores, products, producers of the products, etc. The German Data Forum (RatSWD) could provide assistance with adapting existing rules for data protection and anonymization to the specific case of test market data. In the case of the IP test market Friedrichshafen, the RatSWD could contact the TICC Institute at Zeppelin University in Friedrichshafen to promote further collaboration.

3.2 Initiatives to establish test market quality guidelines and transparency for Germany

Test markets have specific sampling requirements. It is difficult for the users of data to decide whether a test market is a "best practice" sample or not. On the one hand, it is necessary to establish a test environment under controlled

7 See Acxiom (2004). Acxiom Deutschland also offers similar data, especially to direct marketing.

conditions; on the other hand, the results from a test market should be transferable to the real world. Because a test market has multiple "entities," a representative population sample cannot meet the standards for a test market. Additional information – like infrastructure information about the town or region where the test market is located – could help to improve the value of a test market sample. Perhaps it would be useful to discuss the problem in the context of "representative sampling beyond demographics." The German Data Forum (RatSWD) could play a role by suggesting a project bringing together official statistics, the scientific community, and commercial market research with the objective to develop quality guidelines and transparency for German test markets.

References:

BMWI (2007): Konvergenz gestalten – Chancen nutzen. Broschüre zum Arbeitsauftrag der AG 2 „Konvergenz der Medien – Zukunft der Netze und Dienste“ im Rahmen des 2. IT Gipfels.

Högl, S. and Hertle, Th. (2009): MarketingLab – Evaluatives Pretesting mit der GfK Testmarktwelt. In: Bruhn, M./Esch, F.-R./Langner, T. (Eds.): Handbuch Kommunikation. Grundlagen – Innovative Ansätze – Praktische Umsetzungen. Wiesbaden.

Deutsche Telekom (2007): Testmarkt Friedrichshafen. Pressemeldung dpa vom 24.08.2007 14:33 „Friedrichshafen wird Musterstadt – Vertrag mit Telekom“. Mediaversuch am lebenden Objekt. Media & Marketing 6/2004, 44-49.

Acxiom (2004): Which American City Provides The Best Consumer Test Market? “Mirror on America” Provides Ranking of Top 150 Cities.

www.t-city.de IP-Testmarkt Friedrichshafen. [Last visited: 12/03/2010].

www.bonsai-deutschland.de TNS Bonsai-Deutschland, Bremen. [Last visited: 12/03/2010].

http://www.gfk.com/group/services/instruments_and_services/contact_dates/00141/index.en.html GfK Behavior Scan, Hassloch. [Last visited: 12/03/2010].

9. Time Use and Time Budgets

Joachim Merz[*]

* Acknowledgement: I would like to thank Dr. Markus Zwick and Heike Habla from the Research Data Center of the Federal Statistical Office as well as Erlend Holz, Prof. Dr. Manfred Ehling, and Dr. Daniel Vorgrimler (all from the Federal Statistical Office), and Dr. Kimberly Fisher (University of Oxford, UK), for their very helpful support.

Contact:

Joachim Merz
Leuphana University Lueneburg
Research Institute on Professions
(FFB, *Forschungsinstitut Freie Berufe*)
Department of Economics, Behavioural and Law Sciences
Scharnhorststr. 1
21335 Lueneburg
Germany
e-mail: merz[at]uni-lueneburg.de

Abstract

"Time use statistics offer a unique tool for exploring a wide range of policy concerns including social change; division of labor; allocation of time for household work; the estimation of the value of household production; transportation; leisure and recreation; pension plans; and health-care programmes, among others" (United Nations). This advisory report will discuss recent developments, improvements and future challenges of time use and time budgets for policy and research with a focus on international but especially national developments in Germany that have emerged in the wake of the 2001 KVI report.

The topics to be addressed are: recently established international time use institutions, data archives, and surveys; German time use databases and their accessibility, current time use research fields and studies; time use for economic and social policy; new methods in time use survey sampling; future developments; and European and international challenges. The conclusions and recommendations first urge the implementation of the new German Time Use Survey (GTUS 2011/12) and urgently call for its financing and support for its active organization. Specific GTUS improvements, SOEP time use issues, a brand new time use panel, and the permanent establishment of the German Research Data Centers are also recommended.

Keywords: time use, time budgets and time use surveys, time use data
JEL classification: C81, J2D1, I3, O15, O17

1. Time use and time budgets: General concerns

Time is the encompassing and compound dimension and resource of individual activities and living arrangements. Very generally speaking, any characteristic or information is only complete where time is a factor that is considered in addition to the factual socio-economic and geographic attributes. Quantitative-statistical based knowledge about the use of time for all conceivable activities – from the labor market to the leisure world – is thus of central importance not only for the individual but also for the economy, for governmental economic and social policy, and for society at large:

> "Time use statistics offer a unique tool for exploring a wide range of policy concerns including social change; division of labour; allocation of time for household work; the estimation of the value of household production; transportation; leisure and recreation; pension plans; and health-care programmes, among others" (United Nations Statistics Division).[1]

Time use surveys collect information about activity sequences in time spells over a period lasting from one day to a week. At the core of a time use survey is the time use diary, which registers an individual's activity sequence.

1 http://unstats.un.org/unsd/demo graphic/sconcerns/tuse/

For each main activity in such a time period additional information is entered – such as secondary activity – and information about "where" and "with whom" this activity was done. In addition to the diary information, a time use survey typically includes a questionnaire about background socio-economic individual and household variables. Sometimes specific information is included in the questionnaire about less frequent activities for a period longer than a day and/or item-specific questions like a seven-day work schedule proposed by the Harmonised European Time Use Study (HETUS, see Eurostat 2009).

Time budgets in a strong sense refer to activity specific to aggregated time used over the course of the entire day. Time budgets as a set of time taking up activities thus are comparable to income budgets spending for a set of consumption expenditures (Harms and Gershuny 2009: 1). However, the terminus time budget or time budget survey is often synonymous with the diary information itself or with the diary-based complete time use survey (diaries plus socio-economic background); this is the interpretation we will adopt here. The overall advantage of a time budget is its more accurate time use measurement than can be recorded by stylized data, and the temporal location of an activity within a given day. This offers the possibility of analyzing the timing of activities (like working hours); moreover, information about the sequence of activity patterns is an extraordinary surplus when compared to all other surveys asking for daily or weekly individual activities in the labor market or in any field of daily life.

Time use research analyses the individual's use of time. As Andrew Harvey, a longstanding mentor of time use research states,

> "Time use research is the study of how people use their time. Minimally, time use studies show what activities people do week to week or day to day. Maximally, they show what people are doing, where they are, who they are with, and how they feel from minute to minute."[2]

Time use: Background and literature

Some examples of early time use studies are the American study "How Working Men Spend Their Time" (Bevans 1913) and the British studies "Round About a Pound a Week" (Pembers-Reeves 1913). A classic German time use study is the 1933 Marienthal Study "Die Arbeitslosen von Marienthal – Ein soziographischer Versuch über die Wirkungen lang-andauernder Arbeitslosigkeit" (Jahoda, Felix and Lazarsfeld 1933).

Since the beginning of the 20th century, time use research has developed with respect to methodological as well as to substantive issues. Meanwhile

2 http://www.stmarys.ca/partners/turp/pages/whatistimeuse.htm

there are a number of comprehensive studies about the interest in and the international development of time use research. Kramer (2005) has recently provided an historic overview, Harms and Gershuny (2009) focus on time budgets and time use issues, Gershuny (2001) covers time use methods, Harvey (2004, 1999), Harvey, Merz and Mukhopadhyay (2006), Harvey, Szalai, Elliott, Stone and Clark (1984), Gershuny (1995), (Andorka 1987) or the volume on "Time Use – Research, Data and Policy" (Merz and Ehling 1999) give a general overview about the current state of the field.

Although within a time use diary the respondent is characterizing his or her activity in a time spell in his or her own words, only coded activities are available for the data user. Thus, the creation of appropriate coding for all conceivable research interests is a challenging task. However, there are international harmonizing approaches, such as the HETUS project (Eurostat 2009), the United Nations (Bediako and Vanek 1999), or alternative approaches (Hoffmann and Mata 1999). Actual scientific articles with in-depth time use analyses, books and projects can be found in particular in the new electronic *International Journal of Time Use Research*.[3] Andrew Harvey with his TURP project at St. Mary's University in Halifax, Canada, provides a substantial bibliography of time use studies. Since 2007, the Centre for Time Use Research (CTUR) has offered information about current time use publications.[4]

This advisory report will discuss improvements in and future challenges for time use and time budgets with a focus on recent international and, in particular, national developments since 2000 in the wake of the 2001 KVI report.[5] The discussion is organized as follows: section 2 sketches internationally important time use institutions, data archives, and surveys, followed by time use databases and their accessibility in Germany (section 3). Time use research fields with international and national improvements, developments, and studies are presented in section 4. Time use in and for economic and social policy is the topic in section 5. New methods in time use survey sampling are presented in section 6. Section 7 examines future developments within European and international challenges. Section 8 draws conclusions and offers some recommendations.

3 http://www.eIJTUR.org
4 http://www.timeuse.orgh c/information
5 Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik 2001, Merz 2001.

2. Time use international: Institutions, data archives and surveys

The following represent the most significant of the recently established time use institutions, data archives, and international surveys forming the improved international background in which German time use activities are embedded.

International time use institutions. Important international time use institutions are compiled in table 1.

Table 1: International Time Use Institutions

IATUR: The International Association for Time Use Research	www.iatur.org
TURP: Time Use Research Program at St. Mary's University, Halifax, Canada	www.stmarys.ca/partners/turp
UNSTATS: United Nations Statistics: Allocation of Time and Time Use	http://unstats.un.org/unsd/demographic/sconcerns/tuse/
RNTU: Research Network on Time Use at Lüneburg University, Germany	http://ffb.uni-lueneburg.de/rntu
eIJTUR: electronic International Journal of Time Use Research	www.eIJTUR.org
CTUR: Centre for Time Use Research at Oxford University, UK	www.timeuse.org

Major developments. The time use community is growing since 1970[6] and has grown increasingly within the last decade. Its annual conference in 2009 – following earlier conferences in the US and Sydney, Australia – will be at the Leuphana University of Lüneburg, Germany,[7] hosted by our Research Institute on Professions (FBB, *Forschungsinstitut Freie Berufe*) and the German Federal Statistical Office. Since 1985 TURP at St. Mary's University in Halifax (Canada) has provided a worldwide time use bibliography and is a new pioneer in spatial time use research with its 2007–2009 *Halifax Regional Space-Time Activity Research (STAR)* Project, a GPS-assisted household time use survey. Besides the recent UNSTATS activities and the time use research network RNTU activities at Lüneburg, a new peer-reviewed scientific time use journal, the electronic *International Journal of Time Use Research*[8] hosted by FFB (University of Lüneburg) was founded in 2003. Worldwide time use datasets are archived and harmonized by CTUR/MTUS at Oxford University, representing enormous progress in the ability to make international comparisons.

6 http://www.iatur.org
7 http://www.leuphana.de/ffb/iatur 2009
8 http://www.eIJTUR.org

International time use data archives. The first international time budget study was the Multinational Time Budget Study coordinated in the 1960s by Alexander Szalai (1972). This project developed standardized diaries and survey methods and was implemented by twelve countries[9] in 1965. Since then new time use and time budget surveys have increasingly been created. Recent main studies and archives since 2000 are compiled in table 2.

Table 2: International Time Use Data Archives

MTUS: Multinational Time Use Study	www.timeuse.org/mtus
MHES: Multinational Household Expenditures Study	www.economics.unimelb.edu.au/SITE/household/MTUS1.shtml
HETUS: Harmonised European Time Use Study HETUS table generating tool	http://ec.europa.eu/eurostat/ https://www.testh2.scb.se/tus
CHAD: Consolidated Human Activity Database	www.epa.gov/chadnet1/index.html

Major developments. The most comprehensive and enduring data archives of international time use studies since is the Multinational Time Use Study (MTUS) at CTUR now at Oxford University (Prof. Jonathan Gershuny, see Gershuny et al. 2000). MTUS is harmonizing time use studies based on diaries from many countries with now about 60 studies from about twenty-six countries worldwide. MHES, the Multinational Household Expenditures Study (MHES) (Prof. Duncan Ironmonger, University of Melbourne, Australia), provides individual and household information about time use *and* expenditures.

The European Union begun to support the harmonization of time use surveys and statistics in Europe in the early 1990s (HETUS, Eurostat 2009; Rydenstam 1999). Now major European time use surveys are harmonized by HETUS, an enormous advantage for the development of international comparisons. Updated HETUS guidelines are available from 2009. Based on the HETUS, Statistics Finland and Statistics Sweden have developed the *HETUS table generating tool,* an interactive, internet-based, user-friendly tool for producing user-defined statistical tables.[10] The Consolidated Human Activity Database (CHAD) will serve as an example for a specific individual time use

9 USSR, US, BRD, DDR, Bulgaria, Czechoslovakia, Hungary, Poland, Yugoslavia, Belgium, France, Peru.

10 Credentials are necessary to access the tool. Klas Rydenstam, from Statistics Sweden (https://www.testh2.scb.se/tus/tus/ and klas.rydenstam@scb.se) has to be contacted (Rydenstam 2007, 118).

database dedicated to a certain substantive aim, in this case environmental protection.[11]

International time use surveys since 2000. At the turn of the millennium around twenty European countries conducted time use surveys according to the harmonized HETUS guidelines. More than 40 international time use surveys worldwide have been conducted since 2000 (see table 3).[12]

Table 3: International Time Use Surveys since 2000

Country	Time Use Survey	Year
Argentina	Encuesta de Uso del Tiempo de Buenos Aires	2005
Australia	Time Use Survey of New Mothers	2005-2006
Austria	Austrian Time Use Survey 2008-2009	2008-2009
Belgium	Belgian Time Use Survey	2000, 2005
Brazil	Belo Horizonte Time Use Survey	2001
Bulgaria	Time Use Survey	2001-2002
Canada	General Social Survey, 19 Time Use	2000, 2005
Denmark	The Time Use of Households	2001
Estonia	Time Use Survey	1999-2000
European Union	Harmonised European Union Time Use Surveys	1999-2002
Finland	Time Use Survey: Everyday Life in Finland	2000
Germany	German Time Use Study	2001-2002
Guatemala	National Survey of Living Conditions	2002
Hungary	Time Use Survey	2000
Ireland	Adolescent Time Use and Well-Being	2007-2008
Ireland	Time Use in Ireland	2005
Italy	National Time Use Survey	2002-2003
Japan	Japanese Time Use Survey	2000, 01, 05
Mongoloia	Pilot Time Use Survey	2000
Netherlands	Time Budget Survey of the SCP Office	2000

11 CHAD is developed by the US Environmental Protection Agency (EPA) CHAD harmonizes about 10 databases with frequency and duration information of an activity (e.g., under pollution) with further daily and spatial information.

12 Detailed information about earlier harmonized international time use studies are made available by MTUS of the Centre of Time Use Research at Oxford University (http://www.timeuse.org/ information/studies/data). A list of the MTUS harmonized time use activities is available at http://www.timeuse.org/mtus/documentation/appendix.

Country	Time Use Survey	Year
New Zealand	Time Use Study	2008-2009
Norway	Tidsnyttingsundersokelsen	2000-2001
Poland	Time Use Survey	2001
Portugal	Teachers Time	2001-2003
Republic of Kiribati	Time Use Survey Gilbert Island	2001-2002
Republic of Korea	Time Use Survey	2000, 2005
Romania	National Time Use Study	2001
Slovak Republic	Time Use Survey	2006
Slovenia	Time Use Survey	2000-2001
South Africa	Time Use in South Africa	2000
Spain	Encuesta de Empleo del Tiempo	2002-2003
Sweden	Time Use Survey	2000-2001
Switzerland	Emploi du temps en Suisse	2001
Taiwan	National Time Use Survey	2004
Thailand	National Time Use Survey	2000-2001
Turkey	Time Use Survey 2006	2006
United Kingdom	Omnibus, One Day Diary Module	2001, 2005
United Kingdom	The National Survey of Time Use	2000-2001
USA	ATUS: American Time Use Survey	2003-2007

Source: CTUR/MTUS harmonized data (http://www.timeuse.org/information/studies/data) and author research.

Major developments. In addition to these recent, national cross-sectional time use surveys since 2000, other important developments can be noted. First, the Harmonised European Time Use Surveys (HETUS) were a milestone in concerted multinational sampling and activity coding of time use diary data. Second, the new US time use engagement through the annual American Time Use Study (ATUS)[13] includes work on the ATUS ancestor, The American Heritage Time Use Study (AHTUS, 1965, 1975, 1989, 1992–94 and 1998/99) which is harmonized by the Centre for Time Use Research (CTUR) at Oxford University.[14] The American Heritage Time Use Data (AHTUD) is the database for the five respective time use studies and was assessed by a

13 http://www.bls.gov/tus/
14 http://www.timeuse.org/ahtus

multinational group of experts[15] that provided calibration (Merz and Stolze 2008), evaluation, and recommendations for further time use surveys (Harvey 2006)[16]. Third, some countries are following a quinquennial period of collecting new time use surveys (Canada, Japan, Korea). Altogether, the almost exponential increase of new time use studies since 2000 worldwide emphasizes the internationally recognized importance of time use data for research and policy.

3. Time use data in Germany: Databases and data access

The most important development in providing time use diary data nationally is the official German Time Use Survey GTUS 2001/02 (predecessor GTUS 1991/92). In addition, summarized working hour information is provided by the German Microcensus. Average time use data stylized by a “normal day”[17] are part of the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*). Finally, some other topic-specific, smaller-sized surveys and firm time use data have been collected in Germany since 2000.

Time use databases in Germany

German Time Use Survey 2001/02. The 2001/02 German Time Use Survey consists of approximately 5,400 households, 37,700 diary days, and 270 activity codes classified by household work and do-it-yourself activities, paid job or job seeking, voluntary and community work, qualification and education, physiological recreation, social life and contacts, use of media and leisure time activities, child care, taking care of and attending to people, and preparation time and travel time including the means of transport. The GTUS design follows Eurostat’s Guidelines on Harmonised European Time Use Surveys (HETUS). All household members aged ten years and older were asked to fill out diaries based on 10-minute intervals on three days – two days during the week from Monday to Friday, and one day on the weekend. Data were collected on primary and secondary activities, persons involved or present, the location, and mode of transport. A wide range of household and

15 Multinational project “Assessing American Heritage Time Use Studies” by Prof. Dr. Andrew Harvey, St. Mary’s University, Halifax, NS, Canada, Prof. Dr. Dr. Ignace Glorieux, Vrije Universiteit Brussel, Brussels, Belgium, Prof. Dr. Joachim Merz, University of Lüneburg, Germany, Klas Rydenstam, Statistics Sweden.

16 http://pna.yale.edu

17 The benefits and challenges of diary vs. stylized time use information are discussed for example in Robinson 1985, Niemi 1993, and Schulz and Grunow 2007.

individual data (socio-demographic, economic, and other background variables) were collected in additional questionnaires.

The GTUS microdata themselves and information about the survey are available from the Research Data Centers of the Federal Statistical Office and the Statistical Offices of the German *Länder*.[18] In addition, the Institute for the Study of Labor (IZA, *Forschungsinstitut zur Zukunft der Arbeit*) in Bonn offers metadata about this and other surveys.[19] A comprehensive *GTUS-Compass* describing the broad range of GTUS 2001/02 information and its usage is provided by the Federal Statistical Office (2006a).[20]

German Socio-Economic Panel (1984–ongoing). Since 1984, the SOEP of living in Germany has annually collected a broad set of individual subjective and objective information from each household member sixteen years and older.[21] The SOEP, hosted by the German Institute for Economic Research, (DIW Berlin, *Deutsches Institut für Wirtschaftforschung*),[22] registers only "typical" or "normal" work and non-work daytime time use information for each of the following activities: paid work (including commuting and secondary occupational activities), housework and shopping, childcare, do-it-yourself, education/learning, watching television or videos, and hobbies and other leisure activities. In addition, the SOEP asks for information about less frequent activities and how often they were done within different longer time periods.

One advantage of the SOEP (among others) is its truly longitudinal character and its broad range of socio-economic variables for testing behavioral hypotheses. The disadvantage (besides having exclusively stylized information) is that it only permits information on full hours of activity (no minutes or smaller units of time) when collecting data. A simple extension by minutes is strongly recommended for further SOEP waves and for international comparisons.

German Microcensus: The large-scale German Microcensus[23] (1 percent sample of the population) is focused around the labor market and has asked for in-depth information about a variety of "typical" or "normal" working hours since 2005, as well as for current as well as desired working hour arrangement.

18 http://www.forschungsdatenzentrum.de

19 http://idsc.iza.org/metadata/

20 Compass topics: Publications of government, ministries, and research facilities, conferences, journals and other media; Master thesis, final diploma, doctoral dissertations; Eurostat: Harmonised European Time Use Study (HETUS); Federal Statistical Office publications; United Nations (UN); Journals about time use and related topics; Associations, conferences, data archives and research facilities about time use and related topics; General research facilities and data archives; Contact about the Time Use Surveys at the Federal Statistical Office of Germany.

21 http://www.diw.de/soep

22 http://www.diw.de/english

23 http://www.destatis.de

Further studies with time use information. Time use information gathered by private firms, such as Nielsen Marketing or the Society for Consumer Research (GfK, *Gesellschaft für Konsumforschung*) will be discussed in the "Time Use Research Fields" section below. Television and broadcast services (like ARD or ZDF) and other media firms have developed their own large-scale survey system about media use with a significant amount of process-based time use information. The situation and the suitability of diary-based time use research for media use have recently been analyzed by Merz (2009). Smaller-sized or topic-specific studies include the "Berliner Längsschnitt Medien," a project to analyze media use and school performance by the Criminological Research Institute of Lower Saxony (KFN, *Kriminologisches Forschungsinstitut Niedersachsen*)[24] or, in another example, the time use study focused on intra-family relations conducted by the State Institute for Family Research at the University of Bamberg.[25] Although there are important private firms and other institutions that collect time use data in Germany, in general, the data are not available to other institutions or researchers, in general.

Time use microdata access in Germany

While the SOEP and its time use data have been made available for scientists since its inception in 1984 via the DIW Berlin, official microdata have also been provided for some years by new Research Data Centers for the public and the scientific community.[26] The official German Time Use Surveys GTUS 2001/02 and GTUS 1991/92 are provided and serviced by the Research Data Centers of the Federal Statistical Office and the Statistical Offices of the German *Länder*. These Research Data Centers provide four different forms of access to selected microdata of official statistics: Public Use Files (PUFs), Scientific Use Files (SUFs), safe scientific workstations and data laboratories, and controlled remote data processing. These four options differ with regard to both the anonymity of the microdata that can be used and the form of data provision.[27] Access to German official microdata is possible for foreign institutions and scientists not subject to German law.

New microdata access developments after 2000 and in the future. The entire system of microdata access via the Research Data Centers is a new one and has created very successful options for working with official microdata, such as the creation of SUFs. However, SUFs are still anonymized; a "final run" with the original data held within the Federal Statistical Office is

24 http://www.kfn.de
25 http://www.ifb.bayern.de/ forschung /inapf-deu.html
26 http://www.ratswd.de/engl/dat/RDC.html
27 http://www.forschungsdatenzentrum de/en/anonymisierung.asp

necessary for many final results and publications. The new *onsite secured possibilities* (similar to those at the DIW Berlin for geo-coded SOEP data) is a promising avenue for providing advanced access. The possibility for *remote access to micro- and metadata*, which, for instance, is provided by the Luxembourg Income Study (LIS), will also be important in the future. The most important future issue, however, is that the Research Data Center of the Federal Statistical Office must be permanently established in order to continue to provide this necessary, well-accredited service!

4. Time use research fields: International and national improvements, developments, and studies since 2000

In principle, time use research fields encompass the whole range of human activity. However, particularly in the specific time use diary type of data, they focus on and allow for activity analyses incorporating attributes of the timing, duration, and sequence of activities with all its effects and causalities of daily life activities. Stylized time use data also give insight into a normal or average day and/or less frequent activities within a desired period of time.

From this perspective, the international and national time use research fields that have emerged since 2000 can be said to include substantive contributions from economic, sociology, and other sciences and also to have addressed methodological issues on a national and multinational level. Though there are a multitude of studies behind each time use research field over the past decade, and certainly behind those dating before this,[28] in assembling table 4 only one international and one national reference will characterize each issue. My taxonomy of time use research fields tries to capture recent international and national research activities and a variety of sources could be cited.[29]

---

28 See for example Merz and Ehling 1999.

29 National: For GTUS 2001/02 based studies the excellent GTUS-Compass by the German Federal Statistical Office (Statistisches Bundesamt 2006) and further actual information provided by its author Erlend Holz; Research Project Summary and literature from the Reserach Data Center of the Federal Statistical Office (http://www.Forschungsdaten zentrum.de). International: CTUR publication list (http://www.timeuse.org/informa tion/publications/; Information by the Research Network on Time Use Research (RNTU: http://www.rntu.org; electronic International Journal of Time Use Research (http://www. eIJTUR.org) and other Journals.

Table 4: Time Use Research Fields since 2000 – International and National

	International		National	
Time Use Research Field	**Reference**	**Time Use Data**	**Reference**	**Time Use Data**
Labor Market/ Paid Work **Timing, Fragmentation of Work/ Work-life balance/ Sequencing**	Hamermesh 2002/Fisher and Layte 2004/ Lesnard 2004	MTUS Version 5.0.1 (D), British TUS 2000-01 (D), HETUS 2003 (D)/ French TUS 1985-86, 1998-99 (D)/.	Merz and Böhm 2005; Merz and Burgert 2004; Merz, Böhm, Burgert 2004	GTUS 2001/02 (D)
Unpaid work/ Nonmarket Activities/ Household Production	Deding and Lausten 2006, Harvey 2006, Ironmonger 2001	Danish TUS 2001 (D), American (Heritage) TUS (D), Australian TUS (D)	Schäfer 2004	GTUS 2001/02 (D)
Gender Perspectives	World´s Women Report UNIFEM 2009	Multiple time use data worldwide (D/Q)	Cornelißen 2005, Sellach et al. 2004	GTUS 2001/02 (D)
Division of Housework	Anxo and Carlin 2004, Bonke and McIntosh 2005	French TUD 1999 (D)	Gille and Marbach 2004	GTUS 2001/02 (D)
Child Care/ Day Care/ Care giving	Joesch and Spiess 2006, Chalasani 2006	ECHP 1996 (Q)	Kahle 2004, Fendrich and Schillig 2005	GTUS 2001/02 (D)
Family Interactions/ Parental Time and Leisure	Anxo and Carlin 2004/ Guryan, J., Hurst, E. and M.S. Kearney 2008	French TUS 1999/ American TUS 2006	Bundesministe-rium für Familie, Senioren, Frauen und Jugend 2006	GTUS 2001/02 (D)
Nutrition/ Household Economics	US Department of Agriculture	American TUS 2005, 2006	Gwodz et al. 2006	GTUS 2001/02 (D)
Consumption/ Shopping	Jacobson and Kooreman 2004	Netherland SCP Survey 2000 (D/Q)	Merz, Hanglberger and Rucha 2009	GTUS 2001/02 (D)
Education	Guryan et al. 2007	ATUS 2005	Wilhelm and Wingerter 2004	GTUS 2001/02 (D)
Leisure/ Culture/ Quality of Life	Torres et al. 2007	European Quality of Life Survey (25 countries)	Statistisches Bundesamt 2008, Weick 2004	GTUS 2001/02 (D)

	International		National	
Time Use Research Field	**Reference**	**Time Use Data**	**Reference**	**Time Use Data**
Media use/ Play/ IT	Deal 2008	Digital Games Survey 2006 (D/Q)	Merz 2009, Jäckel and Wollscheid 2004, 2007, Fritz and Klingler 2006, Kleinmann and Mößle 2008	GTUS 2000/01 (D), ARD/ZDF-Studie 2005 (Q), BL 2005-2010 (D/Q)
Space/ Geography/ Environment	Harvey 2009	STAR: GPS Time Use Survey 2008	Kramer 2005	GTUS 2001/02 (D)
Mobility/ Transport/ Travel	Keall and Baker 2008	Travel Survey New Zealand 2001 (D)	Kramer 2004	GTUS 2001/02 (D)
Social Contacts/ Networks / Volunteering	Bittman et al. 2005	Australian TUS 1997 (D), Australian SDAC 1998 (Q)	Merz and Osberg 2009, Gabriel et al. 2004	GTUS 2001/02 (D)
Time Crunch/ Time Stress/ Harriedness	Sullivan 2007, Bonke and Gerstoft 2007	Danish TUS 2001 (Q), Home OnLine 1998 (Q/D)	Gille and Marbach 2004	GTUS 2001/02 (D)
Poverty/ Extended well-being/ Inequality	Akarro 2008, Folbre 2009	Time Use Study and Advanced Census Analysis in Tanzania 2002 (D)	Holz 2004, Kettschau et al. 2004, Merz and Rathjen 2009 ISG 2004	GTUS 2001/02 (D)
Special Populations **Children/ Adolescent, Youth/ Elderly, Retirement**	Pääkkönen 2008/ Mulligan, Schneider and Wolfe 2005/ Piekkola and Leijola 2006	Finnish TUS 1999/00 (Q)/ CPS 1992 (Q), NELS 1992, SDAC 1992/93 (ESM)/ MTUS: 1987, 1991, 1995, 1999 2000 (D)	Cornelißen and Blanke 2004, Engstler et al. 2004	GTUS 2001/02 (D)
Economic Accounting/ Valuing/ Sustainable Society	Landefeld and Culla 2000, Eurostat 2003	Country time use study aggregated to Natinal Accounts	Schäfer 2004, Stahmer 2003, Stäglin 2003	GTUS 2001/02 (D). Time-Input-Output Tables

	International		National	
Time Use Research Field	**Reference**	**Time Use Data**	**Reference**	**Time Use Data**
New Methods Visualization/ Sequence Analysis Timing/ Profiling/ Heterogeneity/ Entropy	Michelson and Crouse 2004, Ellegard and Cooper 2004/ Wilson 2001/ Stewart 2006/ Gonzales-Chapela 2006/	ALLBUS 12 1998 (D), FAMITEL 2001 (D/Q), Swedish TUDPS 1996 (D)/ ALLBUS 1998 (D)/ EPA TDS 1992-1994 (D), ATUS 2006 (Q)	Hufnagel 2008	GTUS 2001/02 (D)
Methodology **Diary versus Questionnaire/ Representa-tivity**	Kitterod and Lyngstadt 2005, Niemi 1993/	Norwegian TUS 2000/01 (D/Q)/	Schulz and Grunow 2007/ Merz and Stolze 2008	ifb TUS 2006/ AHTUD 1965-99

AHTUD: American Heritage Time Use Data, ALLBUS: German General Social Survey, ATUS: American Time Use Survey, ARD/ZDF 2005: ARD/ZDF-Studie Massenkommunikation 2005 (Q), BL: Berliner Längsschnitt Medien, CPS: Current population survey, ECHP: European Community Household Panel, HETUS: Harmonised European Time Use Studies, MTUS: Multinational Time Use Study, NELS: National Education Longitudinal Study, TUS: Time Use Survey, SDAC: Survey of Disability, Ageing and Carers, Sloan Study: Study of Youth and Social Development Wave 1, GTUS: German Time Use Survey
Source: Author taxonomy based on various national and international data (see Footnote 10).

Major improvements and developments. Altogether, the table 4 overview shows a wide range of research fields relating to important economic and social issues. For instance, specific time use information provided by diaries allows particular *labor market analyses* that are not available in other labor market surveys: the sequencing, timing, and fragmentation of daily working hour arrangements, multiple jobs per day. These are important for new forms of labor contracts in the development of labor market flexibility. *Unpaid work and nonmarket activities* are significant for understanding the importance of the informal economy and underscore women's economic importance and gender approaches to labor in particular. The *total leisure activities*, including social networking and volunteer work, family interaction, media use, culture, sports, and genuine leisure (to mention only a few) are important in many respects for understanding economic, social, individual, and societal living conditions. For example, recent psychology time use studies (via experience sampling) have been used to study affect regulation (Riediger et al. 2009).

For the German context this overview also demonstrates that the recent German Time Use Study GTUS 2001/02 enabled a broad spectrum of in-depth activity research in a wide range of research fields. However, the primary German database GTUS 2001/02 is no longer up to date; there is an urgent need for a new German time use diary survey. Further information about the over fifty substantive research projects that have been reported to date that are served by the Research Data Centers and based on GTUS 2001/02 (with a great number more using data from GTUS 1991/92) emphasizes the critical importance of the German Time Use Survey for scientific as well as for administrative purposes (see the list of the Research Data Center research projects in the Appendix table A1).

In addition to the spectrum of time-use based scientific research activities that have been discussed, there are many other fields in private enterprises and administrative or governmental activities that ask for or would gain from time use information.

Private firms and time use information. Besides all the working hour time use data within any given private firm, private organizations in the field of consumer surveying also collect item- and time-specific information. To mention only the two important of these private firms: The Nielsen Consumer Panel survey, for example, which now includes 300,000 households in twenty-eight countries, collects information on consumption activities[30] scanned by the respondents via bar-codes. The GfK runs its ConsumerScope with even more explicit time use information, including specific studies on gardening, media use, etc., thus deepening the activity-specific time use information.[31]

Time use and downsizing bureaucracy by reducing administrative burdens: The Standard Cost Model (SCM) of the Federal Statistical Office, a tool for downsizing bureaucracy, measures the administrative costs imposed on businesses and individuals by central government regulation. Specific SCM time use surveys and interviews provide the data to this end and data from GTUS 2001/02 is used for further investigation. The German efforts are integrated in an international SCM network.[32]

Time Use, National Accounts, and Nonmarket Production: Though the main focus of time use research is on individual behavior, there are substantial longstanding international and national efforts to record the contribution of nonmarket production to the national product and national accounts. Emphasis in this area is placed on valuing individual time use using various methods, such as market replacement costs with global or specialized substitutes, opportunity costs, and self-evaluation (Chadeau 1985; Goldschmidt-Clermont 1993). Recent international nonmarket national accounts efforts are described by Landefeld and Culla (2000) and Eurostat (2003). An interesting new way to describe the macro situation of a society is the “Great Day,” an aggregated time use picture proposed by Gershuny (1999).

30 http://www.acnielsen._de/products/cps_homescan.shtml
31 http://www. gfkps.com/scope/infopool/chartofthe week/index.de.html
32 http://www.administrative-burdens.com/

Recent German national accounts by nonmarket satellite systems focus on time pattern in a Social Accounting Framework (see Stahmer 2003, Stahmer and Schaffer 2004, Stäglin and Schindtke 2003 for time input-output tables). Schäfer 2004 provides an estimate of a nonmarket production contribution for the German national accounts based on the GTUS 2001/02.

5. Economic and social policy and time use

Targeted economic and social policy needs accurate individual information about the population. The comprehensive range of time use data on individual activities can provide genuine information to support almost any sound economic and social policy and to accompany the daily temporal coordination of life. Against the substantive background of our time use research field overview (Table 4), one can identify a few main policy areas and new activities – of international importance but cited here with German references – that gain in particular from individual time use information:

- *Family and time use policy.* For almost all activities considered in the recent Seventh Family Report of the Federal Ministry of Family Affairs, Senior Citizens, Women and Youth (BMFSFJ, *Bundesministerium für Familie, Senioren, Frauen und Jugend*) (2006) with time policy for child care, child-rearing allowances, balancing family and work, education and other aspects of individual living conditions.[33]
- *Public transport, traffic, mobility, and time use policy.* There is increasing interest in individual transport and traffic time aspects of working and leisure activities (see Kramer 2005).
- *Bureaucracy downsizing and time use.* Reducing administrative costs and time burden imposed on businesses and individuals (see the discussed SCM project).[34]
- *Poverty and time use policy.* See the reports in this publication for a discussion of the three German Federal Richness and Poverty Reports (Armuts- und Reichtumsberichte der Bundesregierung, Bundesministerium für Arbeit und Soziales 2008; in particular: ISG 2004; Kettschau et al. 2004).
- *Working hours, labor market flexibility and time use policy.* Setting administrative general regulations on working hours and working conditions with particular daily working time regulations.

---

33 http://www.bmfsfj.de
34 http://www.administrative-burdens.com

- *Time policy in urban and rural areas.* To support the temporal coordination of public services and the private and firm sector.

The Time Use Compass by the Federal Statistical Office (2006a) mentioned above provides an additional range of time use information used by the German administration for economic and social policy.

New and future time use policy developments. The temporal aspect regarding family affairs and working-hour arrangements is a longstanding policy focus. Time use policy interests are new with regard to urban and rural temporal coordination of daily life, such as the time policy project for the metropolitan area Hamburg (Mückenberger 2008) and the new time policy of Europe-wide activities (Garhammer 2008). For further examples, see the activities of the German Society for Temporal Governance (DGfZP, *Deutsche Gesellschaft für Zeitpolitik e.V.*).[35]

6. New methods in time use survey sampling

All the substantive time use approaches and research fields are based on the following instruments and methods:

- *Direct time use questions (stylized approach)* record the number of times that an individual participated in a given activity or the amount of time denoted for that activity in a typical day – either time constrained (must cover a defined time period) or time unconstrained.
- *Activity lists* are typically selective rather than exhaustive; mostly time-unconstrained.
- *Beepers* (experienced sampling) collect information via signaling devices that call for immediate information randomly over a given period (day) to register immediate subjective and context-sensitive information.
- *Time use diary* is an exhaustive record of all activities and patterns of associations between people and locations; this allows for sequence analyses; a highly recommended approach.

Time use research uses all kinds of time use data, but the diary is the preferred method of sampling, followed by stylized data. Both have benefits and challenges: diaries allow the investigation of activity timing during a day, stylized data capture less frequent information and disregard the randomness of situations occurring on a single day, to mention only the main issues (see Harvey

35 http://www.zeitpolitik.de

1999 for more). There are some beeper data-based results, like those of the ISR Michigan group, but beeper data is not the dominant sampling instrument that is used. However, experience sampling, by a beeper or another instrument, which collects context-sensitive data by a self-reported momentary experience, by a random or other scheme over a day, for instance, is increasingly used at least in psychological investigations (Riediger 2009).

Within this methodological framework, many new sampling tools connected with the growth of handheld devices and mobile phones have been developed (see the conference volume about new sampling technologies with focus on time use surveying by Ehling and Merz 2002).

Table 5: New time use sampling technologies by surveying principles

Come and Go
PZE-Master [Working hour per terminal] www.zeit-reporter.de/article_info.php?articles_id=154
NovaCHRON [Workers time per web client] http://www.novachron-zeiterfassung.de/personalzeiterfassung.php
diTime [Working hours per web-client] http://www.disoft-solutions.de/
timeCard [Working hours per chipcard/token] www.easy-technology.de/software/timecard/
Micades [Mobile per barcodescanner and GPRS/GSM] www.mobile-zeiterfassung.info/Fahrzeug.html
MOBILDAT [Mobile per software] www.mobile-zeiterfassung.info/Fahrzeug.html
Webalizer [Media/IT use per software] www.tobias-schwarz.net/webalizer_gui.html
Web-Zählpixel [Internet use per plugin/software] www.ivw.de
User tracking [Internet use per cookies/software] www.agof.de/
Project Precise
MobilZeit SERVICE [Working hours per terminal] http://www.mobile-zeiterfassung.info/Fahrzeug.html
TimeLog Project [Working hours per software] http://www.timelog.de/produkte/zeiterfassung.html
TIM / TIM Mobile [Mobile per cell phone (GPRS/GSM) and software] www.pressebox.de/pressemeldungen/echtzeit-zeitmanagement/boxid-108393.html
Task Precise
Zeittagebücher [per diary]
Time-Soft [Working Hours per web-client] www.lewald.com
Micro-Kiosk-System [Working Hours per terminal / PDA] www.softguide.de/prog_g/pg_2252.htm
diTime [per Barcodescanner] http://www.disoft.de/index.htm
SMS-Methode [per cell phone and software]
Mobile Zeiterfassung [per cell phone and software]http://www.virtic.com/?u=mobile_zeiterfassung
Halifax Regional Space-Time Activity Research (Star) Project [activity per cell phone (GPS) and diary] http://www.stmarys.ca/partners/turp/pages/projects/STAR/STAR_Main.htm
TimeCorder [activity per hardware] http://www.paceproductivity.com/timecorder.html
Timeboy [per Hardware] www.datafox.de
mQuest [per PC, PDA or smartphone] www.mquest.info

Source: Merz 2009.

Many new sampling instruments, mainly developed to collect individual working-hour information, can be classified – according to the taxonomy of Merz 2009 – by three principles: Come and Go, Project Precise, and Task

Precise. Come and Go measures the total daily working time (when and how long). Project Precise measures the time information for a certain project (when and how long). Task Precise might measure a certain (sub-)task of a project. Table 5 provides examples of new time use sampling devices for each of these principles.

For a discussion of the benefits and challenges of these new time use data sampling instruments see Merz (2009). They certainly have to be considered and tested before they might be used for a future German Time Use Survey.

7. Future developments: European and international challenges

The worldwide financial and economic crisis accentuates the importance of the effective use of scarce resources. Since time use surveys encompass many (or all) individual activities incorporating temporal information, they are a very efficient "all-in-one" tool that provides a broad scope of detailed individual data in a household context for a multitude of substantive interests with minimal investment. Therefore, one could expect that the current crisis favors the implementation of new time use surveys; however, policy-makers still need to be convinced of its enormous practical value.

In Europe, great efforts have been invested and still have to be invested in order to get a full HETUS every ten years (as Norway, among other countries, has been doing for decades). Following approximately twenty new time use surveys from the beginning of the millennium (2000–2002), the next European Harmonised Time Use Study (HETUS) in 2010–12 will be a cornerstone not only in national surveying and research but also for the development of the European community as a whole.

In the UK, "light" diaries have been discussed for the multiple intervening years between the full-scale surveys every ten years (in Japan and Korea there are only five intervening years). According to the IATUR secretary Dr. Kimberly Fisher, there are a growing number of diary surveys on specific topics linked to longitudinal data – several studies focus on children, for example, notably the Child Development Supplement of the PSID (*US Panel Study of Income Dynamics*)[36] and the "Growing Up Longitudinal Study of Australian Children."[37] These narrowly focused studies represent another way to collect individual time use data.

36 http://psidonline.isr.umich.edu/CDS/
37 http://www.aifs.gov.au/growingup/

Internationally, new countries and new time use surveys are on the agenda worldwide. A new UN-sponsored series of studies in developing countries is being discussed. Based on the experiences and the efforts of the annual American Time Use Studies, the international time use community will succeed in more frequent cross-sectional time use surveys. A comprehensive survey by the German Federal Statistical Office about *Time Use Survey – National Plans for the next wave of surveys 2008-2010 for 32 countries* is included in Appendix 2.

Periodic cross-sectional time use surveys with intervals of five or ten years will be very important in the upcoming years. The invention of an annual time use panel of regularly surveyed individuals and/or households with all its longitudinal information is on the international agenda. The panel option will be an enormous step forward in time use research that will provide – among others things – specific event-driven micro information for up-to-date and targeted policy and research. New electronic devices allow more precise and at the same time less expensive time use data sampling. Future developments and challenges for the time use survey situation in Germany will be outlined in our conclusions and recommendations.

8. Conclusions and recommendations

This advisory report on the current situation in international and national recent time use, recent improvements and future developments has underscored the following: time use surveys – with time as the comprehensive dimension of any individual activity – allow new insight into daily living activities, incorporating the timing and sequence of lived events. The central time diary methodology cues respondents to walk through the sequence of events in a given day, which has significant advantages in ensuring the completeness and consistency of responses. Time use diaries thus support an understanding of causality, and the interdependence that exists between all market and nonmarket activities and their individual synchronization. The disadvantage, however, is the high cost of administration, which mandates relatively few days observed per respondent with the resulting possibility that a survey will miss low frequency events. Therefore, additional summary questions about the “work week” (HETUS) have already been added to the GTUS 2001/02 as well as in some other time use surveys.

Against the background of growing international experience in the field of successful time use survey methodology, the following recommendations are indicated, with a particular focus on Germany. They will support research and targeted policy with more advanced, substantive as well as methodological investigations on modeling individual and household behavior at the micro-level and on developing new and sound national accounts data at the macro-level:

- *Recommendation 1a (GTUS 2011/12):* it is essential that the next official German Time Use Survey (GTUS) is conducted in the years 2011–2012 nationwide by the Federal Statistical Office. The financing for GTUS 2011/12 is not yet assured and it must be organized as soon as possible. The next GTUS has again to be embedded in the European Harmonised Time Use Surveys (HETUS, Eurostat 2009). The next GTUS 2011/12 would assure information in a ten years interval context together with GTUS 1991/92, GTUS 2001/02 with precious time use information including socio-economic background available for targeted policy and research.

- *Recommendation 1b:* new methods in sampling time diaries based on mobile devices – including beeper and/or experience sampling methods for even more context-sensitive questions – should be incorporated in the next GTUS after a proper pilot study. This will fulfill three objectives: first, to gain more context-sensitive data; second, to reduce the burden of filling out a traditional diary booklet; and third, it will reduce the overall expense. The sampling procedure should use mixed-mode data collection (internet, cell- or telephone, mail, pre-coded diaries, etc.) and the advantages of the Access Panel (Körner et al. 2008) with voluntary information from the German Federal Statistical Office.

- *Recommendation 1c:* the single activity spell with its "where" and "with whom" attributes should be extended by expenditure information. This would provide new data about expenditures associated with each activity and the intensity for all related activity fields (transport, shopping, etc.).[38] A suitable way must be found to characterize a second or third job within a daily activity spell.

- *Recommendation 1d:* the time use diary information should be extended by questions concerning less frequent activities. First, with information about the work week consistent with the recent HETUS recommendation (Eurostat 2009, Guidelines Annex VI). Second, by information about a longer period than a day (different week diaries, frequencies, etc.). Third, by information about a "typical" or "normal" period (day, week, month).

- *Recommendation 1e:* the time use diary supplementary information should be extended by more objective background individual and household questions and questions about the living environment. The supplementary data should contain information about the income situation from

38 For example, with brackets for a sequence of equal activity spells.

labor market activities (occupational status, wages, and detailed income including unemployment benefits, etc.) or from other income sources (capital income, further third-party payments, etc.). The environmental information should encompass external child care possibilities and external living conditions (residence environment, exposure to environmental risks, and social life participation including social networks, social “inclusion” etc.).

- *Recommendation 1f:* the time use diary supplementary information should also be extended by subjective information about satisfaction (of life in general and other items beyond time type and stress information) and health (subjective and objective). In addition, the “Big Five” personal characteristics items[39] should be added to create an approximate measurement of unobserved heterogeneity, for instance. All this subjective data will allow researchers to value and qualify the time use information.
- *Recommendation 1g:* the time use diary supplementary information should be closely adjusted and harmonized with the respective socio-economic questions of the SOEP to allow for high quality merged new datasets.
- *Recommendation 2:* a brand new annual Time Use Survey Panel should be started to answer important longitudinal questions. A TUS Panel – for example in the wake of GTUS 2011/12 – will allow the investigation of changing individual time uses and time use profiles in changing environments with extended causality and sequential event analyses. The TUS Panel thus has a different focus than the SOEP.
- *Recommendation 3:* the SOEP should continue to ask for both “typical day” as well as less frequent time use information. First, this will allow continuing longitudinal analyses. Second, it will enable the use of the enormous socio-economic background information on the labor market and additional information present in the SOEP to explain time use behaviour. The SOEP should not only ask for full hours but should allow minutes’ information as well.
- *Recommendation 4:* the Research Data Center of the Federal Statistical Office should in any case be advanced to a permanent standing. However, particularly for its time use data service and its role developing new time use data it should be established permanently. The new onsite secure data access possibilities should be further developed. Particularly,

39 See for a short Big Five Inventory, the SOEP version of the Big Five (Schupp and Gerlitz 2008).

remote access to micro- and metadata should be expanded for fast and secure access.

- *Recommendation 5:* in general, the German Data Forum (RatSWD) should actively support and strengthen all activities related to ensuring that the GTUS 2011/12 will be financed and organized. Because a time use survey provides such a multitude of substantive answers for policy and research in a single, "all-in-one" tool, because it is harmonized now within Europe and offers an efficient use of scarce resources, the next GTUS 2011/12 should be rigorously and tenaciously promoted.

Appendix A1: Current registered research projects registered with the Research Data Center of the Federal Statistical Office and based on GTUS 2001/02

No.	Research Projects: Registered with the Research Data Center of the Federal Statistical Office and based on GTUS 2001/02 (March 2009)
1	Arbeitstitel: Haushalt: Kleine Fabrik oder gender factory
2	Zeitverwendung von Arbeitslosen und Vollzeiterwerbstätigen. Eine vergleichende Analyse mit den Zeitbudgetdaten des Statistischen Bundesamtes von 2002.
3	Inklusionsprofile
4	Zeitverwendung in Haushalten
5	FrauenDatenReport 2005
6	Feiertage, Freizeit und Soziales Kapital
7	Soziale Netzwerke und Hilfebeziehungen im unteren Einkommensbereich
8	Consumption and Time Allocation
9	Female labor market supply and home work in Germany
10	Bayerischer Familienreport 2006 – Schwerpunkt "Väter in Deutschland"
11	Kooperative Demokratie – Kritik der Arbeit und der Arbeitslosigkeit
12	1. Erwerbsverhalten und Home Production / 2. Zeitverwendung im Alter
13	Der soziale Dienstleistungsbereich als Chance für eine höhere Arbeitsmarktintegration und Professionalisierung weiblicher Erwerbskarrieren
14	Zeitverwendung und Work-Life-Balance in Großbritannien und Deutschland
15	Das Arbeitsangebotsverhalten von Frauen in Deutschland
16	A. Mobilitäts- und Freizeitverhalten von Kindern und Jugendlichen B. Verbesserung der Methoden zur Prognose der KFZ-Bemessungsverkehrsstärken
17	Zeitverwendung und soziale Schichten
18	Klartext reden oder Farbe bekennen: Der Einfluss von Sprachkenntnissen und Aussehen auf gesellschaftliche Integration von Migranten in Deutschland
19	Der Einfluss von Kindern auf Zeitallokation von Haushalten
20	Effekt von Zeitverwendung auf die Ausbildung von nicht-kognitiven Fähigkeiten
21	Arbeitszeit & Zeitbudgetanalysen – Analyse täglicher Arbeitszeiten und Nachfragearrangements
22	Soziale Ungleichheit und Prävention
23	Das Konzept der Europäischen Sozioökonomischen Klassifikation und seine Anwendung auf die in der Zeitbudgeterhebung 2001/02 befragten Haushalte
24	Renewbility
25	Substitutability of Partner's Productive Activities
26	Einkommensabhängiges Freizeitverhalten unter älteren Menschen
27	Zeit und soziale Ungleichheit. Die schichtspezifische Strukturierung sozialer Zeit – unter besonderer Beobachtung von Geschlecht und Generation
28	Schulz-Borck/Hofmann: Schadenersatz bei Ausfall von Hausfrauen und Müttern im Haushalt – mit Berechnungstabellen, 6. Aufl.-Karlsruhe: VVW 2000, ISBN 3-58487-89487-894-8
29	"Integration of Rebound Effects into Life-Cycle Assessment" (finanziert durch BFE und Nationalfonds)
30	Ruhestandsmigration in Deutschland
31	Assisted Living – Technisch unterstüztes Wohnen im Alter, Teilprojekt: Sozialwissenschaftliche Begleitforschung
32	Sozioökonomische Berichterstattung (soeb.de)
33	"Einkommen und Freizeit – Eine empirische Analyse des Freizeitverhaltens älterer Menschen mit Daten der Zeitbudgeterhebung des Statistischen Bundesamtes"

34	Ökonomische Analyse der Zeitverwendung für Ernährung
35	Integrierter Survey
36	International Evidence on housework and market work by husbands and wives
37	Entwicklungstendenzen im Online-Printmedienbereich in Deutschland – Arbeitsmarktstatistische und Arbeitsorganisatorische Analyse der Srukturveränderungen durch das Internet für Journalisten, 1990 - Gegenwart
38	Erstellung von Tabellen für das Seminar zur Wirtschaftslehre des Haushalts, in dem Studierende den Zeitaufwand für Kinder in den unterschiedlichen Haushaltstypen vergleichen sollen
39	Stochaistische Modellierung von Nutzerverhalten in Wohngebäuden
40	A cross-cultural analysis of overreporting of socially desirable behavior
41	Bezogenes Verkehrsverhalten von Beschäftigten im sekundären und tertiären Sektor
42	Potentiale der Zeitbudgeterhebung 2001/02 Eine Bestandsaufnahne anhand der Zeitverwendung "Junger Alter"
43	Berichtete und tatsächliche Kirchgangshäufigkeit in Ost- und Westdeutschland
44	Comparative Study on the Double Burden of Working Parents; Gender Differences in Time Poverty
45	Zeitverwendung von Arbeitslosen für Arbeitssuche
46	Soziale Netzwerke und Hilfebeziehungen im unteren Einkommensbereich
47	Zeitbudgeterhebungen – Methodik und Anwendungen
48	Analyse der Verschiebungen zwischen Wegezeiten und Zeiten für andere Aktivitäten in Abhängigkeit von der Raumstruktur
49	PACT (Pathways for carbon transitions)
50	Der zweite demographische Übergang

Source: The Research Data Center of the Federal Statistical Office, Wiesbaden 3/2009.

Appendix A2: Time Use Survey – National plans for the next wave of surveys 2008–2010

Country	Foreseen schedule	Comment
Belgium (BE)	2010	Statistics Belgium collects TUS data and Vrije Universiteit Brussel analyzes them. Next data collection will take place in 2010, analysis in 2011.
Bulgaria (BG)	2009/2010	Survey will be included into the National Program for Statistical Surveys 2009/2010.
Czech Republic (CZ)	Not before 2010	The implementation of TUS has not yet begun (no plan exists). There is a lack of financial resources and human capacity, the respondents' burden is still increasing, and neither TUS nor related activities are the priority of Czech Statistical Office in the area of social statistics.
Denmark (DK)	2008/2009	DTUC-Danish Time Use and Consumption Survey by Rockwool Foundation (Pilot ongoing).
Germany (DE)	No schedule	The next wave of the TUS survey is not yet organized and financed.
Estonia (EE)	2009/2010	EE is planning a TUS by 2009/2010.
Ireland (IE)	Not before 2010	The National Development Plan Gender Equality Unit, which was based in the Department of Justice, Equality and Law Reform, engaged the ESRI to carry out a pilot light diary survey in 2005. The report is available to download at: http://www.justice.ie/en/JELR/Pages/Time_use_survey_report Anonymized microdata is available through the Irish Social Science Data Archive (ISSDA), see: http://www.ucd.ie/issda/ dataset-info/timeuse.htm However, with the exception of this 2005 light diary pilot and a small CSO HETUS pilot carried out in one region of Ireland (Munster) in 1998, to date no national time use study has been carried out in Ireland. There are no definite plans to carry out a HETUS based or light diary survey at present.
Greece (EL)	No schedule	There is a lack of "economic and human resources."
Spain (ES)	2009/2010	ES plans a TUS in 2009/2010. Fieldwork between 10/2009 and 9/2010.
France (FR)	September 2009-August 2010	
Italy (IT)	2008/2009	Fieldwork between February 2008 and January 2009.
Cyprus (CY)	Not before 2013	It is unlikely that TUS will be launched before 2013.
Latvia (LV)	Not before 2011	It is difficult to have a precise plan at this moment. This depends on financial resources.

Country	Foreseen schedule	Comment
Lithuania (LT)	Not before 2010	It is difficult to have precise plan at this moment. This depends on financial resources.
Luxembourg (LU)	Not before 2010/2011	First, they have to integrate the Time Use Survey in their national plan. Thus, it is difficult to have a precise plan for the moment (financing and human resources must be confirmed). It is unlikely that the survey will be launched before 2010–2011.
Hungary (HU)	2009 or 2010	They plan to organize a TUS during 2009 or 2010. Only a pilot (with a n=100 sample) will be made. If it is successful, the results of this pilot can be used to emphasize the importance of such a survey. It is not easy to find financial sources for a survey in Hungary, as it is not compulsory there.
Malta (MT)	No updated information	The previous TUS survey was carried out in 2002.
Netherlands (NL)	2010	Previous TUS surveys: 1. 2005 applying national methodology 2. 2006 according to HETUS guidelines In 2010, they will either apply their national methodology or the Hetus methodology. They have to weight the pros and cons of both methodologies before they reach a decision.
Austria (AT)	2008/2009	Fieldwork from March 2008 until February 2009. The sample for TUS will be a subsample of the Austrian Microcensus. In addition to the Microcensus questionnaire, persons in the selected households will be asked to fill in a diary for one day (aim: net sample of 8,000 persons being 10 years and older). There will be no special TUS questionnaire.
Poland (PL)	(2012) 2014	It is impossible for Poland to carry out TUS in 2010 because of the Agricultural Census in 2010 and the National Census in 2011. The most likely and convenient time for the Polish CSO is 2013/2014, but it will be considered in 2012. This depends on financial resources.
Portugal (PT)	No schedule	It is not planned and depends on financial resources.
Romania (RO)	2009/2010?	The Romanian National Institute of Statistics could not carry out TUS in 2008/2009 due to a lack of financial and human resources. They provisionally planned the survey to be launched in 2009/2010, which depends on financial and human resources.
Slovenia (SI)	No schedule	Slovenia did not plan to incorporate financial resources and employees for the TUS in the medium term plan. A TUS will not be conducted in the near future.
Slovakia (SK)	Not before 2010	Previous TUS surveys: In 2006, the Pilot project on TUS, in accord with the 2004 HETUS guidelines, was carried out. A plan for regular TUS (not earlier than 2010) depends on obtaining of financial resources.

Country	Foreseen schedule	Comment
Finland (FI)	2009/2010	Fieldwork between April 2009 and March 2010.
Sweden (SE)	2010 if resources available	Regarding the next round of TUS, there is an ongoing discussion with the Ministry for Integration and Gender concerning financing. There is a great interest in taking part in the next round.
United Kingdom (UK)	Full survey: not before 2013. Exploring lower cost options (e.g., collecting basic data via an existing survey)	The UK carried out a light diary survey over 4 months in 2005. With regard to a HETUS survey, there appears to be no prospect of funding a full survey in the current planning period (2008-2012) given other priorities and budgetary pressures. ONS is still exploring lower cost options (e.g., collecting basic data via an existing survey), but this will also depend on the provision of financial resources from government and the ESRC.
Croatia (HR)	No schedule	National plan to be confirmed.
FYROM (MK)	2009	According to the working plan 2008-2012, TUS will be carried out in 2009. Fieldwork will start on 1 January 2009.
Turkey (TR)	2011	The previous TUS survey was carried out in 2006 and the results published in July 2007. The Turkish Statistical Institution, TURKSTAT, has planned to carry out TUS for a 5-year-period in line with HETUS guidelines.
Norway (NO)	2010	
Switzerland (CH)	Not before 2011	No TUS is planned at the Swiss Federal Statistical Office (FSO). In the context of the new Statistical System on Households and Persons, the possibility of a mini-TUS added to the omnibus survey is being examined (light diary, CATI-interviews with precoded activities). It would be realized in 2011 at the earliest. The decision is still open.

Source: German Federal Statistical Office 2009 (situation as of November 4, 2008)

References:

Akarro, R. (2008): The Impact of Time Use Differentials on Poverty levels in the Eastern and Northern Zones of Tanzania. European Journal of Economics, Finance and Administrative Sciences 13, 108-120.

Andorka, R. (1987): Time budgets and their uses. Annual Review of Sociology 13, 149-164.

Anxo, D. and Carlin, P. (2004): Intra-family time allocation to housework: French evidence. electronic International Journal of Time Use Research 1, 14-36. www.eIJTUR.org.

Ås, D. (1978): Studies of time use: Problems and prospects. Acta Sociologica 21 (2), 125-141.

Bediako, G. and Vanek, J. (1999): Trial International Classification of Activities for Time Use Statistics. In: Merz, J. and Ehling, M. (Eds.): Time Use – Research, Data and Policy. FFB-Book Series 10, 151-164.

Bevans, G.E. (1913): How Working Men Spend Their Spare Time. New York.

Bundesministerium für Arbeit und Soziales (2008): Der 3. Armuts- und Reichtumsbericht der Bundesregierung. Berlin. http://www.bmas.de/coremedia/generator/26896/lebenslagen-in-deutschland-der-3-armuts-und-reichtums-bericht-der-bundesregierung.html.

Bundesministerium für Familie, Senioren, Frauen und Jugend (2006): Familie zwischen Flexibilität und Verlässlichkeit – Perspektiven für eine lebenslauf-bezogene Familienpolitik. Siebter Familienbericht, Deutscher Bundestag, Drucksache 16/1360.

Chadeau, A. (1985): Measuring Household Activities – Some International Comparisons. Review on Income and Wealth 31, 237-254.

Cornelißen, W. (Ed.) (2005): Gender-Datenreport, 1. Datenreport zur Gleichstellung von Frauen und Männern in der Bundesrepublik Deutschland. Im Auftrag des Bundesministeriums für Familie, Senioren, Frauen und Jugend. 2. Fassung. Munich.

Cornelißen, W. and Blanke, K. (2004): Zeitverwendung von Mädchen und Jungen. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.

Dressel, Chr./Cornelißen, W. and Wol, K. (2005): Vereinbarkeit von Familie und Beruf. In: Cornelißen, W. (Ed.) (2005): Gender-Datenreport, 1. Datenreport zur Gleichstellung von Frauen und Männern in der Bundesrepublik Deutschland. Im Auftrag des Bundesministeriums für Familie, Senioren, Frauen und Jugend. 2. Fassung. Munich. http://www.bmfsfj.de/Publikationen/genderreport/root.html.

Ehling, M. and Merz, J. (Eds.) (2002): Neue Technologien in der Umfrageforschung – Anwendungen bei der Erhebung von Zeitverwendung [New Technologies in Survey Research – Applications for Time Use Studies]. FFB-Schriftenreihe Band 14. Baden-Baden.

Ehling, M. et al. (2001): Zeitbudget in Deutschland – Erfahrungsberichte der Wissenschaft. Spektrum der Bundesstatistik, Bd. 17. Wiesbaden

Engstler, H./Mennin, S./Hoffmann, E. and Tesch-Römer, C. (2004): Die Zeitverwendung älterer Menschen. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.

Eurostat (2009): Harmonised European Time Use Surveys – 2008 Guidelines, EUROSTAT Methodologies and Working papers. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-08-014/EN/KS-RA-08-014-EN.PDF.
Eurostat, Task Force Household Satellite Accounts (2003): Household Production and Consumption, Proposal for a Methodology of the Household Satellite Accounts. Task Force Report for EUROSTAT, Division E1.
Fahr, R. (2005): Loafing or Learning? The Demand for Informal Education. European Economic Review 49 (1), 75-98.
Fendrich, S. and Schilling, M. (2004): Informelle Betreuungssettings in der ausserfamilialen Kinderbetreuung. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.
Folbre, N. (2009): Inequality, Consumption, and Time Use. In: Nolan, B. and Smeeding, T.M. (Eds.): The Oxford Handbook of Economic Inequality. Wiemer Salverda, Chapter 14.
Fritz, I. und Klingler, W. (2006): Medienzeitbudgets und Tagesablaufverhalten. Media Perspektiven 4/2006. http://www.media-perspektiven.de/uploads/tx_mppublications/04-2006_Fritz.pdf.
Gabriel, O./Trüdinger, E.-M. and Völkl, K. (2004): Bürgerengagement in Form von ehrenamtlicher Tätigkeit und sozialen Hilfsleistungen. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.
Garhammer, M. (2008): Arbeitszeit, Zeitnutzung von Familien und Zeitpolitiken in Europa. http://www.opus-bayern.de/ohm-hochschule/frontdoor.php?source_opus=12.
Gershuny, J. (1995): Time budget research in Europe. Statistics in Transition 2 (4), 529-531.
Gershuny, J. (1999): Informal Economic Activity and Time Use Evidence. In: Merz, J. and Ehling, M. (Eds.): Time Use – Research, Data and Policy.
Gershuny, J. (2001): Time Use Research Methods. In: Baltes, P. and Smelser, N. (Eds.): International Encyclopedia of the Social & Behavioral Sciences. Amsterdam.
Gershuny, J. and Sullivan, O. (1998): The Sociological Uses of Time use Diary Analysis. European Sociological Review 14 (1), 69-85.
Gershuny, J./Fisher, K./Jones, S. and Gauthier, A. (2000): A Multinational Longitudinal Time use Data Archive. In: Gershuny, J. (Ed.): Changing Times: Work and Leisure in Postindustrial Society. Oxford.
Gille, M. and Marbach, J. (2004): Arbeitsteilung von Paaren und ihre Belastung mit Zeitstress. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.
Goldschmidt-Clermont, L. (1993): Monetary Valuation of Non-Market Productive Time – Methodological Considerations. Review of Income and Wealth 4, 419-433.
Guryan, J./Hurst, E. and Kearney, M.S. (2008): Parental Education and Parental Time With Children. NBER Working Paper No. 13993. New York. http://www.nber.org/papers/w13993.
Gwozdz, W./Hufnagel, R./Seel, B. and Wahrig, L. (2006): Messung der Entwicklung der geschlechtsspezifischen Arbeitsteilung mit den Daten der Zeitbudgeterhebungen 1991/92 und 2001/02. Hauswirtschaft and Wissenschaft 1, 22-28.
Hamermesh, D. (2002): Timing, togetherness and time windfalls. Journal of Population Economics 15, 321-325.
Hamermesh, D.S. and Pfann, G.A. (2005). The Economics of Time Use. North-Holland.

Harms, T. and Gershuny, J. (2009): Time Budgets and Time Use. RatSWD Working Paper No. 45.
Harvey, A. (1999): Time Use Research: The Roots to the Future. In: Merz, J. and Ehling, M. (Eds.): Time Use – Research, Data and Policy. FFB-Schriftenreihe 10. Baden-Baden.
Harvey, A. (2004): Welcome Address of the IATUR President: eIJTUR and Time use: Past, Present and Future. electronic International Journal of Time Use Research 1 (1), I-IV. www.eIJTUR.org.
Harvey, A. (2009): Halifax Regional Space-Time Activity Research (STAR) Project – A GPS-Assisted Household Time use Survey, St. Mary's University. Halifax. www.stmarys.ca/partners/turp/main.html.
Harvey, A./Szalai, A./Elliott, D.H../Stone, P.J. and Clark, S. (1984): Time Budget Research: An ISSC Workbook in Comparative Analysis. New York.
Harvey, A. (2006): Nonmarket Production and Historical Time use Data: Potential and Issues, Yale Program on Nonmarket Accounts: A Project on Assessing Time Use Survey Datasets. Halifax.
Hoffmann, E. and Mata, A. (1999): Measuring Working Time: An Alterbnative Approach to Classifying Time Use. In: Merz, J. and Ehling, M. (Eds.): Time Use – Research, Data and Policy. FFB-Book Series 10.
Holz, E. (2004): Poverty and family life: Time use of families with and without poverty risk, Proceedings of the Annual IATUR Conference 2004, International Association for Time Use Research – IATUR and Italian National Statistical Institute – ISTAT. Rome.
Institut für Sozialforschung und Gesellschaftspolitik (ISG), Köln (Ed.) (2004): Aspekte der Armuts- und Reichtumsberichterstattung: Reichtum und Eliten – Haushaltsproduktion und Armutsprävention. 2. Wissenschaftliches Kolloquium am 8. und 9. Oktober 2003 in Rüdesheim am Rhein. Veranstaltung des Bundesministeriums für Gesundheit und Soziale Sicherung. Durchführung und Dokumentation: Institut für Sozialforschung und Gesellschaftspolitik. http://www.isg-institut.de/index.php?b=single&id_B=6&id_UB=8&id_Nummer=83.
Ironmonger, D. (2001): Household Production and the Household Economy. Department of Economics Research Paper No. 759. University of Melbourne.
Jäckel, M. and Wollscheid, S. (2004): Mediennutzung im Tagesverlauf: Ausweitung des Angebots und Strukturen der Zeitverwendung. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.
Jäckel, M. and Wollscheid, S. (2007): Time Is Money and Money Needs Time? A Secondary Analysis of Time-Budget Data in Germany. Journal of Leisure Research 39 (1), 86-108.
Jacobsen, J. and Kooreman, P. (2004): Timing Constraints and the Allocation of Time: The Effect of the Changing Shopping Hours Regulations in the Netherlands. IZA Discussion Paper No. 130.
Jahoda, M./Lazarsfeld, P.F. and Zeisel, H. (1933): Marienthal: The Sociography of an Unemployed Community – Translated by the authors with John Reginait and Thomas Elsaesser. Chicago 1971.
Juster, F.Th. (1999): The Future of Research on Time Use. In: Merz, J. and Ehling, M. (Eds.): Time Use – Research, Data and Policy. FFB-Schriftenreihe Band 10. Baden-Baden.

Juster, F.Th. and Stafford, F.P. (1991): The allocation of time: empirical findings, behavioural models and problems of measurement. In: Journal of Economic Literature 29, 471-522.
Kahle, I. (2004): Alleinerziehende im Spannungsfeld zwischen Beruf und Familie. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.
Keall, M.D. and Baker, M.G. (2008): Travel Survey Data to Estimate Time Spent at Important Settings. electronic International Journal of Time Use Research 5, 127-129.
Kettschau, I./Hufnagel, R. and Holz, E. (2004): Lebensgestaltung auf Haushaltsebene – Verknüpfung zwischen Armutsforschung und Zeitbudgetdaten. Expertise für das Bundesministerium für Familie, Senioren, Frauen und Jugend, Beitrag zum Berichtskapitel für den 2. Armuts- und Reichtumsbericht der Bundesregierung. Fachhochschule Münster, Fachbereich Oecotrophologie. Münster.
Kleimann, M. and Mößle, T. (2008): The Logs of Eliza and Other Media Stories. Behavioral and Developmental Effects of a School Based Media Education Program – Berlin Longitudinal Study Media. International Journal of Behavioral Development 32 (6), Supplement 2 (54), 55-59.
Klevmarken, A. (1999): Microeconomic Analysis of Time Use Data: Did We Reach the Promised Land? In: Merz, J. and Ehling, M. (Eds.): Time Use – Research, Data and Policy. FFB-Schriftenreihe Band 10. Baden-Baden.
Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Ed.) (2001): Wege zu einer besseren informationellen Infrastruktur. Baden-Baden
Körner, T./Nimmergut, A./Nökel, J. and Rohloff, S. (2008): Die Dauerstichprobe befragungsbereiter Haushalte – Die neue Auswahlgrundlage für freiwillige Haushaltsbefragungen. Wirtschaft und Statistik 5/2008, 451-487.
Kramer, C. (2004): Zeitverwendung. In: Statistisches Bundesamt (Ed.): Datenreport 2004.
Kramer, C. (2005): Verkehrsverhalten, Wegezeiten und Mobilität. Expertise für den Siebten Familienbericht der Bundesregierung. Berlin.
Kramer, C. (2005): Zeit für Mobilität – Räumliche Disparitäten der individuellen Zeitverwendung für Mobilität in Deutschland. Reihe Erdkundliches Wissen, Band 138. Stuttgart.
Landefeld, J.S. and Culla, S. (2000): Accounting for Nonmarket Household Production within a National Accounts Framework. Review of Income and Wealth 46 (3).
Lindqvist, M. (2001): The Finnish System of Official Social Surveys, Statistics Finland, Official Social Surveys in Europe. Edited by F. Kraus and G. Schmaus. EuReporting Working Paper No. 25.
Meier, U./Küster, Chr. and Zander, U. (2004): Alles wie gehabt? – Geschlechtsspezifische Arbeitsteilung und Mahlzeitenmuster im Zeitvergleich. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.
Merz, J. (1990): Zur Notwendigkeit und Nutzung von Zeitbudgetdaten in der Sozialökonomie. In: v. Schweitzer, R./Ehling, M. and Schäfer, D. (Eds.): Zeitbudgeterhebungen. Ziele, Methoden und neue Konzepte. Stuttgart.

Merz, J. (1992): Zur Dynamik markt- und nichtmarktmäßigen Arbeitsangebots: Zeitverwendung verheirateter Frauen in Beruf und Haushalt – Eine Panelanalyse. In: Hujer, R./Schneider, H. and Zapf, W. (Eds.): Herausforderungen an den Wohlfahrtsstaat im strukturellen Wandel. Frankfurt/New York.

Merz, J. (1997): Zeitverwendung in Erwerbstätigkeit und Haushaltsproduktion – Dynamische Mikroanalysen mit Paneldaten, DFG-Endbericht, FFB-Dokumentation Nr. 7, Fachbereich Wirtschafts- und Sozialwissenschaften, Universität Lüneburg.

Merz, J. (2001): Informationsfeld Zeitverwendung. In: Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Ed.): Wege zu einer besseren informationellen Infrastruktur. Baden-Baden.

Merz, J. (2002): Time Use Research and Time Use Data – Actual Topics and New Frontiers. In: Ehling, M. and Merz, J. (Eds.): Neue Technologien in der Umfrageforschung – Anwendungen bei der Erhebung von Zeitverwendung (New Technologies in Survey Research – Applications for Time Use Studies). FFB-Schriftenreihe Band 14. Baden-Baden.

Merz, J. (2009): Zeitverwendungsforschung und Mediennutzung, Gutachten für das Zweite Deutsche Fernsehen (ZDF). FFB-Diskussionspapier Nr. 75, Leuphana Universität Lüneburg.

Merz, J. and Böhm, P. (2005): Arbeitszeitarrangements und Entlohnung – Ein Treatment-Effects-Ansatz für Freiberufler, Unternehmer und abhängig Beschäftigte. In: Schulte, R. (Ed.): Ergebnisse der MittelstandsForschung. Münster.

Merz, J. and Burgert, D. (2004): Arbeitszeitarrangements – Neue Ergebnisse aus der Zeitbudgeterhebung 2001/02 im Zeitvergleich zu 1991/92. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. FORUM Band 43. Wiesbaden.

Merz, J. and Ehling, M (Eds.) (1999): Time Use – Research, Data and Policy. FFB-Schriftenreihe Band 10. Baden-Baden.

Merz, J. and Rathjen, T. (2009): Time and Economic Poverty – Interdependent Multidimensional Poverty with the German Time Use Survey 2001/02. Manuscript, Research Institute on Professions (FFB), University of Lüneburg.

Merz, J. Hanglberger, D. and Rucha, R. (2009): The Timing of Daily Demand for Goods and Services – Multivariate Probit Estimates and Microsimulation Results for an Aged Population with German Time Use Diary Data. FFB Discussion Paper No. 75, Department of Economics, Behavioural and Law Sciences, Leuphana University.

Merz, J./Böhm, P. and Burgert, D. (2004): Timing, Fragmentation of Work and Income Inequality – An Earnings Treatment Effects Approach. FFB Discussion Paper No. 47, Department of Economics and Social Sciences, University of Lueneburg.

Mückenberger, U. (2008): Zeitpolitik in der Metropolregion Hamburg, Projektantrag „Wissenschaftliche Begleitung des Projektverbunds ‚Zeitpolitik in der Metropolregion Hamburg'". Hamburg.

Niemi, I. (1993): Systematic error in behavioral measurement: Comparing results from interview and time budget studies. Social Indicators Research 30 (2-3), 229-244.

Pember Reeves, M. (1913): Round About a Pound a Week, Bell. London. New Edition: 1990.

Pinl, Cl. (2004): Wo bleibt die Zeit? Die Zeitbudgeterhebung 2001/02 des Statistischen Bundesamts. Aus Politik und Zeitgeschichte Nr. 31-32. http://www.bpb.de/publikationen/CRWJ8M,0,Wo_bleibt_die_Zeit.html.
Riediger, M. (2009): Experience Sampling, Federal Ministry of Education and Research. RatSWD Working Paper No. 62.
Riediger, M. et al. (2009): Seeking Pleaseure and Seeking Pain. Psychological Science 20 (12), 1529-1535.
Robinson, J.P. (1985): The validity and reliability of diaries versus alternative time use measures. In: Juster, F.T. and Stafford, F.P. (Eds.): Time, Goods and Well-Being. Ann Arbour.
Robinson, J.P. and Godbey, G. (1997): Time for Life – The Surprising Ways Americans Use Their Time.
Rydenstam, K. (1999): The EUROSTAT Project on Harmonising Time Use Statistics. In: Merz, J. and Ehling, M. (Eds.): Time Use – Research, Data and Policy. FFB-Book Series, Vol. 10.
Rydenstam, K. (2007): The Harmonised European Time Use Database and the Table Generating Tool. electronic International Journal of Time Use Research (4-1), 118-119. www.eIJTUR.org.
Schäfer, D. (2004): Unbezahlte Arbeit und Haushaltsproduktion im Zeitvergleich. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.
Schäffer, S.M. (2003): Die Zeitverwendung von Konsumenten – Implikationen für das Dienstleistungsmarketing. Wiesbaden.
Schriftreihe Media Perspektiven (2006): Massenkommunikation VII, Eine Langzeitstudie zur Mediennutzung und Medienbewertung 1964 – 2005. Baden-Baden.
Schulz, F. and Grunow, D. (2007): Time-diary versus time-estimation data. A comparison of two different methods of measuring the time spent on housework. Zeitschrift für Familienforschung 1/2007, 106-128.
Schupp, J. and Gerlitz, J.-Y. (2008): Big Five Inventory-SOEP. In: Glöckner-Rist, A. (Ed.): Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 12.00. Bonn.
Sellach, B./Enders-Dragässer, U. and Libuda-Köster, A. (2004): Geschlechtsspezifische Besonderheiten der Zeitverwendung – Zeitstrukturierung im theoretischen Konzept des Lebenslagen-Ansatzes. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.
Stäglin, R. and Schintke, J. (2003): Monetäre, physische und Zeit-Input-Output-Tabellen, Teil 2: Analytische Auswertung, Sozio-ökonomisches Berichtssystem für eine nachhaltige Gesellschaft, Bd. 2. http://www-ec.destatis.de/csp/shop/sfg/vollanzeige.csp?ID=1011818.
Stahmer, C. (2003): Aufbau eines sozio-ökonomischen Berichtssystems für eine nachhaltige Gesellschaft. In: Statistisches Bundesamt (Ed.): Sozialer Wandel – Daten, Analysen, Gesamtrechnungen, Band 41 der Schriftenreihe „Forum der Bundesstatistik“. Wiesbaden.
Stahmer, C. and Schaffer, A. (2004): Time Pattern in a Social Accounting Framework, Paper prepared for International Association for Research in Income and Wealth – IARIW, 28th General Conference, August 22-28, 2004, Cork, Ireland, http://www.iariw.org/papers/2004/axel.pdf.

Stahmer, C./Mecke, I. and Herrchen, I. (2003): Zeit für Kinder. Betreuung und Ausbildung von Kindern und Jugendlichen, Sozio-ökonomisches Berichtssystem für eine nachhaltige Gesellschaft, Bd. 3.

Statistisches Bundesamt (2004): Alltag in Deutschland – Analysen zur Zeitverwendung, Forum der Bundesstatistik. Wiesbaden.

Statistisches Bundesamt (Ed.) (2006a): Compass 2001/02 Time Use Survey: Publications, Links and Prints incl. Data Files and Research Facilities

Statistisches Bundesamt (Ed.) (2006b): Zeitbudgets – Tabellenband I, Zeitbudgeterhebung: Aktivitäten in Stunden und Minuten nach Geschlecht, Alter und Haushaltstyp 2001/02. Wiesbaden. https://www-ec.destatis.de/csp/shop/sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1019589.

Statistisches Bundesamt (Ed.) (2008): Statistisches Jahrbuch 2008, Tabelle 7 Kultur, Freizeit, Sport, 7.15 Zeitaufwand pro Woche für Unterhaltung und Kultur 2001/2002 nach Geschlecht und Alter, 185. https://www-ec.destatis.de/csp/shop sfg/bpm.html.cms.cBroker.cls?cmspath=struktur,vollanzeige.csp&ID=1022845.

Szalai, A. et al. (Eds.) (1972): The use of time: Daily activities of urban and suburban populations in twelve countries. The Hague.

Torres, A./Brites, R./Haas, B. and Steiber, N. (2007): First European Quality of Life Survey: Time use and work–life options over the life course, European Foundation for the Improvement of Living and Working Conditions. Dublin. www.eurofound.europa.eu.

UNIFEM United Nations Development Funds for Women (2008): Who Answers to Women? Progress of the World's Women 2008/2009 – Gender & Accountability. http://www.unifem.org/progress/2008/.

US Department of Agriculture (2009): American Time Use Survey: Food & Eating Module. http://www.ers.usda.gov/emphases/ healthy/atus/questions.htm.

Weick, S. (2004): Lebensbedingungen, Lebensqualität und Zeitverwendung. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.

Wilhelm, R. and Wingerter, Ch. (2004): Lebenslanges Lernen – Statistischer Ansatz und empirische Ergebnisse der Zeitbudgeterhebung 2001/2002. In: Statistisches Bundesamt (Ed.): Alltag in Deutschland – Analysen zur Zeitverwendung. Wiesbaden.

Wilson, C. (1999): Sequence Alignment Analysis of Daily Activities. In: Merz, J. and Ehling, M. (Eds.): Time Use – Research, Data and Policy. FFB-Schriftenreihe Band 10. Baden-Baden.

Wilson, C. (2001): Activity patterns of Canadian women: An application of ClustalG sequence alignment software, Canada Mortgage and Housing Corporation, Paper 01-2292. Ottawa.

II. Methods

1. Survey Methodology: International Developments

Frauke Kreuter

Contact:

Frauke Kreuter
Joint Program in Survey Methodology
1218 LeFrak Hall
College Park, MD 20742
US
e-mail: fkreuter[at]survey.umd.edu

Abstract

Falling response rates and the advancement of technology have shaped discussions in survey methodology for the last few years. Both have led to a notable change in data collection efforts. Survey organizations are currently exploring adaptive recruitment and survey designs and have increased their collection of non-survey data for sampled cases. While the first strategy represents an attempt to increase response rates and save on cost, the latter shift can be seen as part of the effort to reduce the potential bias and response burden of those interviewed. To successfully implement adaptive designs and alternative data collection efforts, researchers need to understand the error properties of mixed-mode and multiple-frame surveys. Randomized experiments might be needed to gain that knowledge. In addition, there is a need for close collaboration between survey organizations and researchers, including the ability and willingness to share data. The expanding options for graduate and post-graduate education in survey methodology could also help to increase the potential for implementing high-quality surveys.

Keywords: survey methodology, responsive design, paradata

1. Introduction

Falling response rates (Schnell 1997; Groves and Couper 1998; de Leeuw and de Heer 2002) and the advancement of technology (Couper 2005) have shaped discussions in survey methodology for the last few years. This report will highlight some of the developments that have resulted from these two trends and discuss the increasing difficulty of conducting surveys in the same way that had been common throughout the 1970s, 1980s, and 1990s. It is impossible to capture all of the changes in survey practice that took place during that time. However, this report will address several of the most prominent developments that have been discussed within the survey methodology research community, and those that are not addressed in the other contributions to this publication.

All of the developments that will be discussed here share an increased flexibility in data collection efforts. At the same time, they illustrate design changes implemented in a controlled or even randomized way in order to assess their effects on individual error sources. The result is less of a streamlined, recipe-style approach to data collection. Unlike in Germany, the data infrastructure in the US and UK allows for this type of flexibility in contexts where survey organizations are closely tied to scientists at universities (e.g., University of Michigan) or in survey research organizations that act as primary investigators (for example, with NORC, the National Opinion Research Center, and the General Social Survey). In both countries, most of the data collection agencies used for social science research are organizations that specialize in surveys for research projects. The companies therefore tend to have an incentive to invest in developing the expertise necessary for conducting high-quality surveys.

The present report begins with a discussion of how response rates have functioned as a quality indicator for surveys, and then summarizes the current discussion of alternatives to response rates as indicators. I will then highlight recent developments within survey operations. Many of these developments are reactions to falling response rates and increased concerns about nonresponse bias; others are motivated more broadly by the larger issue of total survey error (Groves et al. 2004) or as a reaction to technological changes. The main question behind all these developments, however, is: "How can we ensure high-quality data collection in a changing survey environment and increase quality in existing studies?"

2. Response rates and other survey quality indicators

For years, both survey methodologists as well as the general public have focused on response rates as indicators of survey quality (Groves et al. 2008). This focus has changed in recent years. For one thing, even in surveys with traditionally high response rates, participation has fallen below expectations. In addition, empirical evidence over the last decade has increasingly demonstrated that nonresponse rates are poor indicators of nonresponse bias for single survey estimates (Keeter et al. 2000; Curtin et al. 2000; Groves 2006). The shift in focus away from nonresponse rates toward bias is evident in a number of areas. It can be seen, for example, in the guidelines established by the US Office of Management and Budget, which require a detailed plan for the evaluation of nonresponse bias before they approve data collection sponsored by federal statistical agencies.[1] It is also evidenced by

1 All data collections conducted or sponsored by the US federal statistical agencies have to be approved by the Office of Management and Budget (OMB), which ensures that performance standards developed by the Interagency Council on Statistical Policy (ICSP) are met (Graham 2006). Conducting or sponsoring is defined here as any information that the agency collects using (1) its own staff and resources, or (2) another agency or entity through a contract or cooperative agreement. The approval by OMB is not just an attempt to reduce burden on the respondents (see Paperwork Reduction Act) but to ensure "that the concepts that are being measured be well known and understood, and shown to be reliable and valid" (Graham 2006). OMB applications require information from the data collection agency on questionnaire design procedures, field tests of alternative versions of their measures, reinterviews with subsamples of respondents, and the like. Pretests and pilot studies are encouraged, and the OMB guidelines spell out how those can be conducted. No criteria are specified to quantify potential measurement error. The development of a plan to

research initiatives to develop alternative indicators for survey quality (Groves et al. 2008).

Alternative indicators of survey quality can be grouped into two sets (Groves et al. 2008): single indicators at the survey level (which is similar to the current use of the response rate), and individual indicators at the estimate level. Single indicators include *variance functions of nonresponse weights* (e.g., coefficients of variation of nonresponse weights), *variance functions of post-stratification weights* (e.g., coefficients of variation of poststratification weights), *variance functions of response rates* on subgroups defined for all sample cases (both respondents and nonrespondents), *goodness of fit statistics on propensity models,* and *R-indexes* (Shouten and Cobben 2007), which are model-based equivalents of the above. Researchers from the Netherlands, the UK, Belgium, Norway, and Slovenia formed a joint project (RISQ) to develop and study such R-indexes.

The second set of indicators is produced on the survey estimate level. It is evident that nonresponse bias is item-specific (Groves and Peytcheva 2008) and thus estimate-level indicators would have the soundest theoretical basis. Examples of estimate-specific indicators are: *comparisons of respondents and nonrespondents on auxiliary variables; correlation between post-survey nonresponse adjustment weights and the analysis variable of interest (y) measured on the respondent cases; variation of means of a survey variable y within deciles of the survey weights;* and *fraction of missing information on y.* The latter is based on the ratio of the between-imputation variance of an estimate and the total variance of an estimate based on imputing values for all the nonrespondent cases in a sample (Little and Rubin 2002; Wagner 2008; Andrige and Little 2008).

All of these attempts rely heavily on the availability of auxiliary variables, such as enriched sampling frames, interviewer observations, or other paradata correlated with the survey variables of interest. Thus, we cannot revise our survey quality indicators without also changing survey operations.

Survey operations – the procedures of data collection – are themselves subject to quality assessment and quality indicators. O'Muircheartaigh and Heeringa (2008) presented a set of criteria at the 3MC conference in Berlin. Another example for quality assessment of survey operations are the OMB guidelines.[2] Independent of those guidelines, there are a couple of recent developments in survey operations that are informative for the German data collection context.

---

evaluate nonresponse bias is required only in cases where projected unit response rate falls below 80 percent.

2 http://www.whitehouse.gov/omb/inforeg/statpolicy.html#pr. [Last visited: 03/02/2010].

3. Survey operations

While survey methodologists and statisticians are aware of the fact that response rates are a poor indicator of nonresponse error (Keeter et al. 2000; Groves 2006) and are even less suitable as an indicator of the overall survey quality (Groves et al. 2004), a drop in response rates has nevertheless been the catalyst that has engaged survey researchers in rethinking current practices. In the light of the increasing difficulty that has been encountered, growing cooperation has heightened the awareness of potential biases in surveys and created the need to evaluate survey procedures, which are faced with the threat of losing precision through decreasing sample sizes. Changes in fieldwork procedures require cost-quality trade-off decisions.

Surveys conducted with a *responsive design* use *paradata* to carry out these cost-quality trade-off decisions during the fieldwork stage. Such paradata are not only used as criteria for decision-making during field operations, but are increasingly seen as tools for evaluating measurement error or conducting post-survey bias adjustments. *Multiple-mode* surveys are often a response to cost-quality trade-off analyses prior to the start of the survey, but they are also a reaction to coverage problems that arise when mode-specific frames do not cover the entire population. An extreme form of multiple-mode surveys are those where the respondent recruitment is separated from the actual data capture. The most prominent examples are access panels or opt-in polls (discussed in other chapters of this publication and therefore omitted here).

3.1 Responsive design

Survey organizations have been using subsampling and two-phase designs for a long time. However, the design decisions were often only based on estimates of current response rates and qualitative information from field supervisors. These approaches were further hampered by the inability to reach every sample unit in the subsample, and thus the statistical properties of the two-phase design were not necessarily unbiased. Over the last decade, survey organizations in the US and some European countries have begun to systematically base design decisions on quantitative information gathered during early phases of the fieldwork. The most prominent and detailed published example of this comes from the Social Research Center at Institute for Social Research of the University of Michigan, in an article outlining the use of "responsive design" (Groves and Heeringa 2006). Responsive design is characterized by four stages in the survey process. First, design characteristics are identified that may affect survey cost and error. Second, this set of indicators is monitored during the initial stages of data collection. Third, in subsequent phases of data collection, the features of the survey are altered based on cost-error trade-off decision rules. Finally, data from the separate phases are combined into a single estimator. One example of the kinds of data collected are the hours spent by an interviewer calling on sample households, driving to sample areas, conversing with household members, and interviewing individuals in the sample.

One critical element of this type of responsive design is the ability to track key estimates as a function of estimated response propensities (conditioned on a design protocol). If survey variables can be identified that are highly correlated with the response propensity, and if it can be seen that point estimates of such key variables are no longer affected by extending the field period, then one can conclude that the first phase of a survey (with a given protocol) has reached its phase capacity and a switch in recruitment protocol is advisable. Using non-contact error as an example, one can expect that a given recruitment protocol has reached its capacity if the percentage of households with access impediments stabilizes with repeated application of the recruitment protocol (e.g., repeated callbacks). Applying this method, Groves and Heeringa (2006) concluded that, for the National Survey of Family Growth (NSFG) cycle-6 field period, 10–14 calls produced stable cumulative estimates on the vast majority of the key estimates. A necessary condition for tracking key survey estimates concurrently is the ability and willingness of interviewers not only to record respondent data and paradata electronically, but also to submit the data to the survey managers in a timely manner. In the case of NSFG, the submissions occurred every evening (Wagner 2008).

3.2 Paradata

Paradata (data about the process of data collection) were already mentioned as an important tool for guiding fieldwork decisions (see Kreuter 2010). Increasingly, paradata are also used as tools for survey nonresponse adjustment and for the detection and modeling of measurement error. The latter is already more common in online surveys, where keystroke files are readily available due to the nature of the task. Even face-to-face surveys now have the capacity to electronically capture survey process data. Some examples of this include keystroke files obtained from computer-assisted personal interviews (CAPI), the audio computer-assisted self interview (Audio-CASI) surveys (Couper et al. 2008), and digital recordings of the (partial) interviews.

Paradata of potential use for nonresponse adjustments are collected in conjunction with household listings and when contact attempts to sample units are made. Recently, the US Census Bureau began to employ an automated system for collecting contact histories for CAPI surveys (Bates et al. 2008). Other governments have started using similar procedures. For example, the Research Center of the Flemish Government (Belgium) began to use contact forms in their surveys based on the work of Campanelli et al. (1997). The time of contact (day and time), the data collection method (in person or by telephone), and other information is recorded for each contact attempt with each sample unit (Heerwegh et al. 2007). A standard contact form has also been implemented since 2002 (round one) of the European Social Survey, and contact data were recently released publicly by the US National Center for Health Statistics (NCHS) for the 2006 National Health Interview Survey (NHIS). Thus, contact protocol data are increasingly available for each sample unit, which makes those data an attractive source for nonresponse-adjustment variables. Other large survey projects that collect observations of neighborhoods and housing unit characteristics include the 2006 Health and Retirement Study (HRS), Phase IV of the Study of Early Child Care (SECC), the Survey of Consumer Finances (SCF), the National Survey on Drug Use and Health, the British Election Study (BES), the British Crime Survey, the British Social Attitudes Survey (BSA), and the Survey of Health, Ageing and Retirement in Europe (SHARE).

Inspired by Groves and Couper (1998), some researchers have been able to use interviewer observations to assess the likelihood of response. Copas and Farewall (1998) successfully used the interviewer-assessed interest of sample members about participating in the British National Survey of Sexual Attitudes and Lifestyle as a predictor of response. Lynn (2003) demonstrated that the presence of multi-unit structures and door intercoms predicted the amount of effort required to contact sample households in the British Crime Survey. Bates et al. (2006) used contact information from the 2005 NHIS to predict survey participation. They examined the effect of various respondent questions, concerns, and reasons given for reluctance as they were recorded by interviewers on the survey response. For the US National Survey of Family Growth, Groves and Heeringa (2006) used a series of process and auxiliary variables to predict the screening and interview propensity for each active case. The expected screening and interview propensities were summed over all cases within a sample segment and grouped into propensity strata. The propensity strata were used by supervisors to direct the work of interviewers. Propensity models using call record paradata were also estimated for the Wisconsin Divorce Study (Olson 2007) and the US Current Population Survey (Fricker 2007). Both Olson (2007) and Fricker (2007) then examined measurement error as a function of response propensity. Lately, more studies have tried to establish a relationship between paradata collected during the contact process (or as interviewer observations) and key survey variables (Schnell and Kreuter 2000; Asef and Riede 2006; Peytchev and

Olson 2007; Groves et al. 2007; Yan and Raghunathan 2007; Kreuter et al. 2007).
A systematic evaluation of the quality of such paradata, however, is very limited. For example, measurement error properties of these data, collected either through interviewer observation or through digital recordings of timing or speech, are currently being studied by Casas-Cordero (2008) and Jans (2008).

3.3 Auxiliary variables and alternative frames

Next to paradata there is a second set of data sources that is now of increasing interest to survey designers – commercial mass mailing vendors. These lists are of interest for their use in the creation of sampling frames, to enhance survey information and to evaluate nonresponse bias.

In face-to-face surveys in the US, two methods of infield housing unit listing are most common. Traditional listing provides listers with maps showing the selected area and an estimate of the number of housing units they will find. Dependent listing gives listers sheets preprinted with addresses believed to lie inside the selected area. Those addresses come either from a previous listing or from a commercial vendor. Listers travel around the segment and make corrections to the list to match what they see in the field. The latter appears to be less expensive (O'Muircheartaigh et al. 2003). There is a third method of creating a housing unit frame, which involves procuring lists of residential addresses from a commercial vendor and identifying those that fall within the selected areas. Here, geocoding is used instead of actual listings. The coverage properties of such frames are still under study (Iannacchione et al. 2003; O'Muircheartaigh et al. 2006; Dohrmann et al. 2007; O'Muircheartaigh et al. 2007; Eckmann 2008). Survey research organizations are currently exploring the US Postal Service delivery sequence files to replace traditionally used PSUs (Census blocks) with zip codes. While this last development is specific to the US, it is nevertheless of interest as it holds out the potential to stratify with rich datasets, or to inform interviewers in advance about potential residents and their characteristics. This information can be used for tailored designs. In Germany, dependent listing and enhanced stratification was already used for the IAB-PASS study (Schnell 2007).

3.4 Multiple modes

Several US federal statistical agencies have explored the use of mixed mode surveys. The two main reasons that mixed mode studies are usually considered relate to survey cost and response rates. There are three prominent types of multiple-mode studies: modes are administered in sequence, modes are implemented simultaneously, or a primary mode is supplemented with a secondary mode (de Leeuw 2005).

The American Community Survey (ACS), which replaced the Census long form, is an example of a sequential application of modes. Respondents are first contacted by mail, nonrespondents to the mail survey are contacted on the phone (if telephone numbers can be obtained), and finally in-person follow-ups are made to a sample of addresses that have not yet been interviewed. Parallel to the primary data collection, a method sample is available to examine various error sources (Griffin 2008). The Bureau of Labor Statistics (BLS) is currently using multiple modes for the Current Employment Statistics (CES) program. Firms are initiated into the survey via a computer-assisted telephone interview (CATI), kept on CATI for several months, and are then rolled over to touchtone data entry, the internet, fax, etc.[3] Experiments are undertaken to evaluate measurement error separately from nonresponse error for each of these modes (Mockovak 2008). The National Survey of Family Growth has CAPI as its primary mode, although sensitive information (e.g., number of abortions) is collected through Audio-CASI.

With their responsive design and the acknowledgement of imperfect sampling frames, mixed-mode surveys present some attractive advantages. Research is underway to explore the interaction between nonresponse and measurement error for these designs (Voogt and Saris 2005; Krosnick 2005). The European Social Survey program just launched a special mixed-mode design in four countries to examine appropriate ways of tailoring data collection strategies and to disentangle mode effects into elements arising from measurement, coverage, and sample selection. Another large scale study within Europe that experiments with mixed-modes is the UK Household Longitudinal Study (UKHLS), under the supervision of the Institute for Social and Economic Research. On the administrative side, the Social Research Center at the University of Michigan is currently constructing a new sample management system that will allow more efficient ways of carrying out mixed-mode surveys (Axinn et al. 2008). The new system will manage samples across data collection modes (F2F, telephone, Internet, and supplementary data modes such as biomarkers, soil samples, etc.) and will allow easy transfer of samples between modes and interviewers (e.g., between CAPI and centralized CATI).

3 http://www.bls.gov/web/cestn1.htm. [Last visited: 03/02/2010].

3.5 Reduction of response burden

Another development related to measurement error can be seen most recently in the context of large-scale surveys. Researchers at the Bureau of Labor Statistics (BLS) are investigating survey re-design approaches to reduce respondent burden in the Consumer Expenditure Survey (Gonzalez and Eltinge 2008). One proposed method is multiple matrix sampling, a technique for dividing a questionnaire into subsets of questions and then administering them to random subsamples of the initial sample. Matrix sampling has been used for a long time in large-scale educational testing. This method is growing in popularity for other types of surveys (Couper et al. 2008) where respondent burden is an increasing concern. Another method from educational testing that is currently under exploration is adaptive testing. Most applications of this method are currently tested in health surveys but survey issues regarding context effects arise (Kenny-McCoullough 2008).

3.6 Interviewer

All of the above mentioned developments have one feature in common – they alter and extend the task interviewers have to perform. In the past, there was already a tension between the dual role of interviewers. On the one hand, they have to be adaptive and flexible when recruiting respondents into the sample (Groves and McGonagle 2001; Maynard et al. 2002), and on the other hand, interviewers are asked to deliver questions as standardized as possible to reduce interviewer effects (Schnell and Kreuter 2005). Now, however, the number of tasks that one interviewer is required to perform is even higher, including recording observations, bookkeeping, handling technology, explaining technology, switching between different questionnaire flows, etc. Considering this increased burden and the resulting higher expectations placed on the interviewer, a more careful look at interviewer performance seems necessary. Survey organizations (NORC in the US, NatCen and ONS in the UK) have already started to analyze interviewer performance across various surveys (Yan et al. 2008) combined with census data (Durrant et al. 2008) or questionnaires given to the interviewer (Jäckle et al. 2008); others investigate alternatives to conventional interviewers (Conrad and Schober 2007).

Compared to Germany, it seems more common for US data collection firms to employ interviewers that work for one particular survey organization (and thus become acclimated to a particular survey house culture), or, if they do work with other organizations, these would also be social survey research organizations. More importantly, it is common in the US for interviewers to be centrally trained from the survey agency at the beginning of their employment and also at the beginning of new large-scale assignments. Unlike in Germany, face-to-face survey interviewers tend to be paid by the hour rather than by completed cases. This results in a different incentive structure and also opens the possibility for interviewers to spend time on the additional tasks mentioned above. It goes without saying that the cost of face-to-face surveys in the US is often tenfold that of what is typical in Germany.

4. Summary

In conclusion, survey methodologists are conducting new and exciting research into the trade-offs between cost and response rates. As part of these efforts, research is being done on how best to use non-survey data to provide information about nonresponse bias or measurement error, but also to supplement data collection and reduce respondent burden. Research is underway to gain a better understanding of the error properties of mixed-mode and multiple-frame surveys, but conclusive results are still lacking. The German data infrastructure initiative has the potential to contribute to this research. An overarching theme in all of the above mentioned developments has been the increased interest in the relationship between various error sources (Biemer et al. 2008). In Germany, there are several good opportunities to engage in research related to the intersection of error sources, especially given the exceptional data linkage efforts that have been undertaken. In this area, Germany is clearly taking the lead compared to the US. However, what could be improved in Germany is the collaboration between survey organizations and researchers, the amount of data shared between those organizations, and the willingness to systematically allow for randomized experiments in data collection protocols. In short, I would recommend the following:

- Work toward higher quality surveys, particularly in the face-to-face field. One step in this direction would be the development of survey methodology standards and the commitment to adhere to these standards. Those standards should include a minimum set of process indicators (metadata), and variables created in the data collection (paradata).
- Expanding options for graduate and post-graduate education in survey methodology could increase the potential for implementing high-quality surveys.
- Carefully examine interviewer hiring, payment, and training structures in German survey organizations. Recommendations or mini-

mum requirements regarding these issues might also be needed for German government surveys.

- Use the potential inherent in having multiple surveys run within (or across) the same survey organizations for coordinated survey methodology experiments. As we increase the burden on interviewers and try to reduce the burden on respondents, many questions will be left open in the research area of survey methodology, such as the effect of question context through matrix sampling, or the effect of interviewer shortcuts when creating sampling frames, or collecting paradata for nonresponse adjustment.

References:

Andridge, R. and Little, R. (2008): Proxy pattern-mixture analysis for survey non-response. Paper presented at the 2008 Joint Statistical Meeting. Denver.

Asef, D. and Riede, T. (2006): Contact times: what is their impact on the measurement of employment by use of a telephone survey? Paper presented at Q2006 European Conference on Quality in Survey Statistics. Cardiff.

Atrostic, B.K./Bates, N./Burt, G. and Silberstein, A. (2001): Nonresponse in US Government Household Surveys: Consistent Measures, Recent Trends, and New Insights. Journal of Official Statistics 17 (2), 209-226.

Axinn, W. et al. (2008): SRC Building New Software to Facilitate Mixed Mode Data Collection Projects. Center Survey 18 (6), 8.

Bates, N./Dahlhamer, J. and Singer, E. (2008): Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. Journal of Official Statistics 24 (4), 591-612.

Biemer, P./Beerten, R./Japec, L./Karr, A./Mulry, M./Reiter, J. and Tucker, C. (2008): Recent Research in Total Survey Error: A Synopsis of 2008 International Total Survey Error Workshop (ITSEW2008).

Campanelli, P./Sturgis, P. and Purdon, S. (1997): Can you hear me knocking: An investigation into the impact of interviewers on response rates. London.

Casas-Cordero, C. (2008): Errors in Observational Data Collected by Survey Interviewers. Comprehensive Examination Paper, Joint Program in Survey Methodology.

Conrad, F. and Schober, M. (2007): Envisioning the Survey Interview of the Future. New York.

Copas, A.J. and Farewell, V.T. (1998): Dealing with non-ignorable non-response by using an ‘enthusiasm-to-respond’ variable. Journal of the Royal Statistical Society Series A-Statistics In Society 161 (3), 385-396.

Couper, M. (2005): Technology Trends in Survey Data Collection. Social Science Computer Review 23 (4), 486-501.

Couper, M. and Lyberg, L. (2005): The use of paradata in survey research. Paper presented at the 54th Session of the International Statistical Institute. Sydney.

Couper, M./Raghunathan, T./Van Hoewyk, J. and Ziniel, S. (2008): An Application of Matrix Sampling and Multiple Imputation: The Decisions Survey. Paper presented at the 2008 Joint Statistical Meeting. Denver.

Curtin, R./Presser, S. and Singer, E. (2000): The effects of response rate changes on the index of consumer sentiment. Public Opinion Quarterly 64 (4), 413-428.

de Leeuw, E.D. and de Heer, W. (2002): Trends in household survey nonresponse: A longitudinal and international comparison. In: Groves, R.M./Dillman, D.A./Eltinge, J.L. and Little, R.J.A. (Eds.): Survey nonresponse. New York.

de Leeuw, E.D. (2005): To Mix or Not to Mix Data Collection Modes in Surveys. Journal of Official Statistics 21 (2), 233-255.

Dohrmann, S./Han, D. and Mohadjer, L. (2007): Improving coverage of residential address lists in multistage area samples. Proceedings of the Section on Survey Research Methods. American Statistical Association.

Durrant, G.B./Arrigo J. D. and Steele, F. (2008): Using Paradata to Develop Hierarchical Response Propensity Models 19th International Workshop on Household Survey Nonresponse. Ljubljana, Slovenia, 15-17 September 2008.
Eckman, S. (2008): Errors in Housing Unit Listing and their Effects on Survey Estimates, Comprehensive Examination Paper, Joint Program in Survey Methodology.
Fricker, S. (2007): The relationship between response propensity and data quality in the Current Population Survey and the American Time Use Survey. Doctoral dissertation. University of Maryland.
Gonzalez, J. and Eltinge, J. (2008): Adaptive Matrix Sampling for the Consumer Expenditure Interview Survey. Paper presented at the Joint Statistical Meeting, American Statistical Association.
Graham, J.D. (2006): Memorandum for the President's Management Council. January 20, 2006. Executive Office of the President, Office of Management and Budget. Washington, D.C. 20503.
Griffin, D. (2008): Measuring Mode Bias in the American Community Survey. Paper presented at the JSPM Design Seminar Spring.
Groves, R.M. (2006): Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly 70 (5), 646-675.
Groves, R.M. and Couper, M. (1998): Nonresponse in Household Interview Surveys. New York.
Groves, R.M. and Heeringa, S. (2006): Responsive design for household surveys: tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society Series A: Statistics in Society 169 (3), 439-457.
Groves, R.M. and McGonagle, K. (2001): A theory-guided interviewer training protocol regarding survey participation. Journal of Official Statistics 17 (2), 249-266.
Groves, R.M. and Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. Public Opinion Quarterly 72 (2), 167-189.
Groves, R.M./Wagner, J. and Peytcheva, E. (2007): Use of Interviewer Judgments About Attributes of Selected Respondents in Post-Survey Adjustment for Unit Nonresponse: An Illustration with the National Survey of Family Growth. American Statistical Association.
Groves, R.M./Fowler, J.F./Couper, M.P./Lepkowski, J.M./Tourangeau, R. and Singer, E. (2004): Survey Methodology. New York.
Groves, R.M./Brick, J.M./Couper, M./Kalsbeek, W./Harris-Kojetin, B./Kreuter, F./Pennell, B./Raghunathan, T./Schouten, B./Smith, T./Tourangeau, R./Bowers, A./Jans, M./Kennedy, C./Levenstein, R./Olson, K./Peytcheva, E./Ziniel, S. and Wagner, J. (2008): Issues Facing the Field: Alternative Practical Measures of Representativeness of Survey Respondent Pools. Survey Practice, October 2008 http://surveypractice.org/2008/10/30/issues-facing-the-field/ [Last visited: 03/02/2010].
Heerwegh, D./Abts, K. and Loosveldt, G. (2007): Minimizing survey refusal and noncontact rates: do our efforts pay off? Center for Survey Methodology, Katholieke Universiteit Leuven. Survey Research Methods 1 (1), 3-10.
Iannacchione, V.G./Staab, J.M. and Redden, D.T. (2003): Evaluating the use of residential mailing addresses in a metropolitan household survey. Public Opinion Quarterly 67 (2), 202-210.

Jäckle, A./Lynn, P./Nicolaas, G./Sinibaldi, J. and Taylor, R. (2008): Interviewer characteristics, their behaviours, and survey outcomes. Paper presented at the 2008 International Workshop on Household Nonresponse.

Jans, M. (2008): Can Speech Cues and Voice Qualities Predict Item Nonresponse and Inaccuracy in Answers to Income Questions? Prospectus for Dissertation Research, Michigan Program in Survey Methodology.

Keeter, S./Miller, C./Kohut, A./Groves, R. and Presser, S. (2000): Consequences of Reducing Nonresponse in a National Telephone Survey. Public Opinion Quarterly 64 (2), 125-148.

Kenny-McCulloch, S. (2008): Pretesting Surveys with Traditional and Emerging Statistical Methods: Understanding their Differential Contributions and Affects on Question Performance in Standardized and Adaptive Measurement. Prospectus for Dissertation Research, Joint Program in Survey Methodology.

Kreuter, F. (2010): Paradata. [In this publication].

Kreuter, F./Lemay, M. and Casas-Cordero, C. (2007): Using proxy measures of survey outcomes in post-survey adjustments: Examples from the European Social Survey (ESS). Proceedings of the Survey Research Methods Section. American Statistical Association.

Krosnick, J. (2005): Effects of survey data collection mode on response quality: Implications for mixing modes in cross-national studies. Conference on Mixed Mode Methods in Comparative Social Surveys. City University London, 15. Sept. 2005.

Little, R.J.A. and Rubin, D.B. (2002): Statistical Analysis with Missing Data, 2nd edition. New York.

Lynn, P. (2003): PEDAKSI: Methodology for collecting data about survey non-respondents. Quality and Quantity 37 (3), 239-261.

Maynard, D.W./Houtkoop-Steenstra, H./Schaeffer, N.C. and van der Zouwen, J. (2002): Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview. New York.

Mockovak, B. (2008): Mixed-Mode Survey Design. What are the effects on data quality? Challenging Research Issues in Statistics and Survey Methodology at the BLS. U.S. Bureau of Labor Statistics. www.bls.gov/osmr/challening_issues/mixedmode.htm.

Olson, K. (2007): An investigation of the nonresponse – measurement error nexus. Doctoral dissertation. University of Michigan.

Office of Management and Budget (2006): Questions and answers when designing surveys for information collections. www.whitehouse.gov/omb/inforeg/pmc_survey_guidance_2006.pdf. [Last visited: 03/02/2010].

O'Muircheartaigh, C./Eckman, S. and Weiss, C. (2003): Traditional and enhanced listing for probability sampling. Proceedings of the Section on Survey Research Methods. American Statistical Association.

O'Muircheartaigh, C./English, E./Eckman, S./Upchurch, H./Garcia, E. and Lepkowski, J. (2006): Validating a sampling revolution: Benchmarking address lists against traditional listing. Proceedings of the Section on Survey Research Methods American Statistical Association.

O'Muircheartaigh, C./English, E. and Eckman, S. (2007): Predicting the relative quality of alternative sampling frames. Proceedings of the Section on Survey Research Methods. American Statistical Association.

O'Muircheartaigh, C. and Heeringa, S. (2008): Sampling Designs for Cross-cultural and Cross National Studies. Paper presented at Multinational, Multiregional, and Multicultural Contexts. Berlin from June 25–28, 2008.

Peytchev, A., and Olson, K. (2007): Using interviewer observations to improve non-response adjustments: NES 2004. Proceedings of the Survey Research Methods Section. American Statistical Association.

Schnell, R. (1997): Nonresponse in Bevölkerungsumfragen. Opladen.

Schnell, R. (2007): Alternative Verfahren zur Stichprobengewinnung für ein IAB-Haushaltspanel mit Schwerpunkt im Niedrigeinkommens- und Transferleistungsbezug. In: Promberger, M. (Ed.): Neue Daten für die Sozialstaatsforschung: Zur Konzeption der IAB-Panelerhebung „Arbeitsmarkt und Soziale Sicherung". Nuremberg. IAB Forschungsbericht No. 5/2007.

Schnell, R. and Kreuter, F. (2000): Untersuchungen zur Ursache unterschiedlicher Ergebnisse sehr ähnlicher Viktimisierungs-surveys. Kölner Zeitschrift für Soziologie und Sozialpsychologie 52 (1), 96-117.

Schnell, R. and Kreuter, F. (2005): Separating Interviewer and Sampling-Point Effects. Journal of Official Statistics 21 (3), 389-410.

Schouten, B. and Cobben, F. (2007): R-Indexes for the Comparison of Different Fieldwork Strategies and Data Collection Modes. Paper presented at the International Workshop for Household Survey Nonresponse.

Spar, E. (2008): Federal Statistics in the FY 2009 Budget. In: American Association for the Advancement of Science, Report on Research and Development in the Fiscal Year 2009. http://www.aaas.org/spp/rd/rd09main.htm. [Last visited: 03/02/2010].

Staab, J.M. and Iannacchione, V.G. (2003): Evaluating the Use of Residential Mailing Addresses in a National Household Survey, Joint Statistical Meetings – Section on Survey Research Methods.

Voogt, R.J.J. and Saris, W.E. (2005): Mixed Mode Designs: Finding the Balance Between. Nonresponse Bias and Mode Effects. Journal of Official Statistics, 21 (3), 367-387.

Wagner, J. (2008): Adaptive Survey Design to Reduce Nonresponse Bias. Doctoral dissertation. University of Michigan.

Yan, T. and Raghunathan, T. (2007): Using Proxy Measures of the Survey Variables in Post-Survey Adjustments in a Transportation Survey, Joint Statistical Meetings – Section on Survey Research Methods.

Yan, T./Rasinski, K./O'Muircheartaigh, C./Kelly, J./Cagney, P./Jessoe, R. and Euler, G. (2008): The Dual Tasks of Interviewers, International Total Survey Error Workshop (ITSEW), North Carolina.

2. Improving Research Governance through Use of the Total Survey Error Framework

Marek Fuchs

Contact:

Marek Fuchs
Darmstadt University of Technology
Institute of Sociology
Residenzschloss
Marktplatz 15
64283 Darmstadt
Germany
e-mail: fuchs[at]ifs.tu-darmstadt.de

Abstract

Survey research is an integral element of modern social science. The German survey research infrastructure – in terms of research institutes, surveys, conferences, and journals – has greatly improved over the past 20 years, and recently several important European initiatives in this area have gained momentum. This has brought about the need for an integrated theoretical concept to assess and evaluate the quality of surveys and survey estimates. In our view, survey methodology is an interdisciplinary body of knowledge and expertise that describes the "science of conducting and evaluating surveys." It is a theory-driven empirical approach used to assess the quality of survey research. Thus, it applies the principles of survey research and experimental research to the development and assessment of survey methodologies themselves. Even though surveys have been conducted in a highly professional manner for decades, survey methodology offers the opportunity to use a universal theoretical approach when planning and assessing surveys as well as shared terminology. The integrated theoretical concept and joint terminology both foster the professionalization of survey methods and stimulate methodological research on the improvement of survey methods.

One key element of survey methodology is the total survey error framework. This will be described in greater detail below (section 1). Then we will discuss some limitations of this concept (section 2) and mechanisms and organizational issues that arise in promoting the use of this concept (section 3).

1. The total survey error framework

Multiple criteria are used to assess the quality of survey statistics; these include reporting timeliness, the relevance of the findings, the credibility of researchers and results, and finally, the accuracy and precision of the estimates. While timeliness of reporting and the credibility of researchers and results are rather soft indicators that require qualitative assessments, the accuracy of a survey statistic is an objective quantitative quality indicator. It is determined by the survey estimate's distance or deviation from the true population parameter. If, for example, a survey aimed to determine the average household income of a certain population, any deviation of the sample estimate from the true value – the one that would have been obtained if all members of the target population had provided error-free income data – would decrease the survey's accuracy. By contrast, the precision of a survey estimate is determined by the size of the margin of error (or confidence interval) and thus by the standard error. The standard error is a function of the sample size, of the alpha error, and of the variance of the measure in question. Accuracy and precision offer an integrated view of the quality of a survey estimate. While the precision is discussed in almost every introductory statistics textbook, the accuracy is not always considered to the same extent when evaluating the quality of a survey estimate. Rather, most survey researchers generally determine the margin of error or the standard error in order to assess the quality of an estimate. The accuracy of the estimate is considered less rigorous and is also less often determined explicitly. In survey methodology, accuracy and precision are treated as concepts of equal importance. However, given the lack of attention devoted to the accuracy of estimates so far, we focus on this facet in the present paper. In the following, we use the total survey error framework (e.g., Biemer and Lyberg 2003) to provide a comprehensive discussion of a survey estimate's accuracy.

There are two types of survey error that harm the accuracy of a survey estimate: variable or random error, and systematic error. While random errors are assumed to cancel each other out – that is, a negative deviation of the measurement from the true value would be compensated by a positive deviation – systematic errors shift the sample estimate systematically away from the true value. The latter would be the case, for example, if with a certain question wording, all respondents to a survey reported a higher number of doctor visits than actually occurred during a given reference period. For linear estimates (such as means, percentages, and population totals), it is safe to state that an increase in the random error leads to an increased variance, while a rise in any systematic error results in a larger bias of the estimate. Using this terminology, one can state that the accuracy of a survey estimate is affected by an increase in the bias.

From a traditional point of view, the driving factors or sources of survey error fall into two groups: sampling error and non-sampling error. Non-sampling error would then be further differentiated into coverage error, non-response error, and measurement error. A theory-driven modern approach distinguishes between observational errors and non-observational errors. While observational errors are related to the measurement of a particular variable for a particular sample unit, non-observational errors occur when an incomplete net sample is created that is supposed to represent the target population. Building upon this, Groves and colleagues (2009) classify sources of error into two groups: the first sources of error result from the representation of the target population in the weighted net sample ("representation"), and the second from effects on the survey responses obtained from a respondent ("measurement"). This extension of the traditional total survey error concept allows for detailed analysis of the mechanisms, and considers several sources of error as well as possible interaction effects.

1.1 *Total survey error components affecting representation*

(1) Before a sample can be drawn, a sampling frame is necessary that allows access to the members of the target population. The completeness of this frame and possible biases in its composition cause misrepresentations of the population by the sample. If a group is underrepresented in the frame – for example, if individuals who own mobile phones as their only communication device are missing from traditional random digit dialing (RDD) sampling frames because they do not have a landline telephone – the socio-demographic or substantive characteristics of this group are not considered when computing the survey statistic. This underrepresentation of some groups (coverage bias) causes a lack of accuracy of survey estimates (e.g., Blumberg and Luke 2007).

(2) Once a frame is available, one needs to draw a random sample, using a simple random sample, a stratified sample, a cluster sample, or more complex sample designs (Kish 1965; Lohr 1999). Based on this sample, the standard error is computed by taking the square root of the quotient of the variance in the sample and the number of cases in the sample. The standard error is then used to compute the confidence limits and the margin of error – both are indicators for the precision of the estimate. The sampling error depends heavily on the design of the sample: for a fixed number of sample cases, the standard error usually decreases if stratification is applied. By contrast, a clustered sample is generally characterized by larger design effects, which in turn raises the sampling error for a particular estimate. However, on a fixed budget, clustering usually increases the precision since the effective sample size can be increased even though the variance estimate suffers from the design effect caused by clustering.

(3) Unit non-response is probably the form of error that has been studied best of all the bias components in the total survey error framework (Groves and Couper 1998; Groves et al. 2002). Since the early days of survey methodology, researchers have been aware of the fact that some portions of the gross sample cannot be reached in the field phase of a survey or are not willing to comply with the survey request for cooperation. Since the responses of these groups may differ considerably from the responses of those members of the gross sample who can be reached and who are willing to cooperate, unit non-response is considered a serious source of systematic error that yields a non-response bias. The literature provides comprehensive theoretical approaches to explain the various stages of respondent cooperation and also findings that can be generalized beyond particular surveys. In part, this is due to the fact that a potential non-response bias can be assessed for variables for which parameters are

available from official statistics. Compared to other sources of error, this leaves survey researchers in a comfortable situation, since a possible bias can be observed more easily.

(4) Finally, the net sample needs to be adjusted for design effects introduced by the sample design. If the sample design, for example, asked for a disproportional stratified sample, an appropriate weighting procedure would have to compensate for the unequal selection probabilities when estimating the population parameter. In addition, the net sample may be adjusted for a possible non-response bias (redressment), although this procedure is questionable (Schnell 1997). Both procedures require complex computations considering information from the gross sample and from official statistics. While the first approach may potentially increase the random error of the estimate, correcting for bias may introduce systematic errors into the sample and thus bias the estimate.

1.2 Total survey error components affecting measurement

The four sources of error discussed so far are related to the representation of the target population by the weighted net sample. Coverage error, sampling error, non-response error, and adjustment error all potentially contribute to the random error or systematic error of the survey estimate. The next three sources of error are concerned with the measurement process. First, we will discuss the specification error, then the measurement error, and finally the processing error.

(5) Most concepts of interest in survey research cannot be observed directly. Measurement requires researchers to operationalize and translate the concept into questionnaire items that can be asked by interviewers and answered by respondents. For example, the general public's attitudes on illegal immigration need to be decomposed into several items describing various aspect and dimensions of illegal immigration. Respondents are then asked to report their degree of agreement with these items. The combined score of all items on this subject would then be treated as a measurement of the attitudes on illegal immigration. If an important aspect of this concept were missing on the scale, the validity of the operationalization would be compromised because the scale would not measure the defined concept completely and a specification error would occur. Usually, this results in a serious bias because the estimates based on an incomplete scale would not mirror the complete true attitudes of the members of the target population on illegal immigration. Unfortunately, the specification error is hard to determine: it requires a qualitative assessment, and standard procedures are rarely available to date.

(6) Measurement error is a rather complex component of total survey error (Lyberg et al. 1997). It consists of various elements that may cause systematic survey error as well as random survey error, both individually and jointly. Accordingly, measurement error may contribute to an increase in the estimate's variance as well as to its bias. Measurement error arises from the mode of survey administration, from the questionnaire or survey instrument, from the setting in which the instrument is administered, from the interviewers (if present), and also from the respondents (Lyberg et al. 1997).

Survey mode: the traditional trichotomy differentiates among face-to-face surveys, telephone surveys, and self-administered surveys. These modes differ with respect to the presence or absence of an interviewer – this allows for various degrees of standardization of the measurement process and also for different types of motivational support, as well as explanation and help to the respondent – and the dominant communicative channel (audio-visual, audio-only, visual-only). In recent years, many new survey modes have evolved with the introduction of modern information and communication technologies. Some of these modes transfer an established methodology into a computer-assisted mode (Couper et al. 1998); whereas other new modes have evolved as a consequence of merging survey modes (Conrad and Schober 2008). Each of these survey modes has its particular strengths and weaknesses for specific survey topics and survey designs. While a web-based survey might increase the variance of an estimate because respondents tend to answer a frequency question more superficially than in a face-to-face interview, the response to a face-to-face version of the very same questions might be prone to a higher degree of social desirability distortion, which in turn contributes to measurement bias.

Questionnaires: over the past 25 years, questionnaire design has evolved from an "art of asking questions" into a "science of asking questions" (Schaeffer and Presser 2003). This line of research has demonstrated on innumerable occasions that slight modifications in the wording of a question or response categories, in the order of the questions and response categories, and also in the visual design of the whole questionnaire as well as of single questions, affect the answers obtained from the respondents. Since the early days of the CASM movement (CASM=Cognitive Aspect of Survey Measurement), a multiplicity of research papers and textbooks (Sudman et al. 1996; Tourangeau et al. 2000) have contributed to a coherent theoretical approach that helps explain and predict random measurement error and systematic measurement error related to the questionnaire.

Respondent: also within the framework of the CASM movement, a detailed theoretical approach on how respondents consider and answer survey questions has been developed. As a result, the question-answer process has been described in great detail. Using this framework, several systematic and random respondent errors when answering survey questions have been identified. For example, satisficing behavior – as opposed to optimizing response behavior (Krosnick and Alwin 1987) – as well as mood effects and a "need for cognition" have been demonstrated by methodological research.

Interviewer: finally, it has been demonstrated that personal and social characteristics of interviewers – if present in the interview situation – as well as their task-related and non-task-related behaviors may have a considerable influence on the answers obtained from respondents. Accordingly, not only study-specific instructions are needed, but also improved professional interviewer training that focuses on general aspects of the interviewers' duties and responsibilities. However, one has to be aware that it is impossible to avoid individual respondent reactions to an interviewer's personal and/or social characteristics, since interviewer-administered surveys require a personal meeting of respondents and interviewers.

(7) Processing and editing the responses: in addition to the error components mentioned so far, the errors that occur when editing the survey responses obtained from respondents have been included in the total survey error framework. A few examples of possible error in the editing stage of a survey include poor handwriting with open questions, the treatment of inconsistent responses and of answers that were initially not codable, as well as incorrect classification of occupations. Also, scanning paper questionnaires with optical character recognition (OCR) technology and keying the answers from questionnaires into a database are prone to errors. In addition, some crucial responses may be imputed in the presence of item non-response, which is also susceptible to random or systematic error. Accordingly, these survey steps and the errors associated with them may either increase the variance of a variable – which in turn inflates the standard error and the margin of error – or compromise the accuracy of a response because a bias is introduced.

1.3 A simplified formula for the mean squared error

Technically speaking, the total survey error is the difference between a sample estimate and the respective parameter in the target population. This difference is measured by the mean squared error (MSE), which in turn consists of two components: the squared sum of the bias components plus the sum of the variance components (Biemer and Lyberg 2003, for an intuitive discussion of this concept). For the mean squared error, we need to combine the bias and variance from all sources in order to obtain an estimate of the total survey error. Although most sources of error can contribute to bias and variance simultaneously, some sources are primarily responsible for the increase of either variance or bias. Thus, a simplified formula for the mean squared error is as follows:

$$MSE = (B_{spec} + B_{meas} + B_{proc} + B_{cov} + B_{nr})^2 + VAR_{meas} + VAR_{samp} + VAR_{adj}$$

where the abbreviations have the following meaning:

B_{spec}	Specification bias/reduced validity
B_{meas}	Measurement bias
B_{proc}	Processing bias
B_{cov}	Coverage bias
B_{nr}	Non-response bias
VAR_{meas}	Measurement variance
VAR_{samp}	Sampling variance
VAR_{adj}	Adjustment variance

Although it is easy to estimate sampling variance – every introductory statistics textbook outlines the basic approaches – estimating the other types of variance and especially the biases is much more ambitious. The mean squared error as a measure for the total survey error is often only of heuristic value because the exact value of a particular variance or bias component cannot be computed.

The mean squared error offers the opportunity to evaluate survey designs and the estimates computed based on these survey designs. Thus, the "users" of a particular survey can assess the quality of reported results not only based on sampling error and the margin of error, but also based on other error components. This is especially important since the bias component of the mean squared error is assumed to exceed the sampling error. Thus, the sample estimate of the population parameter departs potentially more pronouncedly from the true value than has been assumed based on the sampling error alone.

1.4 Some pros and cons of the total survey error framework

Although total survey error offers a convincing framework to evaluate the accuracy of a survey estimate, it also suffers from a serious drawback. The effort necessary to compute a reasonable estimate of the magnitude of a particular error component usually exceeds the available resources. The estimation of the mean square error requires multiple repetitions of the survey design, which is usually too costly and also not feasible since the target population does not remain unchanged between repetitions. Also, for many survey designs, some error components are not accessible because of the field procedures applied or legal constraints (e.g., privacy laws prohibit extensive non-response follow-up studies in many countries). Also, it should be noted that for the exact computation of the mean squared error the parameter needs to be accessible. Because this is usually not the case, the mean squared error is seldom explicitly determined in practice. More often, only a few key components are estimated or a survey design is rated along the various components of bias and variance on a scale from "low" to "high." The decision for a particular survey design is then made based on a detailed computation of some error components and a rough assessment of the magnitude of the other error components. This leaves the researcher as well as the user of a survey statistic in a situation where a qualitative assessment of the magnitude of the total survey error is the best available assessment.

Regardless of this serious limitation of the total survey error framework, survey research and survey methodology have greatly benefited from the emerging total survey error approach.

(1) The total survey error framework makes researchers aware of possible errors in their survey statistics. If the response rate and the size of the net sample are the only available indicators for a given survey, many likely biases remain undetermined. Here, the total survey error framework allows for systematic reflection on possible limitations to survey quality and thereby fosters professional evaluation of ongoing surveys in terms of data quality and provides a common language and terminology for critical discussion.

(2) In addition, the total survey error framework provides a theoretical explanation for the various types of possible errors (variance and bias) and also for the underlying mechanisms (random error vs. systematic error). It also names a wide range of possible sources for problems in data quality. Hence the total survey error framework puts forward a more comprehensive theoretical approach to further developments of survey methods, beyond the traditional "keep at it" approach. In addition, it provides measurable indicators for evaluating the improvements introduced by these new survey methods.

(3) The total survey error framework also provides a basis for interdisciplinary discourse across the boundaries of traditional disciplines. Among others, surveys have been used for a long time in the fields of sociology, psychology, economy, and educational research. Although it is too early to say that the specific methodologies of these various fields have been completely integrated, one can say that these various methodologies have merged to some extent or are in the process of integration based on the total survey error framework and the survey methodology.

(4) From an international perspective, the integrated concept of total survey error has contributed to the dissemination of high criteria and a set of methods to meet those criteria. International surveys like the Programme for International Student Assessment (PISA), the International Social Survey Programme (ISSP), and the European Social Survey (ESS) would

not be feasible if researchers from diverse cultural and disciplinary backgrounds were not able to interact and cooperate in a common framework. Although there are still many national differences in the design and administration of surveys, the total survey error framework does promote a minimum degree of conformity in the assessment of data quality.

2. Survey error and survey cost

The survey designer's goal is to reduce the total survey error through proper design decisions in the preparatory stages for the survey as well as during the fieldwork. Most of the time, however, design decisions – regarding the mode of administration, question format, interviewer training procedures, and so on – do not only affect one specific source of error, but rather multiple sources. Thus, every improvement in eliminating one error source may be accompanied by an increase in another. Hence, survey designers need to compromise and balance different sources of error.

The total survey error framework offers the opportunity to determine the relative importance and weight of various error components in a given survey. Although not every component can be determined for each survey, an evidence-based assessment of multiple error sources is possible. As the body of literature on the various error components expands, researchers will be able to choose cost-efficient strategies that help reduce the total survey error (Groves 1989). However, in practice, survey designs are not only evaluated in the presence of fixed constraints on time and money. For example, survey design A may be chosen over survey design B despite the fact that it produces data of lower quality in terms of the mean squared error. But because the estimated cost of survey design B is considerably higher, the person responsible nevertheless decides to use survey design A.

Thus, the total survey error framework also relates to cost and requires survey designers to consider the accuracy of their surveys in relation to cost and timeliness of reporting. This raises the danger that researchers will sacrifice the quality of their survey to cost. However, since the total survey error approach requires researchers to document and publish key characteristics of each survey, the scientific community can easily assess to what extent survey quality is compromised to cost constraints. It is hoped that this will prevent researchers from making design decisions solely or predominantly based on the costs involved. The acceptance of the total survey error framework would be greatly increased if funding agencies required applicants to make use of this approach in their proposals.

3. Organizational issues

Although the total survey error approach offers a set of standardized terms, concepts, and measures, it needs to be adapted to the respective surveys, topics, and country-specific conditions. In addition, the total survey error framework involves evidence-based discussions of methodological issues conducive to producing advice and rules based on empirical tests and evaluations. Thus, a thorough country-specific assessment of the various components of the total survey error is needed, either in the form of an evaluation of ongoing surveys in the field or of lab- or field-based independent experimental studies. While experimental studies on methodological issues provide basic knowledge and allow for the testing of methodological concepts, they lack applied results that could directly benefit ongoing surveys. By contrast, in an evaluation embedded in an ongoing survey, researchers are limited in the degree to which the experimental methodological design is able to test innovative approaches since highly risky designs might harm the quality of the production data. Based on this reasoning, improvements in the total survey error approach should be promoted through a strategy combining methodological evaluations of ongoing large-scale surveys and of stand-alone experimental studies or laboratory experiments.

Until recently, the resources allocated to methodological research have not been adequate. Methodological studies have been conducted only as addendums to substantive surveys – which limits the scope and design of the study – or with student populations, or with other factors limiting their generalizability. Of course, the former ZUMA[1] (now continued as a department of GESIS[2]) has a long tradition in methodological research. Nevertheless, given the lack of resources, studies conducted elsewhere have usually been either focused on specific surveys or – if conducted independently – limited in their size and thus in the broader impact of the results in the scientific community. In the past few years, however, two important developments have been taken place. On the one hand, several large-scale surveys have taken over their own survey operation in order to evaluate new modes, innovative instruments, and means of reducing non-response. On the other hand, the projects funded by the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*) since 2008 in the Priority Program 1292 "Survey Methodology" have shown potential to function as a nucleus for a broader movement towards basic methodological research. Based on these experiences it seems advisable to promote a twofold strategy: methodological research should be implemented as part of every large-scale survey funded by public resources. A research plan for methodological studies should

1 Zentrum für Umfragen, Methoden und Analysen.
2 Leibniz Institute for the Social Sciences (*Leibniz-Institut für Sozialwissenschaften*).

already be developed in the design stage of the respective surveys and should be covered by a certain percentage of the overall funding (e.g., 5 percent of the total funds allocated to a particular survey). The research plan for the methodological study should already be specified in the proposal for the survey and evaluated by survey methodology experts according the same high standards as the proposal for the substantive study.

Unlike in the US and some other countries, German academic researchers do not have a wide range of field organizations at their disposal. Although several universities have built small to medium-sized computer-assisted telephone interview (CATI) facilities and some medium-sized online access panels are available as well, the majority of the fieldwork is conducted by private market research institutes. In order to promote the total survey error framework, a universal application of this concept is needed across all sectors including academia, official statistics, and the private sector. At present, the General Online Research conferences (with respect to web surveys), the meetings of the Section on Quantitative Research Methods in the German Sociological Association (DGS, *Deutsche Gesellschaft für Soziologie*) and a few other small-scale events are the only settings in which researchers from academia, the private sector, and official statistics come together and engage in joint methodological discussion. This is completely unsatisfactory. The annual conference of the American Association for Public Opinion Research might serve as a model for a similar conference scheme in Germany.

So far, high ranking permanent academic positions in the field of survey research are usually filled with experts in substantive research areas who are also qualified as survey experts and, in particular, as survey statisticians. Thus, for junior researchers, it is hard to build a career predominantly on survey methodology or even on a specialization in this field (e.g., sampling, measurement, or non-response). However, professionalizing the field of survey methodology will require an infrastructure of experts who focus on the various components of total survey error. Thus, in addition to survey experts in substantive fields and survey statisticians, experts in data collection and survey methodology should be considered more often for permanent academic positions. In the past few years, a few positions have been deliberately offered to this group. Further action should be taken to provide survey methodology with a sufficient human resource basis.

In order to disseminate survey methodology and the total survey error framework, a few specialized master study programs are beginning to emerge in Germany. Given the longstanding tradition of such programs in the US (e.g., Ann Arbor and the Joint Program in Survey Methodology, or JPSM) and the UK (e.g., Essex), one could expect positive effects in Germany as well. Also, doctoral education in the field of methods research has been offered so far on an individual basis only. Accordingly, a structured doctoral program that offers a set of integrated courses in survey methodology needs to be established.

A key challenge for the development of high-quality survey research lies in adopting joint quality indicators and common standards for each of the qualitative measures. The rather disparate use of response rates and measures of non-response in Germany is a good example of how survey research could benefit from an integrated quality concept. Whether we should adopt the English terminology or develop German terminology for the same concepts also needs further discussion. In our view, the use of the English terms has the advantage that the words are clearly identifiable as technical terms. In addition, the use of a shared English language terminology facilitates collaboration in international surveys such as the ESS or EU-SILC. Finally, when using the international terminology it is easier to participate in international discussion at conferences and in journals.

4. Summary of recommendations

In sum, this paper does not suggest a completely new approach to the methodological research on survey methods. Instead, it proposes that the existing work be integrated into the total survey error framework and that this concept and other knowledge from the field of survey methodology be applied rigorously to the planning and assessment of surveys. Also, it recommends the increased use of evidence-based rules and strategies to improve surveys. This will require evaluation and validation studies embedded in ongoing surveys as well as independent experimental studies in the field or in the lab that are not bound by the same limits as ongoing surveys. The following recommendations are the key elements of a strategy for achieving these objectives:

- The total survey error framework should be adopted as standard to describe and assess the quality of surveys. Since this concept requires the documentation of different variance and bias components associated with a particular survey, this will promote the methodological considerations in the planning phase of a survey, in its field phase, as well as during the analysis.
- The error components of a particular survey should be assessed based on evidence from evaluation studies or experimental work.
- Strategies und rules on how to improve the quality of surveys in general should be evidence-based. Experiments in the field and in the lab are key elements in support of evidence-based rules and strategies.
- Funding for methodological research in the total survey error framework should be provided in the context of ongoing large-scale surveys (a minimum of 5 percent of the overall budget for a particular survey) as

well as by national funding agencies for independent experimental studies in the field and in the lab.

- Accordingly, the total survey error framework should be mandated by funding agencies. Applicants should be required to make use of this approach in their proposals.
- For academic positions in the field of survey research (associate and full professorships), universities and similar research institutions should not only recruit candidates from substantive areas or from the field of survey statistics, but should also consider survey methodologists with a record of publications and projects in the various components of total survey error. This will help establish expertise in survey methodology and contribute to the professionalization of survey methodology.
- In order to maintain a consistent flow of graduates and postgraduates in the field of survey methodology, the emerging specialized master programs should be strengthened. Also, at least one structured doctoral program with an international teaching staff should be established in Germany.
- The further development of survey methodology in Germany should be fostered by launching a new international journal that offers survey methodologists a forum by publishing peer-reviewed papers on data collection in English.
- An annual conference of survey methodology experts from the academic sector, the private sector, and official statistical agencies should be established to promote and foster the use of the total survey error framework in survey research across these three sectors in Germany.

As these recommendations are gradually put into practice, survey methodology will evolve as a professional cross-disciplinary discipline contributing to survey research in economics, sociology, political science, health research, educational research, consumer and market research, and many other fields in the academic sector, the private sector, and also in the official statistical agencies.

References:

Biemer, P.P. and Lyberg, L.E. (2003): Introduction to Survey Quality. New York.

Blumberg, S.J. and Luke, J.V. (2007): Coverage bias in traditional telephone surveys of low-income and young adults. Public Opinion Quarterly 71 (5), 734-749.

Conrad, F.G. and Schober, M. (2008): Envisioning the survey interview of the future. Hoboken.

Couper, M.P./Baker, R.P./Bethlehem, J./Clark, C.Z.F./Martin, J./Nicholls, W.L., et al. (Eds.) (1998): Computer Assissted Survey Information Collection. New York.

Groves, R.M. (1989): Survey errors and survey costs. New York.

Groves, R.M. and Couper, M.P. (1998): Nonresponse in household interview surveys. New York.

Groves, R.M./Fowler, F.J./Couper, M.P./Lepkowski, J.M./Singer, E. and Tourangeau, R. (2009): Survey Methodology (2nd edition). New York.

Groves, R.M./Dillman, D.A./Eltinge, J.L. and Little, R.J.A. (Eds.) (2002): Survey Nonresponse. New York.

Kish, L. (1965): Survey Sampling. New York.

Krosnick, J.A. and Alwin, D.F. (1987): An evaluation of a cognitive theory on response-order effects in survey measurement. Public Opinion Quarterly 51 (2), 201.

Lessler, J.T. and Kalsbeek, W.D. (1992): Nonsampling Errors in Surveys. New York.

Lohr, S.L. (1999): Sampling: Design and Analysis. Pacific Grove.

Lyberg, L./Biemer, P./Collins, M./De Leeuw, E./Dippo, C./Schwarz, N. et al. (Eds.) (1997): Survey measurement and process quality. New York.

Schaeffer, N.C. and Presser, S. (2003): The science of asking questions. Annual Review of Sociology 29 (1), 89-113.

Schnell, R. (1997): Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen. Opladen.

Sudman, S./Bradburn, N. and Schwarz, N. (1996): Thinking about answers. The application of cognitive Processes to survey methodology. San Francisco.

Tourangeau, R./Rips, L. and Rasinski, K. (2000): The psychology of survey response. Cambridge.

3. Metadata

Arofan Gregory, Pascal Heus and Jostein Ryssevik

Contact:

Pascal Heus
Open Data Foundation
5335 North Nina Drive
Tucson, AZ 85704
US
e-mail: info[at]opendatafoundation.org

Abstract

Metadata, or data about data, play a crucial role in the social sciences, ensuring that the data collected are accompanied by thorough documentation and grounded in community knowledge across their entire life cycle – from the early stages of data production to secondary analysis by researchers or use by policy-makers and other key stakeholders. This chapter provides an overview of the social sciences metadata landscape, including best practices and related information technologies. It focuses in particular on two measures – the Data Documentation Initiative (DDI) and the Statistical Data and Metadata Exchange (SDMX) Standard – that appear central to a global metadata management framework for social data and official statistics. It also highlights current trends and challenges to integration and provides a set of high-level recommendations for producers, archives, researchers, and sponsors with the aim of fostering the adoption of metadata standards and best practices in the years to come.

Keywords: social sciences, metadata, data, statistics, documentation, data quality, XML, DDI, SDMX, archive, preservation, production, access, dissemination, analysis

1. What is metadata?

Metadata is a difficult term to define; it means many things to many different audiences. If we turn to Wikipedia, we find: "Metadata (meta data, or sometimes metainformation) is 'data about data,' of any sort in any media."[1] While broadly true, the Wikipedia definition does not capture the real importance of metadata to those involved in social science research.

Within any domain, the term *metadata* can be more usefully defined by describing its agreed use. In the case of social science research, there exists a well-developed metadata culture, which allows us to be very specific. Researchers understand what data are: the full range of information that is collected, processed, analyzed, and used in the conduct of research. *Metadata* covers all forms of documentation about this data.

Even so, we are left with a definition of the term that is still incredibly broad. It is sometimes helpful to think about the different types of metadata, using common terms:

- *Structural metadata* describes the structure of datasets, whether these are tabular in nature or simply files of raw data or microdata. Which variable's value appears in which column? Which row represents which case? Are there hierarchical relationships? Etcetera.

1 http://en.wikipedia.org/wiki/Metadata.

- *Reference metadata (also known as "descriptive" metadata)* consists of what is often thought of as "footnote" metadata, whether relating to methodology, sampling, quality measurements, production notes, or other aspects. This is a very broad term that can cover a range of information dealing with everything from individual values to entire collections of data.
- *Administrative metadata* is the data created through the process of administering data, covering its collection, production, publication, and archiving.
- *Behavioral metadata (also known as "paradata")* is information about the reaction and behavior of users when working with data, and that of respondents when the data is being collected (in this case, it is paradata about a collection instrument). This can be of interest to those who act as data librarians, enabling them to better manage their data collections, but can also be of direct interest to researchers seeking to address the questions: what did other researchers do with the data? How did respondents react when asked a question?

It is worth noting that metadata are for human as well as machine consumption. Whereas most of the structural metadata exist to allow software processes to read, manipulate, and exchange data files, the purpose of reference and behavioral metadata is to enable human researchers to find, understand, and assess the quality of the data.

One of the criticisms of metadata as a broad discipline is that it is context-dependent, especially in terms of its use to help navigate the contents of the Internet as a whole. Indeed, there is a long and ongoing debate about the value of metadata. This debate – while both entertaining and instructive – is not particularly useful to those in social sciences research, because very specific definitions of the relevant metadata exist in the form of standard metadata models: the Data Documentation Initiative[2] (DDI), ISO-TS 17369 Statistical Data and Metadata Exchange[3] (SDMX), Dublin Core Metadata Initiative[4] (DCMI), ISO/IEC 11179,[5] the Neuchatel models for variables and classifications, and others.

The benefit of having such standards is that they allow for *direct implementation* of metadata-driven systems and management systems for metadata – and thus realization of the benefits – without having to answer questions about the precise value and meaning of metadata in its broadest sense.

2 http://ddialliance.org/
3 http://sdmx.org/
4 http://dublincore.org/
5 http://metadata-standards.org/

2. Metadata and technology

2.1 Historical technological approaches

Metadata is a natural part of most current data implementations, given the strong focus modern technology places on information. If technology depends on the *exchange and use of information* – or data – then the metadata describing that information can be critical in the creation of systems that perform tasks in an automated way.

Many of the early discussions about metadata dealt with describing the structure of data, whether it be the simple textual format of a data file or the structural information about a relational database schema. Other discussions were concerned more with the content of the data – that is, the type of file and what it contains. This focus arose naturally out of computers' ability to compute at ever increasing speed: the first challenge was to handle the data itself and to perform some operation with it. Once this was achieved, the question was how to retain enough information about the data so that it could be *exchanged with others or used in the future.* This was where the interest in metadata arose.

It is interesting to note how little the metadata capabilities of many statistical tools have grown since the era before the Internet. While many other types of applications have developed the ability to process and understand files from other users based on standard formats and models, statistical processing applications do not share this rich, "networked" view of the world. Many statistical tools today are reminiscent of applications dating from the 1980s – they understand enough metadata to handle specific data files and to interpret their contents and format or perform analytical operations, but have little ability to exchange this information with other systems or describe the context in which the data was produced.

2.2 Metadata and the Internet

The single most important development driving the current interest in metadata is the advent of the Internet. A vast network of interconnected computers requires a large set of standard protocols to allow computers to use files throughout the network. Most of these protocols are metadata.

To give a simple example: when a browser on your computer encounters a webpage, it gets a set of information from the server – metadata – which it uses to properly display that page. The webpage will probably be in HTML[6], but it might also be a Word document or a PDF file, or even a video clip. Each of these files requires a different application behavior. Thus, part of the

6 http://en.wikipedia.org/wiki/HTML.

metadata given to the browser is the MIME-type[7] of the file, which tells the computer which application to launch.

Early Internet protocols provided enough metadata to allow for human users to exchange files, but there was typically insufficient metadata for computer applications to directly perform tasks without human intervention. Because the emphasis was on people viewing files from around the network, there emerged metadata standards that supported this type of application – the best-known of these were a set of citation fields for describing any kind of resource, the Dublin Core.

As the Internet has evolved, there has been an increasing emphasis on interactions between applications – a phenomenon termed "distributed computing." This development revealed that the available metadata – even with the help of standards such as the Dublin Core – were insufficient. In all of its applications, however, the Internet has placed strong emphasis on the use of remote resources without the need for explicit, human-guided integration, thus demanding a large amount of metadata and increasingly requiring metadata standards.

2.3 Metadata and XML-based technologies

One of the biggest developments in the growth of the Internet – and for distributed computing generally – was the advent of the eXtensible Markup Language[8] (XML) and the suite of related technologies and standards. Derived from a technology standard for marking up print documents – the Standard Generalized Markup Language[9] (SGML) – the original focus of XML was to better describe documents of all sorts so they could be used more effectively by applications discovering them on the Internet.

XML is a meta-language used to describe tag sets, effectively injecting additional information into a document. Unlike HTML (which was also based on SGML), however, there was no fixed list of tags – the whole point is that documents could be designed to carry specific additional information about their contents. Thus, XML document types could be designed to carry any sort of metadata in line with the contents of the document.

XML is not only a language but also a collection of technologies available to perform various operations on the underlying data or metadata: the XML schema for describing document structure; XPath[10] and XQuery[11] for

7 http://en.wikipedia.org/wiki/MIME.
8 http://en.wikipedia.org/wiki/XML and http://www.w3.org/XML/
9 http://en.wikipedia.org/wiki/SGML.
10 http://en.wikipedia.org/wiki/XPath.
11 http://en.wikipedia.org/wiki/XQuery.

querying and searching XML; SOAP[12] and REST[13] for facilitating the exchange of information; and many others.

Most importantly, the above technologies are often readily available on most computers and are free to use. The XML standards themselves are maintained by the World Wide Web Consortium[14] and publicly available. This implies that XML not only provides a common language and facilitates metadata management but is also easy to adopt as a technology. While XML does not preclude the existence of legacy metadata management systems, it has shifted the way we model the information structure and expose the metadata to the outside world. Harmonized models have emerged in various fields of expertise, including the social sciences.

The Dublin Core was quickly realized in an XML format, and other standards also used the new format, notably the DDI (see below). At first, these standards were designed very much with human users in mind, but those involved in solving problems related to distributed computing realized that XML was a very powerful tool as well.

These developments led to a set of Web services[15] standards (SOAP, WSDL, etc.) as well as a new type of service-oriented architecture[16] (SOA). The development of Web services technology and service-oriented architectures continued the demand within applications for precisely defined metadata exchanged using standard protocols. Some of the later standards such as SDMX – and later versions of existing standards (such as DDI version 3.0) are designed to leverage these developments.

Today, we have a powerful set of technology tools and metadata models that are directly relevant to the applications used by the social sciences researcher. While not all of the statistical software packages have utilized these developments, we are increasingly seeing these new metadata-rich technologies used to provide researchers and those who support them with functionalities that were not possible in earlier generations of technology.

12 http://en.wikipedia.org/wiki/SOAP.
13 http://en.wikipedia.org/wiki/Representational_State_Transfer.
14 http://www.w3.org/
15 http://en.wikipedia.org/wiki/Web_service.
16 http://en.wikipedia.org/wiki/Service-oriented_architecture.

3. Metadata and the social sciences

3.1 Why metadata?

In the social sciences, data quality has a direct impact on the soundness of policies and the validity of research outputs. Data quality is typically measured using criteria such as accessibility, coherence, relevance, timeliness, integrity, consistency, and coherence. These indicators are generally accepted as a good measure of the overall usefulness of the data. Meeting these criteria not only means making data available but also requires comprehensive documentation of the data structures, production processes, statistical methodologies, data sources, contexts, and many other aspects. This is necessary not only to ensure usability but also for purposes of discovery, accessibility, preservation, and information exchange.

In the social sciences, metadata is essential for several reasons:

- It is needed to ensure that users have sufficient information to properly *understand* and *use* the data. Without relevant documentation, researchers are unable to accurately interpret the meaning of the data. A lack of information also places an extra burden on data providers, who need to be able to respond to users' questions.
- It is required to facilitate *data discovery* and *access* by the intended consumers. The best data in the world is useless if no one is aware of its existence.
- It supports the long-term *preservation* of data by ensuring that the relevant information remains with the data for future use or for conversion into new archival formats.
- Common metadata languages and structures are also essential to support the *exchange* of information between agencies and/or individuals.

In general, better documentation makes for more useful data, and ultimately better research. The usability of data is intricately tied up with issues about how thoroughly it is documented: rich metadata about a dataset allows for easier access and use of the data. Researchers want better data, and one way to help improve data quality is to provide better documentation.

3.2 Metadata and the data lifecycle

The data lifecycle in the social sciences is quite complex as the data flowing from the survey respondents or administrative systems to the researchers and policy-makers goes through several stages and transformation processes involving many different actors. Furthermore, secondary or derivative data and research findings often themselves become data sources for others.

Any description of the purpose of metadata within the data lifecycle should start with an analysis of the users' requirements:

- The majority of data users are not involved in the creation of the data they use.
- Data are frequently used for other research purposes than intended by the creators (secondary analysis).
- Data are frequently used many years after they were created.
- Data users often compare and combine data from a broad range of sources (across time and space).

The common denominator of the four characteristics is their emphasis on the relative distance between the end users of statistical material and the production process. Whereas the creators and primary users of statistics might possess "undocumented" and informal knowledge that will guide them in the analysis process, secondary users must rely on the formal metadata accompanying the data to exploit their full potential. For this reason it might be said that social science data only become accessible through the metadata accompanying the dataset. Without written descriptions of the various elements comprising a dataset, it will appear to the end user as a more or less meaningless collection of numbers. The metadata provides the bridges between the producers of data and their users and convey information that is essential for secondary analysis.

Ideally, data providers should abide by Gary King's replication standard,[17] which holds that "sufficient information exists with which to understand, evaluate, and build upon a prior work if a third party can replicate the results without any additional information from the author." Note that from this perspective, researchers as much as producers are defined as "data providers," and should therefore abide by the same documentation principles.

Traditionally, however, metadata has not been the focus of data producing agencies and the responsibility for documenting data was often left to the data archive, data librarians, or Research Data Centers. Such "after-the-fact" efforts require substantial resources and typically lead to a considerable amount of information loss and sparsely documented data.

This mindset has changed in recent years, and considerable efforts are now being made by data producers and archives to improve the overall quality of metadata. The idea is also being extended to the researchers or end

17 "Replication, Replication," Gary King, PS: Political Science and Politics, Vol. XXVIII, No. 3 (September 1995), 443–499 and "A Revised Proposal, Proposal," Vol. XXVIII, No. 3 (September 1995), 443–499. See also http://gking.harvard.edu/projects/repl.shtml.

users, whose contribution to metadata is often nonexistent. Collecting inputs from the users themselves should lead to a better understanding of data usage, reduce the duplication of efforts, and promote the sharing of knowledge. This shift from a centralized maintenance of metadata by the archive to a distributed approach, where many entities contribute to the knowledge, seems only natural: it is better and easier to capture information about an event at the time of its occurrence than after the fact.

There is another view of the data lifecycle that is not so much concerned with the collection and production of data for research as it is with the aggregation and harmonization of data. This view can be termed the *information chain* because it describes the flow of data from its original micro-level source(s) through the various aggregation and harmonization processes, as the data flows upward from its source through the hierarchy of primary and secondary users. Data collected through surveys or from administrative sources at a regional level might be aggregated at a national level, combined with other sources, and then further aggregated at the international level.

This view of the data lifecycle also places importance on the distance between those collecting the original data, and its eventual use at a higher level of aggregation. Without sufficient documentation about the aggregation and harmonization processes, it is difficult for end users to fully understand the aggregates they are using.

The main goal of capturing metadata at each stage in the lifecycle is to maintain it throughout a single cycle from collection to publication (and hence to archiving), but also to capture each secondary use of the data, so that any dataset will be accompanied by as complete a set of documentation as possible. Information captured as it comes into existence is higher in quality and more complete, which directly benefits the user of the data.

There are also less obvious benefits to having a consistent set of metadata accompanying a dataset through the lifecycle: good metadata can be used to help drive the processing of the data as it goes through its lifecycle, and well-documented data collections make it possible to compare similar datasets. Complete information about the content and processing of a collection of data can provide valuable information to those who want to re-purpose or manage the data within that collection. Thus, the beneficiaries of good metadata, captured as the data is collected, processed, and published, include not only researchers but also secondary users, archivists, and data producers.

Very often, good metadata can form the basis for code generation, whether that code runs inside a statistical package or is used for some other purpose (such as automatic generation of forms for data collection). It can also be used for the automated production of documentation or publications that can be customized to the end user's needs. Although not immediately apparent, the benefit of having good metadata is that the systems which support the researcher, data producer, and archivist can all be made much more efficient and produce higher-quality data.

3.3 *Standard metadata models*

The recent emphasis on the data lifecycle, and on capturing metadata from the beginning, has driven the development of two standard models, each designed around one of the data lifecycle views described above. The DDI is, in its most recent version, based on a lifecycle model that describes the collection and sourcing of data through the stages of publication, archiving, and secondary use. ISO TS-17369, the SDMX standard is based on a view of the information chain, with a stronger focus on aggregate data products. These standards – along with a number of others in various important areas – create a common view of how metadata within the social sciences domain can be described and exchanged to facilitate the flow of metadata accompanying the relevant datasets.

4. The Data Documentation Initiative

4.1 *DDI – early history*

The DDI[18] is an international program to produce a metadata specification for the description of social science data resources. The program was initiated in 1994 by the Inter-University Consortium for Political and Social Research (ICPSR). Contributors to the project come mostly from social science data archives and libraries in the US, Canada, and Europe.

The original aim of the DDI was to replace the widely used OSIRIS codebook specification with a more modern and Web-aware specification that could be used to structure the description of the content of social science data archives. The first preliminary version came in the form of an SGML Document Type Definition[19] (DTD), which in 1997 was converted to an XML DTD. The migration to XML took place just a few months after the W3C released the first working draft of the XML specification. The DDI was consequently one of the very first major metadata initiatives using the new framework. Several data archives started to use the DDI to describe their data collections, and software was developed to support its use. However, it soon became apparent that the first versions of the DDI had several severe limitations:

---

18 http://www.ddialliance.org.

19 http://en.wikipedia.org/wiki/Document_Type_Definition.

4.1.1 A pure "bottom-up" approach

The DDI specification was developed to describe concrete files or products coming out of the statistical production process. Given its roots in social science data archiving, this is quite natural. The information objects in the data archives were final products whose lifeline to the various production processes had been severed and which were given individually to users, outside their original production context.

As a consequence there was a one-to-one relationship between a DDI instance and the physical data it was meant to describe. The DDI was tied to the dataset, and there were no methods to describe abstract statistical concepts that might be represented in more than one concrete study. It was therefore impossible to reference identical variables across datasets, and even series of survey instances where the majority of variables are identical from wave to wave had to be described instance by instance.

4.1.2 Modularity

The first versions of DDI had their roots in a "book" metaphor. It was seen as the digital equivalent of a paper document – the well-established codebook or data dictionary. The specification was not built according to a modular architecture that would have allowed information and application providers to select bits and pieces and "snap" them together on a freer basis.

4.1.3 Extensibility

Another critical limitation was the lack of a proper extensibility mechanism. Within the confines of an XML DTD there is no way to add local extensions without compromising the interoperability of the core specification. You either accept the specification as it is without any additions or you break it. For a big and complex specification like the DDI, this is a major problem that can easily damage the adoption process. Without a mechanism that allows extensions to be made without breaking the standard, the chances are high that application providers will sacrifice interoperability for local efficiency and relevance.

Despite these limitations, the DDI met the fundamental needs of data archives for documenting survey datasets and has been widely adopted by agencies around the world.

4.2 DDI version 3.0

Version 3.0 of the DDI was released in April of 2008, representing a major revision to the standard that solved the problems of earlier versions as described above. Based on a survey lifecycle model, it is designed to describe groups and series of studies, to define degrees of comparison within and across studies, and to allow for reuse of metadata where appropriate. It uses a modular approach, with modules which are related to each step of the data lifecycle. Different types of metadata are organized into packages relating to their contents. All the metadata about a survey instrument, for example, are found in the "data collection" module, represented by an XML namespace.

DDI 3.0 represents an approach to the metadata that is more in line with the capabilities of modern information technology: it is relational in nature rather than document-centric so that metadata can be easily referenced and reused. This is important because modern Web services technology utilizes the idea of distributed computing. DDI 3.0 is designed explicitly to support the concept of having a collection of metadata be distributed and reused by reference.

The combination of the lifecycle approach, a modular design, and metadata reusability has transformed the specification from a product intended for archiving datasets by a single agency into a highly flexible standard that can be used by all actors in the survey lifecycle for different purposes. Expected uses of DDI 3.0 include study design and survey instrumentation, questionnaire generation, support for data collection and processing operations, capturing data aggregation or recoding, managing question or concept banks, data discovery, research projects, data comparability, metadata mining, and probably a number of other purposes that cannot yet be foreseen. For each case, a subset of the specification is used either for the specific purpose or to provide a customized view of the information. A strength of DDI 3 is that it maintains a common language and metadata consistency across the lifecycle stages and among contributors.

The new version has also been designed to work with standards such as SDMX, ISO 11179, Dublin Core, and others, which ensure that the metadata can be connected to other domains or stages of the lifecycle. It takes into account backward compatibility with previous versions of DDI to ensure that current users can continue to use their existing framework or metadata.

Overall, DDI 3.0 has broadened the scope of the specification and made the standard attractive to a broader range of users across the entire survey lifecycle, from data producers to researchers.

4.3 Adoption of the DDI

In its early stage of existence, the DDI specification was primarily used by the data archive community in North America and Europe. With only a handful of tools available, the first DDI users relied on proprietary solutions to manage their metadata or even compiled the metadata by hand! The advent of the Nesstar[20] software played a key role in the adoption and success of the DDI as the only production-grade solution. In 2006, the International Household Survey Network (IHSN) integrated the Nesstar Publisher as one of the components of its Microdata Management Toolkit,[21] a set of tools targeted towards national statistical agencies in developing countries for the preservation and dissemination of survey microdata. Supported by the PARIS21 / World Bank Accelerated Data Program,[22] the toolkit has met with great success and is now in use in dozens of countries across Africa, the Middle East, Latin America, and Asia. DDI is now a truly global specification.

With the publication of DDI version 3.0, the DDI Alliance has broadened the potential user base of the specification to all agencies and individuals involved in the survey lifecycle. While no official implementation of 3.0 is currently in use, several organizations (primarily producers and Research Data Centers) have expressed interest in adopting it or are already in the initial stages of implementation. The availability of generic tools will play a major role in the success of 3.0, but once this initial hurdle is passed, a large uptake of the new version is expected.

5. The Statistical Data and Metadata Exchange

In 2001, seven international and supranational organizations organized the SDMX[23] Initiative: the Bank for International Settlements (BIS), the Organization for Economic Cooperation and Development (OECD), the European Central Bank (ECB), Eurostat, the World Bank, the International Monetary Fund (IMF), and the United Nations Statistical Division (UNSD). The initiative was formed to examine how new technologies could be used to better support the reporting and dissemination of aggregate statistics, which all of these organizations use to support policy and development activities.

In 2005, the first version of the SDMX technical standards (that is, technology standards) became an ISO Technical Specification, ISO TS-

20 http://www.nesstar.com.
21 http://www.surveynetwork.org/toolkit.
22 http://www.surveynetwork.org/adp.
23 http://www.sdmx.org.

17369. They provided an information model and XML formats for all types of aggregate data and related structural metadata, along with guidelines about how Web services should be supported. There is also a legacy format in UN/EDIFACT syntax, formerly known as GESMES/TS (but now SDMX-EDI), which is still supported under the SDMX model.

Having standard XML formats for data and structural metadata made the process of exchanging data more efficient because the data were now predictable and accompanied by rich metadata. SDMX has been implemented by many additional international organizations, and national-level institutions such as central banks and statistical offices. Adoption is global.

In 2008, the SDMX Initiative released two other important sets of products: a second and significantly expanded version of the technical specifications SDMX 2.0 (now being submitted to ISO for acceptance as an International Standard) and a set of content-oriented guidelines, which recommend how various statistical concepts in broad use can be defined, named, represented, and used.

In addition to support for aggregate datasets and related structural metadata, version 2.0 of the technical specifications provide support for all types of reference metadata, including the ability to mimic the contents of other related standards for the purposes of cross-walking. There is also a standard for providing registry services, a feature of Web services architecture that allows for the easy location of data and metadata resources around a distributed network.

It is important to note that both SDMX and DDI were designed to be aligned and to work well with other related standards – SDMX was designed with a knowledge of DDI (version 3.0 and earlier versions), and vice-versa. An effort was made to ensure that these standards are complementary rather than competitive.

6. Other specifications

There are several other standards that are of interest to the social sciences researcher. These will be given a brief mention here, and the list provided is not exhaustive.

- *ISO/IEC 11179:* This standard provides a model for understanding what it terms "data elements," which are as applicable to metadata as they are to data. The model provided gives a standard way of defining terms, the concepts they represent, the value domains they encompass, and how those value domains are represented. Additionally, a model for lifecycle management is provided. Ultimately, this is a powerful model for defining the semantics of different terms and concepts used with social sciences data.

- *ISO 19115:* This standard provides a model for defining geographies and is used by many other systems that care about geography, maps, etc. This model is embedded in DDI, for example, but is widely used.
- *Dublin Core:* Dublin core provides a set of fields for providing the citations of resources and has a core set and an extension mechanism, expressed in XML.
- *METS:* This is a standard from the world of digital archives, which provides for the packaging of a set of related objects (e.g., a webpage and the image files it references). It allows for other standard metadata formats to be embedded in it (DDI is one example of this).
- *PREMIS:* This is an XML format for expressing metadata about the archival lifecycle, and is meant to be used in combination with the OAI archival reference model.

Given the many stages data that and metadata go through in the social sciences and the different perspectives taken by the various actors, it is clear that a single metadata specification cannot be used to cover the entire life cycle. Using the DDI and SDMX as core standards and extending their functionalities through combination with the other standards mentioned above offers data producers, librarians, researchers, and other consumers a robust set of tools for the management of data and metadata across the entire lifecycle. The often non-trivial job of mapping these standards correctly to one another is being undertaken in forums such as the UN/ECE's METIS[24] conference and elsewhere.

One example of this is the use of DDI to document micro-level data sources, with resulting aggregates described using SDMX. Each standard is best suited to a different set of processes – having them well-aligned, and mapped, allows for the combined use of the standards in an efficient and consistent manner.

24 http://www.unece.org/stats/archive/04.01d.e.htm.

7. Metadata in Germany

There has been much involvement from some German organizations in the development and use of metadata standards, and today, Germany is one of the leading countries in terms of adoption of the standards described in this paper. Our impression is that the increased recent interest in DDI and other standards such as SDMX is being driven at least partly by legislative changes regarding the exchange of data between state-sponsored institutes, but we are not familiar enough with German law to make any definite pronouncement. Certainly, German involvement in metadata standards has a long history.

The involvement of Germany in the creation of metadata standards focuses mostly on DDI – some German institutes such as GESIS (*Leibniz Institute for the Social Sciences*) were very involved in both the development of past versions of the standards and also in their implementation. The German Microcensus is a good example of how DDI was – and continues to be – used for data documentation, but there are many others.

More recently, some of the other German institutes involved in social sciences and economics have started using DDI and participating actively in the DDI community. Most notably at the Research Data Centers, where an application must be submitted to gain access to confidential data, there has been an increasing uptake of and interest in the use of DDI 3.0. This reflects an international trend, but thanks to the Research Data Centers and other research institutes, Germany is one of the most active countries in the use of DDI. At the IASSIST 2008[25] conference at Stanford University, the Institute for Employment Research (IAB, *Institut für Arbeitsmarkt und Berufsforschung*)[26] presented a prototype for using the DDI 3.0 metadata model as the basis for a documentation system that will serve both the Research Data Centers and the internal research departments. At the International Data Service Center of the Institute for the Study of Labor (IZA, *Forschungsinstitut zur Zukunft der Arbeit*)[27] in Bonn, DDI 2.1 is used as the standard metadata model, and in the future DDI 3.0 will be used.

One reason for Germany's leadership role within the social sciences metadata community is the hosting of DDI-related events for the past two years at Schloss Daghstul, the Leibniz Center for Informatics. Organized by GESIS, with some co-sponsors, seminars have been held to provide an in-depth understanding of DDI 3.0, and other DDI-related meetings have taken place on related themes (in 2008, the topic was DDI 3.0 best practices). These events took place in the fall of 2007 and 2008, and it appears that they

25 http://iassist08.stanford.edu.
26 http://www.iab.de/
27 http://idsc.iza.org/

will become an annual feature of the DDI community calendar. They have attracted attendees from all over the world.

In 2009, the first European DDI User's Group meeting will be hosted by IZA, which has also played a significant role in organizing the group. Thus, it can be seen that German institutes have had a significant role in the development and use of DDI, and this role appears to be growing with the advent of DDI 3.0.

SDMX has also been supported within Germany. The Federal Statistical Office in Wiesbaden was an early participant in the SDMX Open Data Interchange (SODI) project run by Eurostat, along with a small number of other European national statistical organizations. The European Central Bank in Frankfurt – although not a German organization as such, but a European one – is one of the sponsors of SDMX (along with the BIS,[28] the IMF,[29] the OECD,[30] Eurostat, the World Bank, and the UN Statistical Division), and was also a major user of the standard on which SDMX was based, GESMES/TS.

Increasingly, there is a growing interest in the exchange of research data and statistical data both within countries and across national borders. Metadata standards such as DDI and SDMX are a critical ingredient in facilitating these exchanges. Germany has emerged as one of the more forward-looking countries in this respect.

8. Directions, challenges, and recommendations

The availability of high-quality metadata promises to drive many positive changes within the social sciences in the near future. Better metadata allows for better use of technology, which can fundamentally impact what is possible for researchers: (1) data that is better documented, easier to find and use, and of greater consistency and higher quality; (2) heightened visibility for researchers' findings and the ability to replicate and validate those findings using the actual data and processes; (3) new techniques for identifying comparable datasets and an increased level of granularity in working with data from multiple sources; (4) improved tools for data management to assist data producers, librarians, and archives; (5) and the establishment of virtual research communities.

It is worth noting that important components of the technology suite needed to achieve these benefits are *Web services*[31] *based architectures* and

28 Bank for International Settlements.
29 International Monetary Fund.
30 Organisation for Economic Co-Operation and Development.
31 http://en.wikipedia.org/wiki/Web_service.

registries.[32] The first is the industry standard technology essential for allowing applications to effectively communicate with each other and exchange information. The second implements public catalogs for applications within a domain to facilitate searching and locating data and metadata resources wherever they are located on the Internet or network. This combination is essential to support the establishment of dynamic portals and federated spaces that provide users with a virtual view of the statistical information and effective mechanisms for timely publication of data, documents, and research outputs. It also unlocks powerful features such as notification services (whereby the information automatically flows towards its intended users, not the other way around), comparability and harmonization, researcher feedback, and community-driven knowledge spaces.

Another significant emerging idea is the concept of *enhanced publications*, which combine research findings, data, and metadata as a single package, providing support for the replication standard within the social sciences. Given a collection of such publications, it becomes possible to maintain linkages between primary and secondary datasets and publications, providing for richer comparisons and broader knowledge. Well-packaged information also allows for the use of data at the level of the variable, rather than just the monolithic dataset, supporting more granular comparison and exploration by topic.

These benefits will not be achieved without meeting some significant challenges, however. These can be broken down into three categories: (1) tools, (2) metadata quality, and (3) practice. Most agencies or individuals will likely confront issues in each of these areas, but it is important to know that they do not need to do so in isolation. Organizations such as the Open Data Foundation, the DDI Alliance, the IHSN and others are working towards bringing users together for the purpose of sharing resources and expertise to jointly address metadata challenges.

(1) *Tools:* An XML specification by itself is not something that can be used out of the box. It requires software to allow for the capture, storage, publication, and exchange of the metadata. Building such products can be an expensive effort, and this problem was recognized by the DDI and SDMX sponsors. To address the issue, several initiatives are ongoing for the development of open source solutions to facilitate the use and adoption of DDI and SDMX. The DDI Foundation Tools Program[33] aims at the implementation of a DDI 3.0 core framework and utilities for implementers as well as the production of a generic DDI 3.0 editor. The Open Data Foundation is working with its partners to release a free SDMX browser tool and provides a source code repository to anyone

32 http://en.wikipedia.org/wiki/Metadata_registry.

33 http://tools.ddialliance.org.

interested in developing open source software for social science metadata management. The IHSN has also developed a DDI 2 based Microdata Management Toolkit targeted at statistical agencies in developing countries.

We therefore recommend that anyone interested in adopting a metadata standard check with the relevant organizations regarding the availability of tools and even contribute to the joint development efforts.

(2) *Metadata quality:* Having tools available does not mean that the metadata will be sound and reliable. In the end, it is the content that counts, and compiling high-quality comprehensive metadata also requires good techniques, guidelines, and a significant amount of discipline. While some of the work can be automated or semi-automated using software utilities, it is often necessary to compile information by hand and chase down metadata to find the missing piece of knowledge or document. This is particularly true when the metadata is captured after the fact or after back-logging. This implies that human error and missing information are a factor. Quality assurance is therefore a very important aspect of metadata management, and any organizations adopting standards should thoroughly document these processes. As a general rule, metadata should be treated as an official publication and should therefore follow the same institutional rules.

Harmonization of practices across organizations also plays a major role when the metadata leaves the institution and is shared with users or other partners. If the same metadata elements are documented using different principles, they will no longer be coherent, which can confuse users, impact comparability, and reduce system interoperability.

Agencies such as the DDI Alliance, the IHSN, or SDMX sponsors produce generic guidelines and best practices for the preparation of metadata. They also work closely with metadata producers toward the harmonization of metadata elements. When looking into metadata quality assurance issues, we therefore suggest that users consult the existing websites and literature for references or join existing initiatives. We also recommend that agencies working in smaller communities actively collaborate on metadata harmonization.

(3) *Practice:* Adopting new standards and technologies implies a change in the way the organizations and individuals have been operating. While the benefits of a sound metadata management framework are extensive, this inevitably meets some resistance and requires a certain amount of resources to foster acceptance. Just because the tools and guidelines exist to help realize the benefits does not mean that people will use them. Researchers in particular are often reluctant to recognize that new techniques and discipline are necessary. Awareness, training, and integration

are all adoption issues facing researchers, archives, and data providers. Highlighting the benefits and providing incentives will be necessary to achieve successful integration.

Given the strong interest of data providers in metadata standards, we anticipate the adoption of DDI and SDMX to continue accelerating strongly in the coming years. A key to this success will be the availability of generic software tools. Sponsors and community-driven open source initiatives are expected to contribute a wide range of generic products for the management, publication, and sharing of metadata that will foster adoption of standards. These initial efforts will likely start to produce significant results in 2009–2010. In the meantime, statistical agencies and Research Data Centers with strong internal IT capacities will likely design their own tools in parallel to manage metadata. As the potential market grows in size, it is also possible that statistical packages or other commercial vendors will begin to provide solutions as well.

While the metadata will initially continue to emerge primarily from data archives, the uptake among producers should increase, improving overall quality as the information is captured closer to its source. Researchers will also likely begin to contribute to the metadata knowledge. Such end user adoption may be slow at first, but incentives and benefits should quickly overcome the resistance to change, and we should see an increase in user based metadata. This overall will foster the existence of shared knowledge spaces through metadata and bridge the communication gap that often exists between user and producer.

Given that many actors will now be contributing to the metadata, best practices and harmonization will play a crucial role in the overall quality and consistency of the information. Led by sponsors and major statistical agencies, national and international initiatives will likely emerge to draft metadata management guidelines and work towards the harmonization of common metadata elements. This will not only lead to improved metadata but will also foster better and more comparable data.

As more and more standard metadata is being produced, the need for exchange, sharing, and publication will quickly increase. As end users prefer to have single point of entry, national, regional, and international catalogs or registries will grow in importance. This aggregation of information will support the development of large collections of information that could potentially support complex searches and metadata mining operations. Note that such registries do not store the actual data. They act as "lookup points" that are used to retrieve the location where the information actually is (just like a phone or address book).

In order to foster broad adoption of metadata and related best practices in social sciences, we recommend the following:

(1) Promote the importance of high-quality data documentation and its capture using metadata standards.

(2) Familiarize producers, archives, and researchers with metadata standards, related best practices, and technologies.

(3) Support the development of standards-based tools, preferably under an open source license and aligned on community recommendations.

(4) Do not undertake metadata adoption activities in isolation. Instead, join and sponsor community or government-backed initiatives.

(5) For data and metadata managers and providers, support the establishment of an industry standard, Web service-oriented, and registry-based IT infrastructure to facilitate the management, exchange, reuse, and harmonization of metadata and data.

(6) Integrate metadata capture at all stages of the life cycle. Document events as they happen, not after the fact.

(7) Leverage on the availability of metadata to automate the production of documentation or generation of statistical scripts to reduce the overall production costs, increase quality, and deliver user-customized products.

(8) Support the establishment of virtual research and collaborative spaces to allow for user-driven metadata and foster community knowledge capture.

Overall, the future of social science metadata looks very bright. The availability of robust standards combined with modern technologies has laid the foundation of a global harmonized framework for the management of social science data and documentation. Just as the Internet has revolutionized and connected our world, social science metadata has the potential to open new possibilities for producers, archives, and users.

4. Paradata

Frauke Kreuter and Carolina Casas-Cordero

Contact:

Frauke Kreuter
University of Maryland
Joint Program in Survey Methodology
1218 Lefrak Hall
College Park, MD 20742
US
e-mail: fkreuter[at]survey.umd.edu

Abstract

Paradata – data about the process of survey production – have drawn increasing attention as the statistical world moves towards the implementation of quality metrics and measures to improve quality and save costs. This paper gives examples of various uses of paradata and discusses access to paradata, as well as future developments.

Keywords: paradata, process data, responsive design, measurement error, nonresponse, adjustment

1. Introduction

During the last two decades, survey researchers have begun to use computer-assisted methods to collect social science data. This trend is most obvious in web surveys, but is equally present in telephone surveys that use automated call scheduling systems or mail surveys that take advantage of logs provided by postal services. All of these systems produce data about the survey process as a by-product, which Mick Couper coined paradata in a presentation at the Joint Statistical Meeting in Dallas (Couper, 1998). Inspired by Couper's suggestions to use data automatically generated by computer-aided systems to evaluate survey quality, survey methodologists have since then broadened the concept of paradata to other aspects of the survey process and other modes of collection.

Data about survey process have drawn increasing attention as the statistical world moves towards the implementation of quality metrics, measures to improve quality and save costs, and a framework in which to measure total survey error (Biemer and Caspar 1994; Lyberg et al. 1997; Aitken et al. 2004; Couper and Lyberg 2005). Both data users and data producers are now aware of the potential benefits of paradata. This has been reflected by growing interest at invited paper sessions at international conferences such as the International Workshop on Household Survey Nonresponse, bi-annual conferences of the European Survey Research Association (ESRA), annual conferences of the American Association of Public Opinion Research (AAPOR), Joint Statistical Meetings (JSM), and the Sessions of the International Statistical Institute (ISI), as well as the quality conferences co-organized by Eurostat.

2. Examples for paradata and their use

There is no standard definition in the literature of what constitutes paradata. Several papers attempt to systematize data that are not part of the actual interview (Scheuren 2000; Couper and Lyberg 2005; Scheuren 2005; O'Reilly 2009), but each of these papers varies slightly in terminology and in what is considered paradata. Paradata was originally conceptualized as the data automatically generated as the by-product of the computer-assisted survey process (e.g., call record data and keystrokes), but the term has more recently been expanded to include information that may be recorded by interviewers (e.g., observations), or captured through additional systems (e.g., digital audio recording) (Couper 1998).

For this review we do not seek to provide a fixed definition of paradata. What is important in our opinion is the concept of data collected during and about the survey process. These data can be used to understand and improve the process (and subsequently the end result). Thus, instead of a definition, we give some examples of how paradata is currently being used around the world.

One set of data typically referred to as paradata are call records collected during the process of contacting a sample case. The time of contact (day and time), as well as the outcome of a call (non-contact, refusal, ineligible, interview, appointment, etc.) are almost always available on these call records (Heerwegh et al. 2007; Blom et al. forthcoming). These variables are either recorded by the interviewer (with PAPI or CAPI systems) or automatically, as is commonly the case for call schedulers in computer-aided telephone interviews (CATI). The recording of the date and time of a prior contact allows call schedulers to vary contact attempts with the hope of increasing the probability of a successful contact (Weeks et al. 1987; Kulka and Weeks 1998; Greenberg and Stokes 1990; Stokes and Greenberg 1990; Brick et al. 1996; Sangster and Meekins 2004; Wagner and Raghunathan 2007), and ideally to reduce the cost (Groves 1989; Triplett 2002; Murphy et al. 2003). Prominent examples of call record data collected in face-to-face surveys are the Contact History Instrument (CHI) implemented in surveys by the US Census Bureau (Bates 2003), or the standard contact forms that have been requisite since round one of the European Social Survey (Stoop et al. 2003). In some instances, call record data are used to guide decisions on responsive or two-phase sampling designs (Groves et al. 2003; Kennickell 2003; Groves and Heeringa 2006; Eckman and O'Muircheartaigh 2008), or to gain knowledge about optimal calling patterns in face-to-face surveys in general (Matsuo et al. 2006; Durrant et al. 2009). To our knowledge, there is so far only one survey, the US National Survey of Family Growth (Lepkowski et al. 2009), in which call record data from face-to-face surveys are used to drive centralized day-to-day field decisions similar to those in supervised call centers. For most surveys, face-to-face call record data are analyzed after the fact to assess interviewer efforts and compliance with pre-specified design requests (Billiet and Pleysier 2007; Lipps 2007; Koch et al. 2009).

Regardless of the mode of data collection, survey methodologists use call record data to study various aspects of survey participation. Call record data are available for both respondents and non-respondents to nay given survey and are thus prime candidates for the study of nonresponse bias, for example, through level-of-effort analyses, in which early respondents are compared to late responders assuming that later responders are more similar to non-responders than early responders (Stinchcombe et al. 1981; Smith 1984; Schnell, 1998; Kennickell 1999; Chiu et al. 2001; Duhart et al. 2001; Lynn et al. 2002; Lynn 2003; Wang et al. 2005; Stoop 2005; Voogt and Saris 2005; Billiet et al. 2007; for a meta-analysis of the results, see Olson 2010). With the goal of assessing net quality gains, researchers have used call record data to shed light on the relationship between nonresponse and measurement error (Green 1991; Yan et al. 2004; Olson 2006; Peytchev and Peytcheva 2007; Yu and Yan 2007).

A second set of data subsumed under the concept of paradata is also collected during the initial phase of establishing contact and convincing sample units to participate in the survey. These paradata are observations made by the interviewer. Like call record data, these interviewer observations are available on all sampled cases and thus suitable to inform survey design decisions (Copas and Farewell 1998; Lynn 2003; Groves et al. 2007) and assess nonresponse bias (Maitland et al. 2009). In recent face-to-face surveys, interviewers are charged with collecting observations of neighbor-hoods and housing unit characteristics in a number of surveys usually along the lines suggested by Campanelli et al. (1997), Groves and Couper (1998), or Lynn (2003). Examples are the US Health and Retirement Study, the US Study of Early Child Care, the US Survey of Consumer Finances, the US National Survey on Drug Use and Health, the British Election Study, the British Crime Survey, the British Social Attitudes Survey, the European Social Survey, and the Survey of Health, Ageing and Retirement in Europe. Some rather novel interviewer observations are those that are tailored to the survey topic and thus have higher potential to be useful for adaptive survey design decisions or nonresponse adjustment. Again, a prime example is the National Survey of Family Growth, in which interviewers are asked to guess whether or not the sample person is currently in an active sexual relationship (with an opposite-sex partner), and whether or not children are present in the household Groves et al. (2007). Other sets of interviewer observations made at the doorstep are those capturing the interaction between interviewer and respondent and respondents' reasons for refusal (Campanelli et al. 1997; Bates and Piani 2005; Bates et al. 2008).

Both call record data and interviewer observations have the potential to enhance current nonresponse adjustments. Not only are they available for both respondents and nonrespondents, but ideally are they predictive of the sampled person's probability of responding to a survey and of the survey variables of interest. Over the years, survey methodologists have extensively researched and developed covariates of survey participation (Schnell, 2005; Groves and Couper 1998), many of which are now part of call record and contact data forms. The possibility of using call record data for nonresponse adjustment has been discussed for quite some time (Drew and Fuller 1980; Potthoff et al. 1993), and current papers demonstrate the relationship between information in call records and the probability of responding to a survey request (Beaumont 2005; Biemer and Wang 2007; Blom 2009; Kreuter and Kohler 2009). Interviewer observations of variables close to the survey (such as the presence of children in a fertility survey) can complement call record data in response propensity models due to their likely stronger relationship to survey variables of interest (Kreuter et al. 2010). Difficult issues in modeling may, however, arise when strong predictors of response are combined with strong predictors of survey outcome variables (Kreuter and Olson 2010).

In computer-aided surveys, a third set of paradata can be captured: audio-recordings of the interaction between interviewer and respondent. Researchers have suggested that vocal characteristics of the respondent and interviewer are in part responsible for successful recruitment attempts. Especially during telephone interviews, potential respondents have very little information about the interviewer, aside from how he/she sounds, speaks, and interacts when they decide whether or not to participate in a telephone interview (Groves et al. 2007; Best et al. 2009). Yet interviewers vary widely in how often their invitations lead to participation, suggesting that potential respondents may give considerable weight to interviewers' verbal attributes. Recordings and paradata derived from them are of interest, not only because they can shed light on survey participation, but also because they can be used to assess measurement errors on a question level (Jans 2010). Recordings become more common as digital storage becomes less expensive (Couper 2005; Thissen et al. 2007). However, the post-processing of such recordings into usable paradata is a large task and has been undertaken in only a few methodological studies. Those studies make use of recent developments in the field of acoustical engineering and new software, which makes it possible for researchers to automatically process audio files and obtain objective data on voice characteristics such as disfluencies, pauses, interruptions, speech rate, and pitch (Jans 2010; Conrad et al. 2010).

In addition to audio-recordings, computer-assisted survey instruments facilitate the automated collection of paradata that can be used to assess measurement error at the question level. Most data collection software records the time used to complete a question, a set of questions, or the whole interview (response times), and capture key strokes, with which researchers can, for example, measure how often a respondent backed up and changed an answer and whether supplementary definitions are used (Couper 1998). All of these measures are available for computer-aided personal interviews (CAPI), computer-aided telephone interviews (CATI) and Web surveys. Web surveys also differentiate between paradata that include characteristics of a respondent browser captured from server logs (server-side paradata) and respondent behavior captured by embedding JavaScript code into the instrument (client-side paradata). Response times and key stroke measures have been used to study aspects of the response process (Bassili and Fletcher 1991; Kreuter, 2002; Heerwegh 2003; Kaminska and Billiet 2007; Yan and Tourangeau 2008; Couper et al., 2009; Lenzner et al. 2009; Peytchev 2009), to guide interventions in Web surveys (Conrad et al. 2009), evaluate interviewers (Couper et al. 1997; Mockovak and Powers 2008), and review the performance of questions in pretests (Couper 2000; Stern 2008; Hicks et al. 2009).

Our list of examples is by no means complete, but it does give a flavor of the many uses of data auxiliary to the main data collection that contain information about the process with which the data are collected. There is, in addition, an entirely different usage of paradata beyond monitoring, managing, modeling, and improving the data collection process. Summary statistics of paradata are also used to describe the dataset as a whole: response rates (created out of recordings of the final status in call records) are examples of such survey-level statistics. While paradata contribute to such summary statistics, the summary statistics themselves are usually not referred to as paradata but called metadata instead (Couper and Lyberg 2005; Scheuren 2005).

Auxiliary data available on the case level that come from an entirely different source are also usually not considered paradata (i.e., administrative data, data from commercial lists, or data available on sampling frames). A more borderline case are separate surveys of the interviewers themselves (Siegel and Stimmel 2007). To the extent that information from interviewers can help to understand the survey process, they can be viewed as paradata (like interviewer observations, for example). Metadata and auxiliary data also play increasing roles in monitoring and enhancing data quality. For some recent initiatives in using such auxiliary data, see Smith (2007; 2009).

2.1 Databases and data access

Unlike survey data themselves and metadata about those surveys, paradata are usually not made publicly available for several reasons. For one, it is not common to release gross-sample data, i.e., data records that include all sampled units, both those that respond to the survey request and those that do not. Second, paradata are often not collected on the same unit of analysis as the survey data are, making the release of such datasets more complicated. Call record data are usually collected at each call attempt, which could easily generate up to fifty records for cases fielded in a telephone survey. Response times are collected at an item level and sometimes twice within one item (if the time to administer the item is measured separately from the time the respondent took to answer the question). Vocal properties of an interviewer are recorded on a finer level and could generate several records even within the administration of a single item. Third, the format of these paradata varies a great deal by data collection agency and system: for example, outcome codes on call record data vary across agencies and modes of contact available to the interviewer (Blom et al. 2008). While the lack of standards for the collection and release of paradata is not a problem per se (except for making data preparation work more burdensome for analysts), it does require proper documentation, which is usually not covered by data collection grants. Fourth, for some of the paradata, there are open legal and ethical questions. Detailed observations of the neighborhood or housing unit might facilitate the de-identification of survey respondents. For Web surveys, Couper and Singer (2009) raise the question of whether respondents should be informed about the capturing of client-side paradata in particular if they are used to understand or even control respondent behavior, and not just used for improvement of the design or performance of the instrument.

Some important surveys do release their paradata to the public. Examples are contact protocol data from the European Social Survey, paradata from the US National Health Interview Survey, and paradata from the American National Election Survey (the latter being available for secondary analysis upon request).

3. Future developments

3.1 Data provision

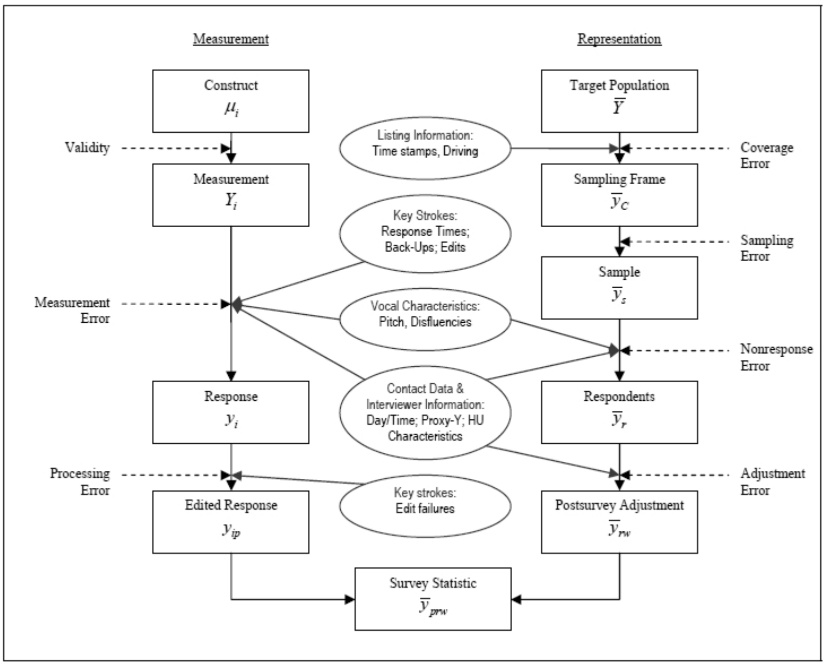
As the previous section showed, the potential uses of paradata are wide-ranging. Survey methodologists have started to exploit paradata to guide intervention decisions during data collection and to provide opportunities for cost savings. To the extent that errors cannot be prevented, paradata also help us to detect errors after the fact (thus providing guidance for the next survey) and to model and adjust for them. So far, a series of paradata have been used to assess or model measurement error, nonresponse error, and even the interaction of the two. Until now, very few paradata have been collected for other parts of the process. If we match the most commonly collected paradata to the various error sources in a total survey error framework (see figure 1), we see that for several process steps in the generation of survey statistics, no paradata are currently available. The systematic documentation of questionnaire development by Schnell et al. (2008) could lead to new paradata for the creation of measurement indicators.

From a quality monitoring and improvement perspective, a more structured approach towards the selection, measurement, and analysis of key process variables would be desirable (Morganstein and Marker 1997). Ideally, survey researchers would specify a set of product characteristics and underlying processes associated with these characteristics, and then these processes would be checked by means of key process variables.

The virtue of paradata as a by-product of the survey process is that they come cheap to the data collector. If paradata are used systematically for process improvement and postprocess analyses, then their structure will probably change: variables will be added (e.g., new interviewer observations) and requests for standardization might turn out to conflict with existing collection systems. Paradata might then no longer be just a by-product, but a product with costs attached to it. It is up to survey methodologists to prove that paradata provide the cost control (or even cost savings) and performance increases that they have promised. Without the demonstration of repeated and successful use, survey methodologists will face difficulties in convincing data collection agencies to routinely collect such data.

One obstacle to demonstrating the usefulness of paradata is the quality of the data itself. While paradata might help to address some of the errors present in survey data, the data may suffer from measurement error, missing data, etc. Interviewers can erroneously record certain housing unit characteristics, can misjudge features about the respondents, or can fail to record a contact attempt altogether (Casas-Cordero 2010; Sinibaldi 2010; West 2010). For example, it is possible that paradata are subject to high variation in the way the information is recorded by different interviewers (e.g., evaluation of the condition of the house relative to other houses in the area) or some interviewers may simply not place high priority on filling in the interviewer observation questionnaires because they are not paid for doing so. Some studies have shown high levels of missing data in interviewer observations, indicating a lack of data quality (Kreuter et al. 2007; Durrant et al. 2009). Such missings may occur, for example, if the interviewer does not have enough time or does not feel the need to fully record every contact attempt to the household. Likewise, scripts embedded in Web surveys can fail to install properly and client-side data are not captured as intended, and recordings of interviewer administered surveys can be inaudible due to background noise or loose microphones (McGee and Gray 2007; Sala et al. 2008).

Figure 1: Total Survey Error components and paradata for their assessment (modified graph from Groves et al. 2004)



As long as these recording errors and missing data patterns are not systematic, they will reduce the effectiveness of paradata for process improvement and error modeling, but should not threaten them altogether. If errors appear systematically (e.g., savvy users in Web surveys prevent scripts from capturing key strokes), resulting conclusions are threatened to be biased. Currently, not enough is known about the measurement error properties of paradata.

3.2 Data usage

As mentioned before, a key challenge to the use of paradata is their unusual data structure, with time-dependent observations on multiple levels collected through various modes with varying instruments. If we again take call record data as an example, the literature is still dominated by analyses using case-level aggregate statistics of call-level data (e.g., total number of contact attempts, total number of refusals), while some more recent examples take advantage of the multilevel structure by using survival models or multilevel discrete time event history models in predicting propensities to respond (Durrant and Steele 2009; Olson and Groves 2009; Wagner 2009).

Many methodological questions concerning how to make best use of paradata are still unsolved. In the estimation of response propensity models, we do not know yet if time should be modeled discretely as active day in the field or relative to the time since beginning of the field period. Nor is it clear how to best combine paradata into nonresponse propensity models with the aim of adjusting for survey nonresponse (Kreuter and Olson 2010). When dealing with response latencies, we do not yet know how best to handle unusually long response times, how best to model time dependency within the process of answering multiple subsequent survey questions, etc. Closer collaboration among survey methodologists, statisticians, and econometric modelers could benefit the research in this area.

Methodologists who use paradata for management and monitoring are still experimenting with tools for displaying the constant flow of process information. A "dashboard" was developed at the Institute for Social Research in Michigan (Groves et al. 2008; Lepkowski et al. 2009) to provide survey managers and principal investigators with timely access to data, and tools to facilitate decision-making – but there is still room for improvement (Couper 2009). The use of process control charts has been proposed before (Deming 1986; Morganstein and Marker 1997; Couper and Lyberg 2005), but so far, no standard charts have been developed to monitor survey data collection. Increased access to paradata and in particular timely update of such data streams will increase the need for good tools to display and analyze paradata.

3.3 Data access

To address the risk of de-identification of respondents, the paradata that pose this danger could be made available in Research Data Centers where access and usage of data is monitored. Given the potential of certain paradata to improve nonresponse adjustment, an entirely new data retrieval system might be worth considering. Given appropriate paradata, nonresponse adjustment can be tailored to individual analyses. Usually, only one set of nonresponse adjustment weights is created and distributed with survey data. Growing nonresponse has made the assumption that a single adjustment strategy is sufficient for all statistics produced by a survey less tenable. A data retrieval system could be conceptualized that allows the on-demand creation of adjustment weights based on the planned analysis.

Public access to paradata also allows a post-hoc examination of the procedures followed by the data collection institutes. If survey organizations are aware that process information will become public, this might lead overall to a higher data collection standard. Obviously higher-quality work will come with a price. However, some survey organizations might not want to release paradata, as it discloses information about their fieldwork procedures. If these procedures are considered to be proprietary, the disclosure could be seen as an impingement on their comparative advantage.

4. Discussion

Survey data collection is essentially a production process with a product. Surveys do not differ in this respect from other organizations that produce products or services and are concerned about their quality. Management strategies for such organizations have moved to what are called continuous quality improvement methods (Imai 1986; Deming 1986), in which measures of the process are monitored along the way so that error sources can be located and interventions planned (examples of such strategies are Total Quality Method, TQM, or Six Sigma). Several researchers have suggested the application of such strategies to the process of survey operations (Biemer and Caspar 1994; Morganstein and Marker 1997). Paradata, as discussed here, can play an important role in the application of such strategies. The European Statistical System has developed a handbook on improving quality through the analysis of paradata (Aitken et al. 2004), but the work is still not done, and individual surveys might do well to identify key process variables for their specific circumstances (Couper and Lyberg 2005).

Survey data collection faces major uncertainties in the planning stages. It is difficult to estimate the effectiveness of measures taken to establish contact with households, identify eligible persons, select a respondent, gain that person's cooperation, and complete the interview. Likewise, estimates of the cost implications of any of these steps are often difficult to make. Responsive designs (Groves and Heeringa 2006) seek to address this uncertainty by measuring the results of various survey design features, often experimentally, and then use these measurements to intervene in the field data collection process. This monitoring includes both the paradata as well as key survey estimates. To the extent that the paradata provide information about the risk of nonresponse bias, the responsive design is capable of reducing the risk of this bias. Much more effort is needed to manage the costs of alternative design features.

To increase the conditions for high-quality collection of paradata, a survey climate is necessary that allows for experimental manipulation within the field process. Pooling data across studies can also help to disentangle confounding elements; for this, some standardization of paradata would be necessary (Blom et al. 2008). Panel data enjoy the luxury of repeated measures of observations. Researchers only recently started to explore the potential of paradata to examine attrition (Lepkowski and Couper 2002; Kreuter and Jäckle 2008) and measurement error in relation to interviewer characteristics (Jaeckle et al. 2009; Weinhardt and Kreuter 2009; Yan and Datta 2009).

Compared to other countries, data collection in Germany is not as "paradata-rich" as it could be. Since 1995, Schnell and his colleagues suggested the inclusion of contact protocol data for the gross sample to be a standard deliverable (Schnell et al. 1995). Very few surveys followed this suggestion. Furthermore, systems should be developed and put in place that allow data collection agencies to engage in data-driven interventions into the fieldwork process. For a single survey, the start-up costs might be too high and survey organizations might not see the need for such investments. If, however, the German social science data community as a whole demands paradata for process controls, investments in the respective systems might be economical. Investment into the development of new statistical tools and methods is also needed to help make sense of the vast amount of unstructured paradata generated by modern survey process. The standard analytic tools we use for survey data are not appropriate for much of the paradata we need to analyze. Here, too, collaboration throughout the social science data community would be a good first step.

References:

Aitken, A./Hörngren, J./Jones, N./Lewis, D. and Zilhão, M.J. (2004): Handbook on improving quality by analysis of process variables. Technical report, Eurostat.

Bassili, J.N. and Fletcher, J.F. (1991): Response-time measurement in survey research a method for CATI and a new look at nonattitudes. Public Opinion Quarterly 55 (3), 331-346.

Bates, N. (2003): Contact histories in personal visit surveys: The Survey of Income and Program Participation (SIPP) methods panel. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Bates, N./Dahlhamer, J. and Singer, E. (2008): Privacy concerns, too busy, or just not interested: Using doorstep concerns to predict survey nonresponse. Journal of Official Statistics 24 (4), 591-612.

Bates, N. and Piani, A. (2005): Participation in the National Health Interview Survey: Exploring reasons for reluctance using contact history process data. In: Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.

Beaumont, J. (2005): On the use of data collection process information for the treatment of unit nonresponse through weight adjustment. Survey Methodology 31 (2), 227-231.

Best, H./Bauer, G. and Steinkopf, L. (2009): Interviewer voice characteristics and productivity in telephone surveys. Paper presented at the European Survey Research Association (ESRA) Conference, Warsaw, Poland.

Biemer, P. and Caspar, R. (1994): Continuous quality improvement for survey operations: Some general principles and applications. Journal of Official Statistics 10 (3), 307-326.

Biemer, P. and Wang, K. (2007): Using callback models to adjust for nonignorable nonresponse in face-to-face surveys. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Billiet, J./Philippens, M./Fitzgerald, R. and Stoop, I. (2007): Estimation of nonresponse bias in the European Social Survey: Using information from reluctant respondents. Journal of Official Statistics 23 (2), 135-162.

Billiet, J. and Pleysier, S. (2007): Response based quality assessment in the ESS – Round 2. An update for 26 countries. Technical report, Center for Sociological Research (CeSO), K.U. Leuven.

Blom, A. (2009): Nonresponse bias adjustments: What can process data contribute? Iser working paper 2009-21, Institute of Social research (ISER), Essex University.

Blom, A./Lynn, P. and Jäckle, A. (2008): Understanding cross-national differences in unit nonresponse: The role of contact data. Technical report, Institute for Social and Economic Research ISER.

Blom, A./Lynn, P. and Jäckle, A. (forthcoming): Understanding cross-national differences in unit nonresponse: The role of contact data. In: Harkness, J.A./Edwards, B./Braun, M./Johnson, T.P./Lyberg, L.E./Mohler, P.P./Pennell, B.-E. and Smith, T. (Eds.): Survey Methods in Multinational, Multiregional, and Multicultural Contexts. New York.

Brick, J.M./Allen, B./Cunningham, P. and Maklan, D. (1996): Outcomes of a calling protocol in a telephone survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Campanelli, P./Sturgis, P. and Purdon, S. (1997): Can you hear me knocking: An investigation into the impact of interviewers on survey response rates. Technical report, The Survey Methods Centre at SCPR, London.

Casas-Cordero, C. (2010): Testing competing neighborhood mechanisms influencing participation in household surveys. Paper to be presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Chicago, IL.

Chiu, P./Riddick, H. and Hardy, A. (2001): A comparison of characteristics between late/difficult and non-late/difficult interviews in the National Health Interview Survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Conrad, F./Broome, J./Benki, J./Groves, R.M./Kreuter, F. and Vannette, D. (2010): To agree or not to agree: Effects of spoken language on survey participation decisions. Paper to be presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Chicago, IL.

Conrad, F./Couper, M./Tourangeau, R./Galesic, M. and Yan, T. (2009): Interactive feedback can improve accuracy of responses in web surveys. Paper presented at the European Survey Research Association (ESRA) Conference, Warsaw, Poland.

Copas, A. and Farewell, V. (1998): Dealing with non-ignorable nonresponse by using an 'enthusiasm-to-respond' variable. Journal of the Royal Statistical Society, Series A 161 (3), 385-396.

Couper, M. (1998): Measuring survey quality in a CASIC environment. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Couper, M. (2000): Usability evaluation of computer-assisted survey instruments. Social Science Computer Review 18 (4), 384-396.

Couper, M. (2005): Technology trends in survey data collection. Social Science Computer Review 23, 486-501.

Couper, M. (2009): Measurement error in objective and subjective interviewer observations. Modernisation of Statistics Production, Standardisation, Efficiency, Quality assurance and Customer Focus, 2-4 November 2009, Stockholm, Sweden.

Couper, M./Hansen, S. and Sadosky, S. (1997): Evaluating interviewer performance in a CAPI survey. In: Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N. and Trewin, D. (Eds.): Survey Measurement and Process Quality. New York.

Couper, M. and Lyberg, L. (2005): The use of paradata in survey research. In: Proceedings of the 55th Session of the International Statistical Institute, Sydney, Australia.

Couper, M. and Singer, E. (2009): Ethical considerations in the use of paradata in web surveys. Paper presented at the European Survey Research Association (ESRA) Conference, Warsaw, Poland.

Couper, M./Tourangeau, R. and Marvin, T. (2009): Taking the audio out of Audio-CASI. Public Opinion Quarterly 73 (2), 281-303.

Deming, W. (1986): Out of the Crisis. Cambridge.
Drew, J. and Fuller, W. (1980): Modeling nonresponse in surveys with callbacks. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Duhart, D./Bates, N./Williams, B./Diffendal, G. and Chiu, P. (2001): Are late/difficult cases in demographic survey interviews worth the effort? A review of several federal surveys. In: Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.
Durrant, G./D'Arrigo, J.and Steele, F. (2009): Using field process data to predict best times of contact conditioning on household and interviewer influences. Technical report, Southampton Statistical Science Research Institute.
Durrant, G. and Steele, F. (2009): Multilevel modelling of refusal and noncontact in household surveys: Evidence from six UK government surveys. Journal of the Royal Statistical Society, Series A 172 (2), 361-381.
Eckman, S. and O'Muircheartaigh, C. (2008): Optimal subsampling strategies in the General Social Survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Eifler, S./Thume, S. and Schnell, R. (2009): Unterschiede zwischen subjektiven und objektiven Messungen von Zeichen öffentlicher Unordnung. In: Weichbold, M./Bacher, J. and Wolf, Ch. (Eds.): Umfrageforschung. Herausforderungen und Grenzen. Wiesbaden. (Sonderheft 9 der Österreichischen Zeitschrift für Soziologie).
Green, K.E. (1991): Relunctant respondents: Differences between early, late and nonrespondents to a mail survey. Journal of Experimental Education 56, 268-276.
Greenberg, B. and Stokes, S. (1990): Developing an optimal call scheduling strategy for a telephone survey. Journal of Official Statistics 6 (4), 421-435.
Groves, R.M. (1989): Survey Errors and Survey Costs. New York.
Groves, R.M. and Couper, M. (1998): Nonresponse in Household Interview Surveys. New York.
Groves, R.M./Fowler, F./Couper, M. and Lepkowski, J. (2004): Survey Methodology. New York.
Groves, R.M. and Heeringa, S. (2006): Responsive design for household surveys: Tools for actively controlling survey errors and costs. Journal of the Royal Statistical Society, Series A 169 (3), 439-457.
Groves, R.M./Wagner, J. and Peytcheva, E. (2007): Use of interviewer judgments about attributes of selected respondents in post-survey adjustment for unit nonresponse: An illustration with the National Survey of Family Growth. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Groves, R.M./Kirgis, N./Peytcheva, E./Wagner, J./Axinn, B. and Mosher, W. (2008): Responsive design for household surveys: Illustration of management interventions based on survey paradata. NCRM Research Methods Festival, St Catherine's College, Oxford, UK.
Groves, R.M./Van Hoewyk, J./Benson, F./Schultz, P./Maher, P./Hoelter, L./Mosher, W./Abma, J. and Chandra, A. (2003): Using process data from computer-assisted face to face surveys to help make survey management decisions. Paper to be

presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Nashville, TN.
Heerwegh, D. (2003): Explaining response latencies and changing answers using client-side paradata from a web survey. Social Science Computer Review 21 (3), 360-373.
Heerwegh, D./Abts, K. and Loosveldt, G. (2007): Minimizing survey refusal and noncontact rates: Do our efforts pay off? Survey Research Methods 1 (1), 3-10.
Herget, D./Biemer, P./Morton, J. and Sand, K. (2005): Computer audio recorded interviewing (CARI): Additional feasibility efforts of monitoring field interview performance. Paper presented at the U.S. Federal Conference on Statistical Method.
Hicks, W./Edwards, B./Tourangeau, K./Branden, L./Kistler, D./McBride, B./Harris-Kojetin, L. and Moss, A. (2009): A system approach for using CARI in pretesting, evaluation and training. Paper presented at the FedCasic Conference, Delray Beach, FL.
Imai, M. (1986): Kaizen: The Key to Japan's Competitive Success. New York.
Jaeckle, A./Sinibaldi, J./Tipping, S./Lynn, P. and Nicolaas, G. (2009): Interviewer characteristics, their behaviours, and survey outcomes.
Jans, M. (2010): Verbal Paradata and Survey Error: Respondent Speech, Voice, and Question-Answering Behavior Can Predict Income Item Nonresponse. Ph.D. thesis, University of Michigan, United States.
Kaminska, O. and Billiet, J. (2007): Satisficing for reluctant respondents in a cross-national context. Paper presented at the European Survey Research Association (ESRA) Conference, Prague, Czech Republic.
Kennickell, A. (1999): What do the 'late' cases tell us? Evidence from the 1998 Survey of Consumer Finances. Paper presented at the International Conference on Survey Nonresponse, Portland, OR.
Kennickell, A. (2003): Reordering the darkness: Application of effort and unit nonrepsonse in the Survey of Consumer Finances. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Koch, A./Blom, A./Stoop, I. and Kappelhof, J. (2009): Data collection quality assurance in cross-national surveys: The example of the ESS. Methoden, Daten, Analysen 3 (2), 219-247.
Kreuter, F. (2002): Kriminalitätsfurcht: Messung und methodische Probleme. Opladen.
Kreuter, F. and Jäckle, A. (2008): Are contact protocol data informative for non-response bias in panel studies? A case study form the Northern Ireland subset of the British Household Panel Survey. Panel Survey Methods Workshop, University of Essex, Colchester, UK.
Kreuter, F. and Kohler, U. (2009): Analyzing contact sequences in call record data. Potential and limitation of sequence indicators for nonresponse adjustment in the European Social Survey. Journal of Official Statistics 25 (2), 203-226.
Kreuter, F./Lemay, M. and Casas-Cordero, C. (2007): Using proxy measures of survey outcomes in post-survey adjustments: Examples from the European Social Survey (ESS). In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Kreuter, F. and Olson, K. (2010): Multiple auxiliary variables in nonresponse adjustment. [Manuscript under review].

Kreuter, F./Olson, K./Wagner, J./Yan, T./Ezzati-Rice, T./Casas-Cordero, C./Lemay, M./Peytchev, A./Groves, R.M. and Raghunathan, T. (2010): Using proxy measures and other correlates of survey outcomes to adjust for nonresponse: Examples from multiple surveys. Journal of the Royal Statistical Society, Series A.

Kulka, R. and Weeks, M. (1998): Toward the development of optimal calling protocols for telephone surveys: A conditional probabilities approach. Journal of Official Statistics 4 (4), 319-332.

Lenzner, T./Kaczmirek, L. and Lenzner, A. (2009): Cognitive burden of survey questions and response times: A psycholinguistic experiment. Applied Cognitive Psychology.

Lepkowski, J. and Couper, M. (2002): Nonresponse in longitudinal household surveys. In: Groves, R.M./Dillman, D./Eltinge, J. and Little, R. (Eds.): Survey Nonresponse. New York.

Lepkowski, J./Groves, R.M./Axinn, W./Kirgis, N. and Mosher, W. (2009): Use of paradata to manage a field data collection. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Lipps, O. (2007): Cooperation in centralised CATI panels – A contact based multilevel analysis to examine interviewer, respondent, and contact effects. Paper presented at the European Survey Research Association (ESRA) Conference, Prague, Czech Republic.

Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N. and Trewin, D. (1997): Survey Measurement and Process Quality. New York.

Lynn, P. (2003): PEDAKSI: Methodology for collecting data about survey non-respondents. Quality and Quantity 37 (3), 239-261.

Lynn, P./Clarke, P./Martin, J. and Sturgis, P. (2002): The effects of extended interviewer efforts on nonresponse bias. In: Dillman, D./Eltinge, J./Groves, R.M. and Little, R. (Eds.): Survey nonresponse. New York.

Maitland, A./Casas-Cordero, C. and Kreuter, F. (2009): An evaluation of nonresponse bias using paradata from a health survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Matsuo, H./Loosveldt, G. and Billiet, J. (2006): The history of the contact procedure and survey cooperation – Applying demographic methods to European Social Survey contact forms round 2 in Belgium. Louvain-la-Neuve, Belgium. Paper presented at the Quetelet Conference.

McGee, A. and Gray, M. (2007): Designing and using a behaviour code frame to assess multiple styles of survey items. Technical report, National Centre for Social Research (NatCen), London, UK.

Mockovak, W. and Powers, R. (2008): The use of paradata for evaluating interviewer training and performance. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Morganstein, D. and Marker, D. (1997): Continuous quality improvement in statistical agencies. In: Lyberg, L./Biemer, P./Collins, M./de Leeuw, E./Dippo, C./Schwarz, N. and Trewin, D. (Eds.): Survey Measurement and Process Quality. New York.

Murphy, W./O'Muircheartaigh, C./Emmons, C./Pedlow, S. and Harter, R. (2003): Optimizing call strategies in RDD: Differential nonresponse bias and costs in

REACH 2010. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Olson, K. (2006): Survey participation, nonresponse bias, measurement error bias, and total bias. Public Opinion Quarterly 70 (5), 737-758.
Olson, K. (2010): When do nonresponse follow-ups improve or reduce data quality? A synthesis of the existing literature. [Manuscript].
Olson, K. and Groves, R.M. (2009): The lifecycle of response propensities in fertility and family demography surveys. Paper presented at the annual meeting of the Population Association of America, Detroit, MI.
O'Reilly, J. (2009): Paradata and Blaise: A review of recent applications and research. Paper presented at the International Blaise Users Conference (IBUC), Latvia.
Peytchev, A. (2009): Survey breakoff. Public Opinion Quarterly 73 (1), 74-97.
Peytchev, A. and Peytcheva, E. (2007): Relationship between measurement error and unit nonresponse in household surveys: An approach in the absence of validation data. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Potthoff, R.,/Manton, K. and Woodbury, M. (1993): Correcting for nonavailability bias in surveys by weighting based on number of callbacks. Journal of the American Statistical Association 88 (424), 1197-1207.
Sala, E./Uhrig, S. and Lynn, P. (2008): The development and implementation of a coding scheme to analyse interview dynamics in the British Household Panel Survey. Technical report, Institute for Social and Economic Research ISER, University of Essex.
Sangster, R. and Meekins, B. (2004): Modeling the likelihood of interviews and refusals: Using call history data to improve efficiency of effort in a national RDD survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.
Scheuren, F. (2000): Macro and micro paradata for survey assessment. [Manuscript].
Scheuren, F. (2005): Paradata from concept to completion. In: Proceedings of the Statistics Canada Symposium. Methodological Challenges for Future Information Needs.
Schnell, R. (1998): Besuchs- und Berichtsverhalten der Interviewer. In: Statistisches Bundesamt (Ed.): Interviewereinsatz und –qualifikation. Stuttgart.
Schnell, R. (2005): Nonresponse in Bevölkerungsumfragen. Opladen. www.ub.uni-konstanz.de/kops/volltexte/2008/5614. [Last visited: 03/10/2010].
Schnell, R./Hill, P. and Esser, E. (1995): Methoden der empirischen Sozialforschung. 5th Edition [latest edition 2008]. Munic.
Schnell, R./Krause, J./Stempfhuber, M./Zwingenberger, A. and Hopt, O. (2008): Softwarewerkzeuge zur Dokumentation der Fragebogenentwicklung – QDDS – 2. Technical report, Endbericht des Projekts an die DFG.
Siegel, N. and Stimmel, S. (2007): SOEP-Interviewerbefragung 2006. Methodenbericht. Technical report, Munich.
Sinibaldi, J. (2010): Measurement error in objective and subjective interviewer observations. Paper to be presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Chicago, MI.
Smith, T. (1984): Estimating nonresponse bias with temporary refusals. Sociological Perspectives 27 (4), 473-489.

Smith, T. (2007): The Multi-level Integrated Database Approach (MIDA) for improving response rates: Adjusting for nonresponse error, and contextualizing analysis. Paper presented at the European Survey Research Association (ESRA) Conference, Prague, Czeck Republic.

Smith, T. (2009): The Multi-level Integrated Database Approach for detecting and adjusting for nonresponse bias. Paper presented at the European Survey Research Association (ESRA) Conference, Warsaw, Poland.

Stern, M. (2008): The use of client-side paradata in analyzing the effects of visual layout on changing responses in web surveys. Field Methods 20 (4), 377-398.

Stinchcombe, A./Jones, C. and Sheatsley, P. (1981): Nonresponse bias for attitude questions. Public Opinion Quarterly 45, 359-375.

Stokes, S. and Greenberg, B. (1990): A priority system to improve callback success in telephone surveys. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Stoop, I. (2005): The Hunt for the Last Respondent. The Hague.

Stoop, I./Devacht, S./Billiet, J./Loosveldt, G. and Philippens, M. (2003): The development of a uniform contact description form in the ESS. Paper presented at the International Workshop for Household Survey Nonresponse, Leuven, Belgium.

Thissen, R./Sattaluri, S./McFarlane, E. and Biemer, P.P. (2007): Evolution of audio recording in field surveys. Paper presented at The American Association for Public Opinion Research (AAPOR) 62th Annual Conference.

Triplett, T. (2002): What is gained from additional call attempts and refusal conversion and what are the cost implications? Technical report, The Urban Institute, Washington DC.

Voogt, R. and Saris, W. (2005): Mixed mode designs: Finding the balance between nonresponse bias and mode effects. Journal of Official Statistics 21, 367-387.

Wagner, J. (2009): Adaptive contact strategies in a telephone survey. In: Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.

Wagner, J. and Raghunathan, T. (2007): Bayesian approaches to sequential selection of survey design protocols. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Wang, K./Murphy, J./Baxter, R. and Aldworth, J. (2005): Are two feet in the door better than one? Using process data to examine interviewer effort and nonresponse bias. In: Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference.

Weeks, M./Kulka, R. and Pierson, S. (1987): Optimal call scheduling for a telephone survey. Public Opinion Quarterly 51, 540-549.

Weinhardt, M. and Kreuter, F. (2009): The different roles of interviewers: How does interviewer personality affect respondents' survey participation and response behavior?

West, B. (2010): An examination of the quality and utility of interviewer estimates of household characteristics in the National Survey of Family Growth. Paper to be presented at the Annual Conference of the American Association of Public Opinion and Research (AAPOR), Chicago, MI.

Yan, T. and Datta, R. (2009): Estimating the value of project-specific and respondent specific interviewer experience: Evidence from longitudinal and repeated cross-section surveys.

Yan, T. and Tourangeau, R. (2008): Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. Applied Cognitive Psychology 22 (1), 51-68.

Yan, T./Tourangeau, R. and Arens, Z. (2004): When less is more: Are reluctant respondents poor reporters? In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

Yu, M. and Yan, T. (2007): Are nonrespondents necessarily bad reporters? Using imputation techniques to investigate measurement error of nonrespondents in an alumni survey. In: Proceedings of the Section on Survey Research Methods of the American Statistical Association.

5. Record Linkage from a Technical Point of View

Rainer Schnell

Contact:

Rainer Schnell
University of Duisburg-Essen
Institute for Sociology
Lotharstr. 65
47057 Duisburg
Germany
e-Mail: rainer.schnell[at]uni-due.de

Abstract

Record linkage is used in preparing sampling frames, deduplicating lists, and combining information from two different databases on the same object. If the identifiers of the same objects in two different databases have error-free unique common identifiers like personal identification numbers (PID), record linkage is a simple file merge operation. If the identifiers contain errors, record linkage is a challenging task. In many applications, the numbers of observations in the files differ widely: a sample survey may contain a few thousand records while an administrative database of social security numbers may contain a few million. Available software, privacy issues, and future research topics are discussed.

Keywords: record linkage, data mining, privacy-preserving protocols

1. Introduction

Record linkage seeks to identify the same objects in two different databases using a set of common identifiers.[1] If the files have error-free unique common identifiers like personal identification numbers (PID), record linkage is a simple file merge operation. If the identifiers contain errors, record linkage is a challenging task. In many applications, the numbers of observations in the files differ widely: a sample survey may contain a few thousand records while an administrative database of social security numbers may contain a few million. Most research applications of record linkage use the linking process to prepare sampling frames, deduplicate lists, and combine information from two different databases on the same object.[2]

2. Current applications

Searching for the keyword "record linkage" will currently yield a few thousand papers on applications in medicine (mainly epidemiology), but only a few dozen papers in the social sciences. Nevertheless, record linkage is often used by social science research organizations as part of their fieldwork

1 The term "record linkage" is the one most commonly used by statisticians. In computer science, there are a variety of different terms for this process: "deduplication," "reconciliation," and "merge/purge processing."

2 Record linkage tries to identify the same objects in two databases. Do not confuse record linkage with statistical matching: statistical matching (or data fusion) tries to find records of very similar values for different objects; thereby deliberately joining data files with no common objects. For applications of statistical matching, see D'Orazio et al. (2006).

activities; in many such cases the client does not even know that a record linkage process has been used. In practice, constructing sampling frames often implies linking records from different databases referring to the same entities, such as names, addresses, birthdates, phone numbers, and geodata.[3] Record linkage is often used to combine information based on a survey with information from a database. This is often the case with business surveys, where information on the performance, size, and type of business are combined with business survey data through record linkage.[4]

Record linkage may be used to build panels after data collection, for example by using historical data as in the Victorian Panel Study (VPS). The VPS is intended as a longitudinal dataset based on the British censuses from 1851 to 1901 (Crockett et al.: 2006). Such linkages are possible in many cases, even without the use of unique personal identifiers. One such application is the Statistical Longitudinal Census Dataset (SLCD). The Australian Bureau of Statistics (ABS) will build the SLCD by linking a 5 percent sample of people from the 2006 population census to subsequent censuses. To minimize privacy problems, the ABS will link records without using names or addresses (Bishop and Khoo 2006). Record linkage is also an essential tool for conducting general censuses, and indeed is the most important tool used in registry-based censuses – like the German Census in 2011 – where record linkage is required to estimate coverage rates.[5] As a final example, in nonresponse research, linking the data of nonrespondents to administrative data files is one of the few methods of assessing nonresponse bias with empirical data.

3. Record linkage process

Record linkage is the process of linking two files that contain data on the same entity using common identifiers. This process follows a standard sequence (see figure 1). Usually, the identifiers must be standardized, which is called "pre-processing." Since the number of comparisons is generally too high to be computed directly, the computations are split up between disjunct subsets of observations (called "blocks") and repeated for different blocking criteria.[6] The similarity of records within a block is computed using similarity functions, most often today either with an edit-distance or Jaro-Winkler

3 Some examples for German surveys may be found in Schnell (2008).

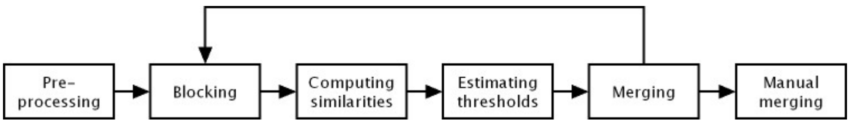
4 Details on such application can be found in a paper by Winkler (1995).

5 There is a rich literature on using record linkage for census undercount estimates, starting with Winkler/Thibaudeau (1991) and Ding/Feinberg (1996).

6 For example, in a cancer registry, persons living within an area with a common postal code are treated as a block.

string similarity function.[7] Then a decision has to be made on thresholds of similarity: records above a threshold are considered as a link; records below the threshold are considered as a non-link. Records between the thresholds are usually submitted for clerical review. The statistically most interesting part of the process is the decision on which pairs of elements in the two files to consider as true links. This decision may be made based on different computational models, for example, classification trees (CART), support vector classifiers (SVM), or statistical decision rules.[8] Most record linkage programs today use a probabilistic decision rule based on the model proposed by Fellegi/Sunter (1969). The parameters of the model are usually estimated by some variant of an EM algorithm (Herzog et al.: 2007). Special situations (for example, a known one-to-one correspondence between the two files) require modifications of the decision rules.

Figure 1: The linking process



4. Available software

There are many record linkage systems available. Most of these are special purpose programs for use in official statistics or cancer registries.[9] Furthermore, there are a few commercial programs for office applications. Of course, there also exist academic proof-of-concept implementations of special algorithms. In the following, the historically most important program and three contemporary programs in the public domain will be described in some detail.

7 Details on the computation and performance of string similarity functions can be found in Herzog et al. (2007) and Schnell et al. (2003).

8 Detail on SVMs and CART can be found in any textbook on statistical learning, for example, Bishop (2006).

9 A highly selective review from an official statistics point of view can be found in Herzog et al. (2007), which also includes a list of criteria that should be used in evaluations of record-linkage software.

4.1 Automatch

The most widely known probabilistic record linkage program is "Automatch." The last version (4.2) was released in 1992. Automatch is now a part of a large collection of programs (IBM's "WebSphere QualityStage") and cannot be licensed or purchased as a stand-alone program. The cost of the IBM Web-Sphere is far beyond the scope of research groups; therefore Automatch is no longer used in research contexts. Only a few cancer registries use the old DOS version of Automatch with special permission from IBM. Automatch is often used to validate the other programs. It should be noted that the limitations of old DOS programs have been evaded by some clever programming shortcuts; therefore Automatch is not a perfect baseline for comparisons.

4.2 Link Plus

Link Plus is primarily a probabilistic record linkage program for cancer registries. The program has been developed for the "National Program of Cancer Registries" (NPCR) of the Center for Disease Control and Prevention. It is a Windows-based program for detecting duplicates and linking cancer registry files with external files.[10] The program offers different similarity functions and phonetic encodings. Furthermore, it handles missing data and special cases like middle initials.[11]

4.3 Link King

"Link King" is an SAS-based probabilistic record linkage program developed by Kevin M. Campbell. The program requires a base SAS license. The program can work with SAS files, SPSS portable files, and CSV files. The most interesting features are nickname matching, gender imputation for 20,000 (American) names, and the calculation of distances between (American) zip codes.[12]

10 Since the development team wants to include the Microsoft.NET framework and Access databases, the binding of Link Plus to windows will be even closer in the future.

11 The program is available for no charge at http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm

12 The program is available for no charge at http://www.the-link-king.com

4.4 The Merge Toolbox: MTB

A project group of the author (funded by a research grant from the German Research Foundation) has developed a "merge toolbox" (MTB) for probabilistic record linkage (Schnell et al.: 2005). MTB is written in JAVA and is therefore highly portable to any modern computer system. The program consists of a preprocessing module, a linkage module, and a manual editing module. The program can read and write STATA and CSV files, computes nearly all known string similarity functions, and can perform deterministic and probabilistic record linkage. MTB is being used by cancer registries and research groups in epidemiology, sociology, and economics in Germany.[13]

4.5 Empirical comparisons of programs

Since most record-linkage programs for probabilistic linkage use the same algorithms for making link decisions, the programs should yield very similar results, given the same input. Since the programs differ in preprocessing, some studies compare different parts of the linkage process. Only identically preprocessed data files should be used for linking; but this is often of no practical relevance. For practical applications, the complete linkage process between optimally tuned programs should be compared: this is no small task, and as a result, such studies are rare (Campbell et al.: 2008). From a theoretical point of view, it would be interesting to compare different programs using different decision rules (for example, CART, SVM, and Fellegi-Sunter) on non-preprocessed data and identically preprocessed data. However, systematic studies of this kind are still lacking. For the future, it seems more promising to work on an optimized combination or sequence of decision rules after extensive standardization and preprocessing than to make naive empirical comparisons.

13 A restricted version of the program is available for no charge at http://www.uni-konstanz.de/FuF/Verwiss/Schnell/mtb. For scientific purposes, the full program is available for no charge by writing to the author.

5. Privacy issues

Record linkage may be misused for de-anonymization of scientific research files. This possibility of misuse is simply due to the fact that the programs try to minimize distances between objects in a high-dimensional space. Therefore, de-anonymization by minimizing distances can be done by every program for cluster analysis.[14] This misuse is therefore not specific to record-linkage programs.

The result of a successful record linkage is a dataset C with more known characteristics of the objects than in the original data files A and B. Using this enhanced data file C to compare these characteristics with another data file D makes identification of objects in D much more likely than identification using A or B alone, since the number of observations with a given combination of characteristics is declining with every added variable.[15] The risk of disclosure is therefore higher after record linkage. It might be necessary to use additional standard disclosurerisk measures for the enhanced data file C.[16]

6. Research perspectives

From a statistical perspective, the theoretical problems of record linkage are well defined and some interesting solutions have been found. Many applied researchers consider record linkage a simple task. In practice, it is not. In fact, the lay user is often disappointed with the performance of record linkage programs.[17] The main reason for this poor performance is usually the quality of the input data: if many identifiers are missing or poorly standardized, any automatic method will fail. Therefore, we need more work on preprocessing of identifiers. Since preprocessing depends on language- and country-specific details, programs and algorithms must be fine-tuned with local datasets and expert systems. Experts from the fields of statistics and computer science need to use real data from actual data-generating processes.

14 For an application, see Torra et al. (2006).

15 This can be seen as a direct consequence of the definition of k-anonymity: in a k-anonymized dataset, each record is indistinguishable from at least (k-1) other records.

16 Examples of such techniques can be found in Willenborg/de Waal (1996) and Domingo-Ferrer (2002); for record linkage and privacy issues in general, see United States General Accounting Office (2001).

17 For example, Gomatam et al. (2002) note higher sensitivity and a higher match rate but a lower positive predicted value of Automatch in comparison to a stepwise deterministic strategy. These results could be changed easily by changing the matching parameters and the preprocessing.

6.1 Real-world test datasets

Interestingly, a standard dataset for comparing record linkage procedures has not been published. Instead, some research groups build data generators with specified error-generating mechanisms. Since such error structures may be different from those of real-world applications, a collection of test datasets based on real world data would be highly desirable. Since the details of name conventions, addresses, postal codes, etc. differ between countries and databases, a German reference database is needed.

6.2 Expert systems and key standardization

Database fields contain many different ways of storing information of key values used for record linkage. These fields must be standardized based on expertise with the distinctive features of German addresses, phone numbers (land lines and mobile), name conventions (for example, historical rules for name changes after marriage), academic titles, historical hereditary titles, legal business forms, etc. Compiling such lists and generating transformation rules is a tedious and labor-intensive task. Currently, the huge amount of work required to generate such exhaustive lists and standardization rules is only done by private companies.[18] Of course, the cumulated commercial knowledge bases are not available for academic use. Therefore, German official statistics will have to buy such standardization services for large-scale operations like the 2011 Census on the commercial market with obvious consequences. In the long run, statistical offices, cancer registries, and other publicly funded research organizations will need common knowledge bases for key standardization.

6.3 Reference databases

For practical record linkage, several reference databases are needed that are currently not publicly available for research purposes. At present, simple lists of all German municipalities with old and new German zip codes, correspondence lists of zip codes and phone numbers, regional identifiers like city codes (*Gemeindekennziffer*), Gauss-Krüger coordinates, and street addresses are not available for public use. Every record linkage group has to compile

18 The unit on "Postal Automation" of Siemens I&S (Constance) employs more mathematicians and computer scientists for producing such expert systems than all German cancer registries together. Given the published lists of customers of other companies in the same sector in Germany (for example, "Fuzzy Informatik," a spin-off of Daimler), it is safe to assume that currently more than 50 experts in Germany are working on such standardization tasks.

its own rough version of these reference lists. Since some of these lists are quite expensive, there should be a common scientific license for this data.[19] Furthermore, frequency tables of names and surnames conditioned on gender, nationality, and year of birth would be very useful for imputing gender, nationality, and age based on a given name. Other databases can be used for the same purpose, for example, gender can be imputed with certain ICD or ISCO codes. This imputed information can be used for record linkage with incomplete keys.

6.4 Candidate generation

One interesting idea that has not been studied in detail so far is the generation of candidates for matching based on a search string. The candidates can be generated by introducing random errors or according to pre-specified rules (Arusu et al.: 2008). The resulting candidates will be compared to the existing identifiers. This step should follow unsuccessful standard linkage attempts.

6.5 Blocking

Data files for record linkage are usually quite large. In many applications, we have a small file (for example, a survey) with about 1000 observations and an administrative database with, for example, 10 million records. This would result in 10^{10} comparisons, taking 278 hours at 10,000 comparisons per second. Using standard hardware and standard programs, this is unacceptable. The computation time is usually reduced by using a simple idea: compute the similarity matrix only within subgroups. These subgroups are called "blocks" and the strategy is called "blocking." For example, instead of comparing all company names in Germany with one another, we compare all pairs of company names within each city. Using a suitable blocking variable reduces the computing time for one typical record linkage run (10,000 observations linked to a five million record database) to less than a hour. Of course, this speed comes at a price. The variable used for blocking must be considered a perfect classification variable: exhaustive and disjunct- and error-free. Since blocking variables are in many cases proxy variables of geographical identifiers like dial prefixes, postal codes, or administrative units, there is no guarantee for error-free perfect classification of units. Currently, there is a great deal of research activity in computer science on modifications of blocking algorithms to im-

19 For example, a list of all the geo-coordinates of all German buildings, which would be useful for many research purposes in record linkage and epidemiology, is a considerable expense, amounting to about the cost of one research assistant per year.

prove on simple blocking schemes (for example, “adaptive blocking,” Bilenko et al.: 2006). These new blocking techniques still have yet to be implemented in production software for record linkage.

6.6 Algorithms for large similarity matrices

As an alternative to blocking, algorithms for computing approximate similarity matrices could be used. Such algorithms have been proposed in the technical literature, for example, “Sparsemap” (Hristescu and Farach-Colton: 1999), “Boostmap” (Athitsos et al.: 2004) and “WEBSOM” (Lagus et al.: 2004). Another interesting approximation was recently suggested by Brandes/Pich (2007). None of these techniques has been systematically used for record linkage up to now. Special data structures or algorithms used for high-dimensional indexing (Yu 2002) have rarely been applied to large-scale record linkage projects.

6.7 Special hardware

Since the blocking of datasets reduces the task of computing an $n*n$ similarity matrix to the independent computation of k matrices of size $m*m$, the computation can be done by several independent machines or processors. This is a very simple version of a parallel computing process, which requires only a minor modification of existing programs. Of course, parallel searching of similarity index structures by special algorithms (Zezula et al: 2006: Chapter 5) or the separate standardization of each record may also be done with such hardware. However, the resulting program can be run on the shelf hardware like standard PC boards. Since such a system should be portable, a compact server rack can be used. Currently available server boards house four processors with four cores each, so a special machine with 64 cores can be built by using only four server boards. In order to reduce power consumption, smaller mobile processor boards may be used instead, requiring eight boards with two quad-core mobile processors. Such a system will drain less than 1000 watts in total, so it does not require special cooling or power supply. The machine should be equipped with at least 1 Gbyte RAM for each processor. In order to minimize the risk of data leaking, the machine can be built as a diskless server: it needs no hard disk at all, since the operating system can be booted from a memory stick and the data to be processed may be kept on removable memory sticks.[20] The sticks should be destroyed after reading; the linked data file should be saved to an empty new stick. In slightly less de-

20 Even a data file with 30 million records and 100 bytes of ID-information per record fits on a 10 EUR 4 Gbyte USB stick.

manding computer security environments, the input files may be copied to the machine by using VPN. Such a portable secure special purpose record-linkage machine can be built for the price of three small enterprise servers. It would be highly desirable to have at least one such machine within a trusted computing center with restricted access, for example, within one of the Research Data Centers.

6.8 Privacy-preserving record linkage

In most practical applications, record linkage has to be done with the standard keys: name, surname, gender, date of birth, and place of birth. Since people hesitate to use such identifiers, in many applications encrypted keys have to be used. Since the input data for encryption is prone to errors, a slight deviation between the keys of a true link pair is probable. Such slight deviations result in keys that cannot be matched, since similarity distances between encrypted keys are pointless. Therefore, privacy-preserving record linkage requires special algorithms. Starting with the publication by Churches/Christen (2004), some protocols for record linkage with encrypted alphanumeric keys with errors have been suggested (Pang and Hansen: 2006; Scannapieco et al.: 2007). Independent comparisons of these protocols have not been published and are badly needed. All protocols seem to be awkward to implement with mistrustful database owners. To overcome these problems, we have developed a new protocol that has proven to be fast and reliable (Schnell et al.: 2009). We are currently testing the protocol on different simulated datasets. A complete record linkage solution for encrypted keys must include a protocol for computing distances between encrypted metric data. One very interesting protocol has been proposed by Inan et al. (2006). A highly secure record linkage program for error prone numeric and alphanumeric keys will require a few years of testing and programming. This seems to be the most important research task still to be carried out before record linkage can be used widely given the increasing privacy concerns in western populations.

7. Three recommendations

7.1 Training datasets and reference datasets

In order to improve the performance of record-linkage programs and algorithms, large training and reference datasets should be produced. These should be real-life datasets containing only linkage variables. The links have to be established with a common error-free key or through careful clerical work. Simulated datasets are no substitute for such datasets. Therefore, privacy concerns must be addressed by standard procedures of statistical disclosure control.

7.2 Research program on preprocessing and privacy-preserving record linkage

We need a European research program on preprocessing keys for privacy preserving record linkage. Such a research program should be multinational, since European countries differ in ethnic composition and therefore in the distribution of ethnic surnames. Furthermore, the legal situation of record linkage differs widely within Europe. A multinational and multi-disciplinary research group of computer scientists, lawyers, linguists, historians, and social scientists is therefore needed to solve the problems of privacy-preserving record linkage using standard identifiers like names and surnames.

7.3 National Record Linkage Center

We currently have no research centers for record linkage in Germany, only the cancer registries, which perform a very limited kind of record linkage for a single purpose. Every research team in criminology, sociology, medicine and economics must build its own record linkage infrastructure. In many cases, the cost of doing so exceeds the available research funds. Therefore, at least one National Record Linkage Center is needed. This center should have special machines (massive parallel processors), a team trained in record linkage, and the data protection facilities necessary to act as a data trustee for large-scale projects.

References:

Arasu, A./Chaudhuri, S. and Kaushik, R. (2008): Transformation-based Framework for Record Matching, International Conference on Data Engineering.

Athitsos, V./Alon, J./Sclaroff, S. and Kollios, G. (2004): BoostMap: a method for efficient approximate similarity rankings. CVPR, 268-275.

Bilenko, M./Kamath, B. and Mooney, R.J. (2006): Adaptive Blocking: Learning to Scale Up Record Linkage, ICDM06: Sixth International Conference on Data Mining, 87-96.

Bishop, C.M. (2006): Pattern Recognition and Machine Learning. Berlin.

Bishop, G. and Khoo, J. (2006): Methodology of Evaluating the Quality of Probabilistic Linking, Proceedings of Statistics Canada Symposium 2006. www.statcan.ca/english/freepub/11-522-XIE/2006001/article/10401-en.pdf. [Last visited: 02/10/2010].

Brandes, U. and Pich, C. (2007): Eigensolver Methods for Progressive Multidimensional Scaling of Large Data. Proc. 14th Intl. Symp. Graph Drawing (GD '06). LNCS 4372, 42-53.

Campbell, K.M./Deck, D. and Krupski, A. (2008): Record linkage software in the public domain: a comparison of Link Plus, The Link King, and a 'basic' deterministic algorithm. Health Informatics Journal 14, 1-15.

Churches, T. and Christen, P. (2004): Some methods for blindfolded record linkage. BMC Medical Informatics and Decision Making 4, 1-17.

Crockett, A./Jones, C.E. and Schürer, K. (2006): The Victorian Panel Study. University of Essex. [Unpublished manuscript].

Ding, Y. and Feinberg, S.E. (1996): Multiple sample estimation of population and census undercount in the presence of matching errors. Survey Methodology 22, 55-64.

Domingo-Ferrer, J. (Ed.) (2002): Inference control in statistical databases: from theory to practice. Lecture notes in computer science Vol. 2316. Berlin.

D'Orazio, M./DiZio, M. and Scanu, M. (2006): Statistical Matching: Theory and Practice. New York.

Elmagarmid, I.A./Ipeirotis, P.G. and Verykios, V. (2006): Duplicate Record Detection: A Survey. IEEE Transactions on Knowledge and Data Engineering, 19/2006, 1-16.

Fellegi, I.P. and Sunter, A.B. (1969): A Theory for Record Linkage. Journal of the American Statistical Association 64.

Gomatam, S./Carter, R./Ariet, M. and Mitchell, G. (2002): An empirical comparison of record linkage procedures. Statistics in Medicine 21 (10), 1485-1496.

CMIS Technical Report No. 03/83, CSIRO Mathematical and Information Sciences.

Herzog, T.N./Scheuren, J.J. and Winkler, W.E. (2007): Data Quality and Record Linkage Techniques. New York/Berlin.

Hristescu, G. and Farach-Colton, M. (1999): Cluster-preserving embedding of proteins. Technical report. Rutgers University.

Inan, A./Saygin, Y./Savas, E./Hintoglu, A.A. and Levi, A. (2006): Privacy Preserving Clustering on Horizontally Partitioned Data. Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW'06).

Lagus, K./Kaski, S. and Kohonen, T. (2004): Mining massive document collections by the WEBSOM method. Information Sciences 163 (1-3), 135-156.

Pang, C. and Hansen, D. (2006): Improved record linkage for encrypted identifying data. HIC 2006 and HINZ 2006 Proceedings. Brunswick, 164-168.

Scannapieco, M./Figotin, I./Bertino, E. and Elmagarmid, A.K. (2007): Privacy preserving schema and data matching, Proceedings of the 2007 ACM SIGMOD international conference on Management of data, 653-664.

Schnell, R./Bachteler, T. and Bender, S. (2003): Record linkage using error prone strings. American Statistical Association, Proceedings of the Joint Statistical Meetings, 3713-3717.

Schnell, R./Bachteler, T. and Reiher, J. (2005): MTB: Ein Record-Linkage-Programm für die empirische Sozialforschung. ZA-Information 56, 93-103.

Schnell, R. (2008): Avoiding Problems of Traditional Sampling Strategies for Household Surveys in Germany: Some New Suggestions. Discussion paper for the GSOEP. www.diw.de/documents/publikationen/73/86107/diw_datadoc_2008 -033.pdf. [Last visited: 02/10/2010].

Schnell, R./Bachteler, T. and Reiher, J. (2009): Privacy-preserving record linkage using Bloom filters. BMC Medical Informatics and Decision Making 2009, 9-41.

Torra, V./Abowd, J.M. and Domingo-Ferrer, J. (2006): Using Mahalanobis Distance-Based Record Linkage for Disclosure Risk Assessment. Privacy in Statistical Databases, 233-242.

United States General Accounting Office (GAO) (2001): Record Linkage and Privacy. Issues in Creating New Federal Research and Statistical Information, April 2001, GAO-01-126SP. Washington.

Willenborg, L. and de Waal, T. (1996): Statistical Disclosure Control in Practice. New York.

Winkler, W.E. (1995): Matching and Record Linkage. In: Cox, B. et al. (Eds.): Business Survey Methods. New York, 355-384

Winkler, W.E. and Thibaudeau, Y. (1991): An application of the Fellegi-Sunter model of record linkage to the 1990 U.S. decennial census technical report, US Bureau of the Census. www.census.gov/srd/papers/pdf/rr91-9.pdf. [Last visited: 02/10/2010].

Yu, C. (2002): High-Dimensional Indexing. Transformational Approaches to High-Dimensional Range and Similarity Searches. Berlin.

Zezula, P./Amato, G./Dohnal, V. and Batko, M. (2006): Similarity Search. The Metric Space Approach. New York.

6. Experiments, Surveys and the Use of Representative Samples as Reference Data

Thomas Siedler and Bettina Sonnenberg

Contact:

Thomas Siedler
Bettina Sonnenberg
DIW Berlin
Mohrenstr. 58
10117 Berlin
Germany
e-mail: tsiedler[at]diw.de
bsonnenberg[at]diw.de

Abstract

During the last two decades, laboratory experiments have come into increasing prominence and constitute a popular method of research to examine behavioral outcomes and social preferences. However, it has been debated whether results from these experiments can be extrapolated to the real world and whether, for example, sample selection into the experiment might constitute a major shortcoming of this methodology. This note discusses potential benefits of combining experimental methods and representative datasets as a means to overcome some of the limitations of lab experiments. We also outline how large representative surveys can serve as reference data for researchers collecting their own datasets in order to explore potential sample selection biases.

Keywords: experiments, survey, representativity
JEL-Classification: C01, C52, C8, C9, D0, D6, D81, D84

1. Introduction

During the last two decades, laboratory experiments have come into increasing prominence and now constitute a popular method of research to examine behavioral outcomes and social preferences. There are obvious advantages of laboratory experiments. First, researchers can control the environment under which individuals make their decisions and allow causal inferences by exogenously varying one parameter while holding all others constant. Second, the simplicity of many such experiments makes it easy to explain the findings to non-academics and policy-makers. However, major limitations of most experiments are that they are administered to students, who usually self-select themselves into the study and are therefore not representative of the entire adult population. In fact, due to self-selection, experimental studies with student subjects might not even be representative of the entire student population. For example, Eckel and Grossman (2000) investigate the impact of recruitment methods on behavior in a series of dictator experiments with a charitable organization as a recipient in laboratory sessions. The authors compare altruistic behavior among student subjects recruited voluntarily through announcements in graduate and undergraduate courses ("voluntary sample") with students in which the experiment was conducted during class time ("pseudo-voluntary sample"). They find that pseudo-volunteers are significantly more generous on average than their volunteer counterparts, and that socio-economic characteristics such as religion or survey measures of altruistic preferences have a larger effect on giving behavior among students recruited pseudo-voluntarily. Similarly, Harrison et al. (2007) examine potential self-selection bias in both a field experiment and a laboratory experiment with student subjects. The authors start with the observation that samples observed in the experiment might suffer from randomization bias (Heckman and Smith 1995). Being interested in individuals' risk attitudes, the authors note that the likelihood to participate in the experiment might be higher for individuals with on average higher risk attitudes than among the general population. On the other hand, the researchers offer participants a fixed show-up fee that might encourage individuals that are more risk-averse to participate in the experiment, potentially outweighing sample selection into the experiment in their study due to randomization bias. The authors report significant self-selection into both the field experiment and the laboratory experiment with adult subjects drawn from the general Dutch population, arguing that their sample is on average more risk-averse than the general population (see also Roe et al. 2009). In addition, most laboratory experiments are conducted on very homogenous samples (typically students studying the same subject at the same university) and often information on potentially important socio-economic background characteristics is missing or lacks sufficient variation. Another shortcoming of laboratory experiments is the lack of anonymity. In most laboratory studies, students play against each other and know that the other player is a student. Hence, the degree of anonymity is rather low. Both the degree of homogeneity and anonymity in the subject pool might influence revealed social preferences (Sapienza et al. 2007). The question has also been raised whether laboratory experiments are externally valid and to what extent laboratory findings can be extrapolated to the general population (Levitt and List 2007). A branch of the recent literature examines the external validity of laboratory experiments by comparing behavior in laboratory sessions with experimental outcomes in more heterogeneous and representative samples (Bosch-Domenech et al. 2002; Haigh and List 2005; Benz and Meier 2006). The majority of these studies report that the behavior in the lab differs from that observed in other contexts. For a detailed discussion of potential limitations of laboratory experiments measuring social preferences, see Levitt and List (2007). For a recent discussion regarding potential improvements and future challenges in the field of experimental economics, see Gächter (2009).

Another strand of research in economics and the social sciences makes use of survey questions from large representative cross-sectional or household panel datasets. One criticism of using attitudinal questions from these surveys concerns the lack of behavioral underpinnings and the absence of meaningful survey questions in certain contexts. For example, Glaeser et al. (2000) and Ermisch et al. (2009) discuss the difficulties of measuring respondents' trustworthiness by means of survey questions. Combining attitudinal survey questions with behavioral experiments that include monetary rewards can potentially provide a fuller understanding of economic behavior and help to overcome some of these shortcomings. This note briefly discusses potential benefits of combining experimental methods and representative datasets when studying economic outcomes and social behavior. We also provide a short overview about the recent literature combining the experimental approach with survey methods. Finally, we discuss potential benefits of using large representative surveys as reference data for researchers collecting their own datasets. An overview of recent selected studies combining behavioral experiments with survey questions or using representative surveys as reference datasets is provided in table 1.

2. Combining behavioral experiments and survey methods

2.1 Trust and trustworthiness

A new research strand combines behavioral experiments and survey methods. Fehr et al. (2002) incorporate the standard trust-game experiment (Berg et al. 1995) into a representative survey of the German population and asked respondents several survey measures of trust. Fehr et al. (2002) find a positive association between attitudinal survey measures of trust and sender's behavior, but no significant correlation between survey-based measures of trust and trustworthiness in the experiment. In addition, the authors report that individuals aged 65 and above, highly skilled workers, and those living in larger households exhibit less trusting behavior in the experiment.

Using nationally representative data for Germany, Naef and Schupp (2009) compare survey and behavioral measures of trust. The authors create a new survey measure of trust and find that it is significantly correlated with the experimental trust measure. Moreover, they report that their experimental measure of trust is not subject to a social desirability bias and is robust to variations in stakes and the use of strategy method. This study demonstrates how survey measures can be tested by combining the experimental approach with survey methods.

In a representative sample of the Dutch population, Bellemare and Kröger (2007) measure levels of trust and trustworthiness elicited through an experiment similar to those presented by Berg et al. (1995) in a representative sample of the Dutch population. The authors also compare their representative trust experiment with a sample of college students in an equivalent laboratory experiment. They find that college students have considerably lower levels of trust and trustworthiness than individuals in the representative sample and that these differences can be explained mainly by differences in socio-economic and background characteristics, in particular age, gender, and education. For example, the authors find that women have higher levels of trust than men, but display lower levels of trustworthiness. In line with

Fehr et al. (2002), Bellemare and Kröger (2007) find a positive, inverted U-shaped association between age and trust. The authors do not find evidence of a participation bias in their trust experiment with student subjects, and therefore argue that trust and trustworthiness as measured in the laboratory are informative about the behavior in the general population.

Ermisch et al. (2009) integrate a new experimental trust design into a sample of the British population. The authors' rationale for using an alternative trust design is based on observations that the sender's behavior in the standard trust-game experiment (Berg et al. 1995) is not only influenced by trust but also depends on other motivations such as sender's reciprocity, risk aversion, altruism, or inequality aversion (Cox 2004; Karlan 2005; Ermisch and Gambetta 2006 and Sapienza et al. 2007). In their "one-shot" trust experiment, the sender faces the decision as to whether or not to pass on a fixed amount of money (e.g., whether or not to send £10. If £10 are sent, the experimenter increases it by £30 so that the second person receives £40) and the receiver must decide whether or not to pay back a fixed amount of money (e.g., the sender has the choice of either paying back £22 or keeping all £40). Thus, the players cannot choose whether or not to transfer a certain amount of money between, say, £1-£10; rather they face the decision whether to transfer the entire amount or nothing. The authors argue that this binary trust game is more likely to measure revealed trust and trustworthiness than the standard trust game experiment, in which the possibility of sending "any amount favours the intrusion of other motives such as 'gift giving', 'let's risk part of it' and 'I like to gamble'." Ermisch et al. (2009) find that the experiment is more likely to reveal trust if people are older, if they are home-owners, if their financial situation is "comfortable," or if they are divorced or separated. Trustworthiness is lower if a person's financial situation is perceived by them as difficult or as "just getting by."

2.2 Risk attitudes

Another recent example demonstrating the benefits of combining incentive-compatible experimental measures with survey methods is the study by Dohmen et al. (2009). In a previous related study, Dohmen et al. (2007) examine the relationship between individual's risk aversion, impatience, and cognitive abilities. They find that lower cognitive abilities are significantly associated with greater risk aversion and more pronounced impatience. These relationships are found to be robust to controlling for a broad set of socio-economic characteristics, such as age, gender, education, and income, which are measured through standard survey questions. In their study, both risk aversion and impatience are measured by choice experiments that involve real monetary choices and relatively large stakes. Respondents were told in advance that the experiment was about financial decisions, that they would have the chance to win money, and that the earned amount would depend on their choices in the experiment. Subjects were also informed that every seventh participant would win. For instance, in the lottery experiment, a financial decision is represented by the choice between a certain payoff (*Option A*) and a risky lottery (*Option B*). Participants were also informed that, for each paired lottery, *Option B* always implies a 50 percent chance of winning €300 and a 50 percent chance of winning nothing. The experiment starts with the following lottery choice: respondents can choose between a certain payoff of €0 (*Option A)* and *Option B*. If participants choose *Option B*, the amount of *Option A* is increased by €10 in the next decision round. Thus, the second lottery choice is between the "safe" payoff of €10 and Option *B*. Similarly, conditional on prior decisions, a third lottery choice is between a certain payoff of €20 and *Option B*. The experiment ends when subjects choose *Option A* for the first time, or when the maximum amount of €190 for *Option A* is reached. This study is another example demonstrating the potential benefits of combining experimental and survey measures in a representative sample of the population.

3. Using representative surveys as reference data

In this section, we briefly discuss potential benefits of using large representative surveys as reference datasets for researchers collecting their own data. Household panels might offer a useful reference point for experimental studies, thanks to their longitudinal character and the sampling of all household members – for example, the British Household Panel Study (BHPS), the new household panel study *Understanding Society* in the United Kingdom, and the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*). Register data can constitute another fruitful source of reference data (Harrison et al. 2007). The basic idea here is that large representative surveys can serve as reference data for researchers collecting datasets that do not represent the full universe of the population of interest (e.g., through clinical trials, intervention studies, laboratory and behavioral experiments, and cohort studies). An important issue when investigators collect their own data is whether the sample represents the general population, or conversely, whether it is selective (for example, by design or through choice-based sampling). This approach might offer several benefits. First, by asking participants similar questions to those in representative surveys, researchers can compare their sample with either a sub-sample or the whole representative survey. Second, in contrast to many of the scales and questionnaire instruments developed by psychologists, for instance, questions in household panel surveys like the SOEP or BHPS are not copyrighted and can be used by other researchers free of charge. Thus, these datasets can be a valuable point of reference for designing new questionnaires. Combining experimental sessions with a questionnaire collecting basic individuals' socio-demographic characteristics used in representative surveys gives researchers valid information as to the representativeness of their sample with respect to the individual characteristics surveyed.

Two recent studies exemplify the potential for using questions from a panel survey when researchers collect their own data. In Germany, Geyer et al. (2009) examine whether individuals aged 17-45 with operated congenital heart disease have adverse employment chances compared to people without heart problems. The authors compare their sample of patients (N=314; treatment group) with a sample drawn from the SOEP, which serves as a comparison group. The treatment group consisted of women and men who had a congenital heart disease and were operated on at the University Hospital of Göttingen. The authors conducted a face-to-face interview with patients using several SOEP questions. Comparing their hospital sample with the SOEP as reference data they found considerable differences between the two samples with respect to gender, age, and employment status.

Two recent projects that also follow the idea of using a representative household panel study (SOEP) as reference data are the Berlin Aging Study II and the Brain Gene Behavior Project. The Berlin Aging Study II, collecting data on objective socio-economic and biological characteristics like objective health, functional capacity, subjective health, and well-being, draws on SOEP questions with regard to health and life satisfaction to enable comparisons with the SOEP data (Max Planck Institute for Human Development 2009). Likewise, the Brain Gene Behavior Project, a large-scale study on the molecular genetic basis for personality, cognitive, and individual behavioral differences, makes use of the SOEP questionnaire to exploit comparable reference data (Neuroeconomics Lab Bonn and Socio-Economic Panel 2009).

In the United Kingdom, the study by Ermisch et al. (2009) demonstrates how a panel survey can help in determining the extent to which a particular sample is representative of the general population. The authors integrate a new experimental trust design into a former sample of the British population and compare their trust sample with a sample from the BHPS. By using a questionnaire similar to the BHPS, the authors are able to determine that their trust sample over-represents women, people who are retired, older, divorced, or separated. Together, these two studies show that household panel studies can serve as useful reference data for researchers collecting their own samples and can help to reveal the representativeness of their own collected data.

4. Conclusion

The studies reviewed demonstrate that enormous academic benefits can be derived from combining experimental studies with representative surveys.[1] First, experiments based on representative samples help to assess potential biases of studies based on student subjects who self-select themselves into the sample. This advances our knowledge on whether and to what extent experimental studies on student samples can be generalized. Second, research measuring both revealed preferences and stated preferences allows researchers to validate their measures. For example, Fehr et al. (2002), Ermisch et al. (2009), and Naef and Schupp (2009) report that answers to attitudinal questions on trust toward strangers do predict real trusting behavior in the experiment.

The recent studies by Eckel and Grossman (2000) and Roe et al. (2009) demonstrate the importance of self-selection into experimental studies, and their studies suggest that results from laboratory experiments might not be generalized to the entire population. In this note, we briefly discussed potential benefits of using large representative survey as reference data for researchers who are collecting their own datasets and point readers to two recent examples in the literature.

1 See also Falk et al. (2009).

Table 1: Studies Combining an Experimental Design with Survey Methods

Author(s)	Topic	Method
Bellemare and Kröger 2007	Measure levels of trust and trustworthiness elicited through an experiment similar to those presented by Berg et al. (1995) in a representative sample of the Dutch population.	Trust and trustworthiness measured by an invest- and-reward experiment.
Benz and Meier 2006	Explore the correlation between individual behavior in laboratory experiments and in a similar situation in the field.	Donation lab experiments with college students.
Dohmen et al. 2009	Investigate the relevance of survey questions on risk-taking behavior in field experiments and actual behavior in the real world.	Risk-taking measured by a lottery game in a field experiment and SOEP survey questions with a representative sample of 450 participants.
Eckel and Grossman 2000	Compare the effect of recruitment method in dictator experiments with student subjects.	Altruism measured by means of dictator games.
Ermisch et al. 2009	Measure trust and trustworthiness in Great Britain using an experimental and survey design.	One-shot trust experiment with former respondents of the BHPS in combination with survey questionnaires.
Fehr et al. 2002	Investigate trust and trustworthiness by comparing behavioral experimental outcomes and representative survey data.	Implementation of a trust experiment in a representative survey of the German population in 2002.
Gächter et al. 2004	The authors present survey and experimental evidence on trust and voluntary cooperation in Russia using both a student and a non-student sample.	One-shot public goods experiment.
Geyer et al. 2009	Examine the effect of congenital heart disease on employment status.	Sample of 628 patients surveyed in clinic combined with medical check-up (treatment group).

Data	Finding
Representative sample of the Dutch population and a laboratory sample with college students.	The smaller amount of students' investments predominantly demonstrates differences in socio-economic and background characteristics. While these characteristics can explain different revealed behavior, they have almost no impact on stated trust. Return ratios are significantly lower in the lab sample as well.
Secret use of the real donation spending behavior of the students.	The authors find a rather moderate or weak relationship between lab and field behavior.
Comparison with representative data of the whole SOEP sample on seven different survey questions with regard to risk attitudes.	The general risk attitude survey questions are significantly correlated with behavior in the lottery game as well as with actual behavior in the real world, e.g., with regard to financial, sports, and health-related behavior. Simultaneously, specific behavior is best predicted by context-specific risk survey measures in the respective domain.
Laboratory experiment with self-recruited students (voluntary sample) and in classroom recruited college students during the class period (pseudo-voluntary sample).	Volunteers are less generous in distributing endowments and are more motivated by incentives than classroom-recruited students. Respondents' characteristics such as sex, religion, and altruism influence the behavior of pseudo-volunteers more than that of volunteers. The authors conclude that self-selection into the sample matters.
Comparison with representative BHPS sample allows the authors to examine whether their experimental sample is representative of the general population.	For example, the authors report that their experimental sample over-represents women, people who are retired, divorced, or separated. Individual behavior in experiments is found to be a reliable and superior measure compared to standard common trust survey questions.
–	Trust in strangers and past trusting behavior correlate with trust behavior in the experiment, but no survey measure predicts trustworthiness.
Not fully representative survey data of Russian non-students and a student subject pool.	Non-students display higher levels of trust than students, and also contribute more to the public good as long as socio-economic background is not controlled for. Individuals who believe that most other people are fair contribute significantly more to the public good game than those without such beliefs. Likewise, optimists make higher contributions than pessimists.
Their comparison group is a 10 percent sample drawn from the German SOEP.	The authors find significant differences between male patients and male control subjects. Those with congenital heart disease are less likely to be employed full-time, more likely to be employed part-time, and in marginal employment. The differences between treatment and control group depend on the severity of the disease.

Author(s)	Topic	Method
Glaeser et al. 2000	Examine the validity of trust survey questions with a behavioral trust experiment.	Laboratory experiment with Harvard undergraduates.
Harrison et al. 2007	Investigate whether experiment samples are biased because of the risk of randomization. The authors undertake both a laboratory experiment and a field experiment to examine whether selection into the experiment influences measures of risk attitudes.	Eliciting individual risk attitudes through an experimental lottery game in both a field experiment and a laboratory experiment.
Levitt and List 2007	Discuss whether estimates on pro-social behavior from laboratory experiments can be extrapolated to the real world.	Literature review.
Naef and Schupp 2009	Test the correlation and validity of trust survey questions with experimental measures of trust.	Trust experiment with survey respondents, representative for Germany.

Data	Finding
Survey measures on trust (self-reported attitudes and behavior) of 258 Harvard undergraduates.	Ten out of 12 GSS trust questions do not predict trust, but are related to trustworthiness as measured in the experiment. Trust in the experiment is associated with past trusting behavior. Trust and trustworthiness rise with closer social distance.
First, the authors collect information on subjects' socio-economic characteristics by means of questionnaires and use this information to correct for potential self-selection into the field experiment. Second, in their laboratory experiment, they investigate the impact of variation in recruitment information on individual risk attitudes.	The authors find that the use of show-up fees generates a more risk-averse sample. Participants in both the field and laboratory experiment are found to be more risk-averse than the general population once they control for selection into the experiment.
	The authors argue that pro-social behavior in experiments depends on a number of experimental situation and design factors, e.g., stakes, sample recruitment, anonymity, as well as unobserved respondents' characteristics. They caution against generalizing results from laboratory to real-world situations.
Self-reported trust and trustworthiness by different measurements with a representative survey sample.	GSS Survey question do not measure trust in the experiment. However, the authors find a significant correlation between self-reported SOEP trust measures and experimental measures of trust. Students are found to be slightly more trustful than non-students.

References:

Bosch-Domenech, A./Montalvo, J.G./Nagel, R. and Satorra, A. (2002): One, two, (three), infinity,...: Newspaper and lab beauty-contest experiments. American Economic Review 92 (5), 1687–1701.

Bellemare, C. and Kroeger, S. (2007): On representative social capital. European Economic Review 51, 183-202.

Benz, M. and Meier, S. (2006): Do people behave in experiments as in the field? Evidence from donations. Working Paper No. 248, Institute for Empirical Research in Economics, University of Zurich.

Berg, J./Dickhaut, J. and McCabe, K. (1995): Trust, reciprocity, and social history. Games and Economic Behavior 10, 122-142.

Cox, J.C. (2004): How to identify trust and reciprocity. Games and Economic Behavior 46, 260-281.

Dhami, M.K./Hertwig, R. and Hoffrage, U. (2004): The Role of Representative Design in an Ecological Approach to Cognition. Psychological Bulletin 130 (6), 959-988.

Dohmen, T./Falk, A./Huffman, D. and Sunde, U. (2007): Are risk aversion and impatience related to cognitive ability? IZA discussion paper No. 2735. [Forthcoming in the American Economic Review].

Dohmen, T./Falk, A./Huffman, D./Sunde, U./Schupp, J. and Wagner, G.G. (2009): Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey. [Forthcoming in Journal of the European Economic Association].

Eckel, C.E. and Grossman, P.J. (2000): Volunteers and pseudo-volunteers: The effect of recruitment method in dictator experiments. Experimental Economics 3, 107-120.

Ermisch, J. and Gambetta, D. (2006): People’s trust: the design of a survey-based experiment. ISER Working Paper 34/2006.

Ermisch, J./Gambetta, D./Laurie, H./Siedler, T. and Uhrig, N. (2009): Measuring people’s trust. Journal of the Royal Statistical Society: Series A (Statistics in Society) 172 (4), 749-769.

Falk, A./Dohmen, T. and Sunde, U. (2009): Kontrolliert und repräsentativ: Beispiele zur Komplementarität von Labor- und Felddaten. Perspektiven der Wirtschaftspolitik 10 (1), 54-74.

Fehr, E./Fischerbach, U./von Rosenbladt, B./Schupp, J. and Wagner, G.G. (2002): A Nation-Wide Laboratory. Schmollers Jahrbuch 122 (4), 519-542.

Gächter, S./Herrmann, B. and Thöni, C. (2004): Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. Journal of Behavior & Organisation 55, 505-531.

Gächter, S. (2009): Experimental Economics. [In this publication].

Geyer, S./Norozi, K./Buchhorn, R. and Wessel, A. (2009): Chances of employment in a population of women and men after surgery of congenital heart disease: Gender-specific comparisons between patients and the general population. Congenital Heart Disease 4 (1), 25-33.

Glaeser E.L./Laibson, D.I./Scheinkman, J.A. and Soutter, C.L. (2000): Measuring trust. Quarterly Journal of Economics 115, 811-846.

Haigh, M. and List, J. (2005): Do professional traders exhibit myopic loss aversion? Journal of Finance 60 (1), 523-534.

Harrison, G.W./Lau, M.I. and Rutström, E.E. (2007): Risk attitudes, randomization to treatment, and self-selection into experiments. Working Paper 05-01, Department of Economics, University of Central Florida.

Heckman, J.J. and Smith, J.A. (1995): Assessing the case for social experiments. Journal of Economic Perspectives 9 (2), 85-110.

Hogarth, R.M. (2005): The challenge of representative design in psychology and economics. Journal of Economic Methodology 12 (2), 253-263.

Karlan, D.S. (2005): Using experimental economics to measure social capital and predict financial decisions. American Economic Review 95 (5), 1688-1699.

Levitt, S.D. and List, J.A. (2007): What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives 21, 153-174.

Max-Planck-Institute for Human Development (2009): Berlin Aging Study II (BASE II): Health and Cognitive Functioning across the Lifespan. http://www.base-berlin.mpg.de/BASE_II.html. [Last visited: 02/10/2010].

Naef, M. and Schupp, J. (2009): Measuring trust: Experiments and surveys in contrast and combination. IZA discussion paper No. 4087.

Neuroeconomics Lab Bonn, Socio-Economic Panel (2009): Bonn Gene Brain Behavior Project. http://www.neuroeconomics-bonn.org/cms/front_content.php?idcat=4. [Last visited: 02/10/2010].

Roe, B.E./Haab, T.C./Beversdorf, D.Q./Gu, H.H. and Tilley, M.R. (2009): Risk-Attitude Selection Bias in Subject Pools for Experiments involving Neuroimaging and Blood Samples. Journal of Economic Psychology 30 (2), 181-189.

Sapienza, P./Toldra, A. and Zingales, L. (2007): Understanding Trust. NBER Working Paper No. 13387.

Siedler, T./Schupp, J./Spieß, C.K. and Wagner, G.G. (2009): The German Socio-Economic Panel as Reference Data Set. Schmollers Jahrbuch 129 (2), 367-374.

7. Experimental Economics

Simon Gächter

Contact:

Simon Gächter
Centre for Decision Research and Experimental Economics
School of Economics
Sir Clive Granger Building, University Park
NG7 2RD
Nottingham
UK
e-mail: simon.gaechter[at]nottingham.ac.uk.

Abstract

Experimental economics has become an established method for generating controlled and replicable empirical information that is complementary to other empirical methods in the social sciences. There is a strong research infrastructure for laboratory experimentation in Europe and also in Germany. A valuable instrument in the development of this methodology would be the creation of a short socio-economic survey integrating questions already used in existing surveys, which experimental economists could then administer to their participants. This would make it relatively easy to analyze the selectivity of subject pools. However, among experimental economists there is as yet no existing standard questionnaire for collecting this information, which limits the ability to compare respective datasets. The effort shall be made, therefore, to create such a common questionnaire. Furthermore, there is at present no across-the-board standard for data reporting in this area. There is one data repository in the United States that currently does collect experimental data and makes them freely available. Building up a data archive that integrates (merges) existing data, however, is a very laborious undertaking and requires substantial scientific input from interested researchers.

Keywords: experimental economics, data archives, selectivity of subject pools
JEL Classification: C81, C9

Key points and recommendations

(1) Experimental economics is an established method for the generation of controlled and replicable empirical knowledge that is complementary to other empirical methods in the social sciences. There is a strong research infrastructure for laboratory experimentation in Europe and also in Germany.

(2) Most of the experiments that have been conducted by experimental economists have used students as subjects. A recent research interest to emerge thus raises the question of whether the results from the laboratory (i.e., using students) can be generalized to other social groups. Of particular interest in answering this question are experiments conducted as part of representative surveys, such as the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*) or the British Household Panel Study (BHPS, part of the longitudinal study Understanding Society). The advantage of these studies is that representative socio-demographic information can be connected to experimentally observed behavior. This method has enormous future potential and research has only just begun. In Europe, the German SOEP has played a pioneering role.

(3) There is currently no general standard for data reporting. The release of data after publication is voluntary, with the exception of two top professional journals that require accepted papers to publish their data. Only one data repository exists (in the United States) where experimental data are collected and made freely available.

(4) Building up a data archive that integrates (merges) existing data is very laborious and requires substantial scientific input from interested researchers. The construction of such a database is complex due to the multidimensionality of the data, the different interests of researchers, and their various property rights to the use of data.

(5) One valuable instrument in the development of this methodology would be the creation of a short socio-economic questionnaire drawing from questions used in existing surveys – such as the BHPS or SOEP – which experimental economists could administer to their participants. It would then be relatively simple to analyze the selectivity of subject pools. However, since no standard yet exists among experimental economists, there is limited comparability among respective datasets. An effort shall be undertaken to create such a common questionnaire.

In order to provide the necessary background for an understanding of the issues being addressed in subsequent sections, this article will first introduce the nature of experimental data. In section 2, I will then describe what I see as the current situation in experimental economics and discuss the state of data reporting and recording today. Section 3 explores some interesting future developments. Section 4 describes what I see as the main challenges facing experimental economics, and section 5 offers some concluding recommendations.

1. Research questions and data in experimental economics

Economic experiments are a method of observing economic decision making under controlled conditions. Thus, experimental economics is not a subfield of economics but rather an empirical method used to answer specific research questions. These questions come from all parts of the discipline of economics (Kagel and Roth 1995; Plott and Smith 2008; Camerer 2003; Duffy 2008; Falk and Gächter 2008; Shogren 2008). Experiments have been used to test theories, to uncover empirical regularities, to test the behavioral implications of institutions and incentives, to uncover the structure of peoples' attitudes towards risk and uncertainty, their time preferences and their social preferences. Many of these experiments can be considered basic research, but research on how experimental data might be used for consulting, policy advice, and economic engineering is growing (Roth 2002).

The methods of experimental economics are used not only within economics, but also increasingly in management science, anthropology, political science, biology, social neuroscience, and psychology. As such, experimental economics is a platform for interdisciplinary research. There are also close links to psychology, not least because experimental economics is frequently used as a toolbox by behavioral economists interested in improving the psychological realism of economics (Camerer et al. 2004). Although experimental economics and experimental approaches in psychology have a great deal in common, there are some significant differences in their respective methodologies (Hertwig and Ortmann 2001).

A large part of empirical research in economics is the use of field data, that is, naturally occurring data which accrue in daily economic life. These data are typically collected for recording purposes (e.g., by statistical offices) and are often not directly useful for answering scientific questions, in particular those that are motivated by economic theory. The reason is that economic theories (and most research questions derived from them) are typically "if-then" statements, and naturally occurring data do not exist in this fashion. In experiments these "if-then" conditions can be implemented by way of experimental design.

In addition to laboratory experiments, field experiments are also conducted where the experiment takes place in the natural decision making environment of the participants (Harrison and List 2004). Of particular interest are experiments conducted as part of representative surveys, where the advantages of experiments and survey data are combined (Fehr et al. 2002). Some recent studies also take advantage of the new range of possibilities offered by the Internet.

In the following section, I describe the procedures typically followed in a laboratory experiment. In a large majority of cases the participants are undergraduate students at the respective university. Specialized web-based software is now available for managing recruitment (Greiner 2004).[1] When participants decide to take part they normally do not know what the experiment will be about; they are invited "to take part in an experiment on economic decision making." Thus, self-selection depending on the type of experiment is not an issue. Upon arrival at the lab, the participants receive written instructions which contain the complete rules for the particular experiments.

The large majority of experiments are those conducted in networked computer laboratories and the interactions take place using specially-designed professional software, such as, for example, the popular toolbox "z-Tree" (Fischbacher 2004). In addition to being fully scripted (written

1 Exlab. University of Central Florida. http://exlab.bus.ucf.edu/.

instructions and rules ensure that experiments are always conducted in a comparable way), two further standards exist for conducting experiments: first, participants get paid on the basis of their decisions; and second, the use of deception in the design of experiments is forbidden (Hertwig and Ortmann 2001; Friedman and Sunder).[2] Thus, experiments are real decisions, not hypothetical ones, as they are in questionnaire-based research, or in simulations.

The ability to control and replicate the data generating process is one of the decisive advantages of experimentation over other methods of data collection. Naturally occurring decision situations are complex; many conditions under which natural decisions occur are unknown to the researcher and cannot be influenced or occur simultaneously with other conditions, such that it becomes impossible to say anything about causality. By contrast, in an experiment the experimenter designs ("controls") the decision situation and therefore causal inferences can be made when conditions ("treatments") change.

Replicability refers to the degree to which it is possible to run the exact same experiment – whether in the same research lab or in any other lab. This is a very important feature that is normally not feasible with other methods of data generation. There are various forms of replication. Researchers typically replicate the same experiments several times, simply to collect enough data. Sometimes researchers replicate their experiments in different participant pools (within and even across cultures) to see the robustness of findings across different social groups (Güth et al. 2003; Gächter et al. 2004; Carpenter et al. 2005; Hermann et al. 2008). Another type of replication occurs if other scientists want to run the same experiment in their own lab. This is usually quite easy, because it is an established standard of good practice to include documentation of the instructions used in the appendix of the research paper. Similarly, the software code is also frequently available. Exact replication is quite rare because it is hard to publish, but it is common to replicate previous results alongside new treatments, for instance, to create comparisons (Smith 1994). The ability to replicate results is a particular advantage of laboratory data and may not be feasible with field experiments because they take place in naturally occurring decision-making situations that may change over time in a way that cannot be controlled.

A common critique of laboratory experiments (i.e., those using undergraduates as subjects) is that undergraduates are a very specific portion of the population. Furthermore, laboratory experiments are associated with the potential drawback of being artificial situations that do not greatly resemble natural decision-making contexts (this can also be a decisive advantage of the lab experiment, however). For these reasons, it has become increasingly

2 Experiments in which participants are not paid on the basis of their decisions or that employ deception are normally not publishable in economics journals.

popular for researchers to conduct experiments using non-student participant pools and outside university labs.

Doing experiments with non-students, cross-culturally, and in a much noisier "field" has consequences for both the design of experiments and how they are statistically analyzed. Simple comparisons of means often do not suffice because the use of varied participant pools requires controlling for their characteristics. To the extent that subject pool characteristics are important, or even the focus of research there are two suggested implications: first, the requirements concerning the amount of data collected increases and, second, simple non-parametric statistics are not powerful enough for the data analysis. Multivariate regression techniques are needed. The rapid development of microeconometrics is certainly very valuable here but these techniques have to be adapted to the nature of experimental data (Andersen et al. 2007).

2. Status quo

In this section I will address the following issues: (1) the status of experimental economics, (2) the standards used in conducting experiments, and (3) the current situation in reporting data. Finally, I will describe one repository of experimental data, called "ExLab."

Status of experimental economics. Experimental economics is now an established method of empirical economic research.[3] The number of publications in this area has increased tremendously since the mid-1980s. Experimental papers are now published in all of the major journals as well as in field journals in the discipline. Since 1998 there has also been a specialized field journal (*Experimental Economics*) devoted to the development of experimental economics, broadly conceived.[4] Meanwhile, there are also textbooks (Friedman and Sunder 1994; Davis and Holt 1993), monographs (Camerer 2003; Guala 2005), and handbooks (Kagel and Roth; Plott and Smith 2008). There is a professional association of experimental economists, the "Economic Science Association,"[5] to which most experimental economists belong. Many universities, too, now run experimental economics laboratories, and the European infrastructure, including Germany, is excellent, generally speaking, and competitive with the existing infrastructure in the US.

3 The contribution of experimental economics to the economic sciences was further recognized when the 2002 Nobel Prize was awarded to Vernon Smith and Daniel Kahneman.

4 Experimental Economics. Springer. http://www.springer.com/economics/economic+theory/journal/10683.

5 Economic Science Association. https://www.economicscience.org/.

Standards for conducting experiments. I have described the current situation with regard to the rules of conduct for the types of experimentation mentioned above. The standard is quite uniform and is normally enforced through editorial policies. There is no standard for eliciting socio-demographic background information. In the past, these variables were often of little interest to researchers, because the related experiments were focused on testing behavioral theories and used convenience samples of sociologically homogeneous undergraduates for that purpose. The only notable exception to this was where there was a particular interest in gender differences. The situation today is somewhat different. Many researchers now routinely collect socio-demographic data, in particular if they are using non-student participant pools. Thanks to software that is both specialized and easy-to-use (like "z-Tree"),[6] administering these questionnaires has become relatively easy. However, no standard questionnaire for gathering background data has yet emerged.

Status quo for data reporting. It is common practice to attach the written instructions of an experiment to the manuscript when submitting it to a journal. The instructions are important in the evaluation of the validity of a given experimental design. Often these instructions are published alongside with the article or on the website of either the journal in which it is published or the author. It is uncommon, however, to submit the data itself at the review stage.

There is currently no uniform standard for reporting the data of *published* papers. At present, the three top journals in the field – the *American Economic Review*, *Econometrica*, and the *Review of Economic Studies* – publish the data (from any empirical paper, not only experimental ones) and require authors to submit the data (raw data, software, and code for analyzing the data) for publication on their websites.[7] Apart from these journals I am not aware of any other economics journal that publishes the data of empirical studies on its website. However, since the *American Economic Review* and the *Review of Economic Studies* are highly respected journals, other journals may adopt the same standard.

Some researchers publish the instructions, software, and data on their websites voluntarily alongside the paper itself, yet no homogeneous standard has emerged. There is an informal expectation that the instructions, software, and raw data from *published* papers will be supplied if requested by another researcher. It appears that people normally comply with this expectation as a social norm. When they do not agree to release data, it is usually because they intend to utilize the collected data further in new research projects.

6 Z-Tree. http://www.iew.unizh.ch/ztree/index.php.

7 American Economic Review. http://www.aeaweb.org/articles/issues_datasets.php; Review of Economic Studies. http://www.restud.com/supplementary.asp; Econometric Society. http://www.econometricsociety.org/

The ExLab data repository. To my knowledge, "ExLab" is the only repository for experimental data currently in existence. ExLab is run by the College of Business Administration of the University of Central Florida. It may be used by all researchers in the experimental social sciences.[8]

ExLab consists of three modules. The "Experiment Manager" provides a platform for organizing experiments (scheduling sessions, recruitment, registration of participants, etc.). The "Questionnaire Builder" can be used to develop online questionnaires. The most interesting function in the context of this report is the "Digital Library" module. Here, registered researchers can upload their data, instructions, software, and paper, whether the experiment is published or not. It is also possible simply to download selected materials.

There are roughly 150 projects currently registered (the projects are usually published papers). Many of them contain raw data; however, there is no common format. Some data are just a pdf-file, some are xls-files, some are Stata data files, and some refer the viewer to an external website. The quality of data documentation is variable, partly depending on how old the data are. Because the "Digital Library" is not centrally managed, the quality of data documentation depends on the researchers who upload data. In some cases socio-demographic information of participants is available.

3. Future developments

Experimental economics is clearly here to stay. It has become a valuable tool for economic research that complements existing tools. An important task of previous research was testing theories, and undergraduates were often sufficient for this purpose. Many experiments returned highly regular results, raising the important issue of whether they are generalizable to other social groups. Some developments on the horizon are a response to this question. Here I will discuss future developments (1) in field experiments, (2) in the integration of experiments into representative surveys, and (3) in the cross-fertilization with other behavioral sciences. A recent development (4) is the use of the Internet for conducting experiments.

(1) *Field experiments* are certainly the fastest growing area of experimental economics. Researchers conduct field experiments in almost all areas in the field of economics, with the possible exception of experiments that are purely theoretical that are best conducted in the lab. Field experiments are an important addition to the methodological toolbox because they enhance our understanding of economic decision making outside

8 Exlab. University of Central Florida. http://exlab.bus.ucf.edu/.

the artificial (though indispensible!) worlds of lab experiments. Field experiments can also give us a richer picture of the importance of socio-demographic variables in economic decision making. Therefore, I expect field experiments to continue to grow in importance.

(2) *Integration of experiments into representative surveys.* While running experiments in the field with non-student participants can give us important insight into the generalizability of laboratory findings, only representative samples allow us to draw more general conclusions. The integration of experiments into representative surveys is an exciting development. The SOEP[9] has played a pioneering role in this area. In the Netherlands, CentERdata has also facilitated studies with representative participant pools.[10] In the US, TESS (*Time-Sharing Experiments for the Social Sciences*) allows researchers to run experiments on representative participant pools.[11]

Recent experimental research has focused on issues of trust, fairness, and attitudes toward risk (Fehr et al. 2002; Bellemare and Kröger 2007; Mellemare et al. 2008; Dohmen et al. 2005; Naef et al. 2007). Research in this area is a promising new development and I expect it to expand rapidly, especially considering the ever-expanding body of experience with the process of conducting experiments in the surveys.

(3) *Cross-fertilization of experiments from other behavioral sciences.* Economic experiments (in particular, in simple games) are now used in all of the behavioral sciences. The datasets produced depend on the specific research environment and questions of the respective science. For example, anthropologists have run experiments in small-scale societies where people naturally have significantly different socio-economic backgrounds from those people living in modern, highly developed societies (Heinrich et al. 2006; 2005). But apart from these exceptional instances, the data are not that different than those we already know.

The situation is somewhat different in the emerging field of neuroeconomics and the closely related field of social neuroscience, both of which represent exciting new directions in the field (Sanfey et al. 2006; Fehr and Camerer 2007). Up to this point, the datasets have typically been relatively small, in particular where scanning methods (e.g., fMRI) are used. Representativeness (with regard to socio-demographics) has not yet become an issue because most research has simply tried to establish some basic facts. In this respect neuroeconomics is in the same pioneering situation that standard experimental economics was in fifteen

9 German Socio-Economic Panel. http://www.diw.de/english/soep/29012.html.

10 Centerdata. http://www.centerdata.nl/en.

11 Time-Sharing Experiments for the Social Sciences. http://www.experimentcentral.org/.

to twenty years ago. For example, research at this time sought to establish basic facts about trust and reciprocity (in rather small-scale lab studies using student-subjects). Today, experiments are run with potentially thousands of participants in representative surveys like the SOEP. It is conceivable that a similar development will occur in neuro-economics, provided some of its basic findings are replicated in other studies and appropriate techniques (e.g., biomarkers) are developed.

(4) *Experiments using the Internet*. In principle, the Internet offers the possibility of reaching large (worldwide) participant pools, in some cases of several thousand participants (Drehmann et al. 2005) who come from diverse socio-economic backgrounds (Egas and Riedl 2008). Thus, Internet experiments present a potentially attractive research tool. The drawback is that an Internet experiment allows for less control than a lab experiment. Participants might also perceive the decision making situation as more anonymous, compared to a lab environment where there are usually other people are in the room. Whether increased anonymity is a problem or perhaps an advantage depends on the research question. Some research has started to compare decision making in the lab and on the Internet (Güth et al. 2003; Anderhub et al. 2001; Charness et al. 2007). As the Internet gains in importance, combining lab and online experiments will be a fruitful area of research. The lab can provide the (small-scale) benchmark and be used to generate hypotheses about what should happen in the (large-scale) Internet experiment (or in a representative experiment).

A novel area that seems very promising consists of experiments conducted using virtual interactive platforms such as "Second Life."[12] Some researchers see great potential in using such virtual worlds for economic (Castranova 2008) or social science research (Bainbridge 2007) because experiments that are not feasible in the real world can be conducted on the Internet, and because these virtual worlds have millions of users. From the perspective of experimental economics, the question is whether experiments that are set within virtual platforms have scientific value, due to the potential for selection biases of virtual world participants and the inability to control who actually participates. Research on the comparability of results from well-known laboratory findings has just begun, but seems encouraging (Chesney et al. 2007). Thus, I expect research on virtual platforms to continue and to produce some important findings in the near future.

12 Second Life. http://secondlife.com/.

4. European and international challenges

The challenges of conducting cross-national research exist on two levels – funding and comparability of methods. The funding issue is beyond the scope of this particular report, but the question of methodology deserves some comment.

Some of the most serious challenges to methodology in experimental economics arise from conducting cross-cultural research. Ensuring the comparability of procedures and participant pools are the key problems that need to be solved in order to move forward. Of the two, comparability of participant pools is the more challenging problem. If representative experiments are not feasible, one approach is to maximize participant pool comparability by running all experiments with the same social groups (Herrmann et al. 2008).

Since participant pools will never be perfectly homogenous across locations it is important to control statistically for the socio-demographic background characteristics. For a proposal on such questions, see Siedler et al. (2008). If representative experiments are feasible the challenge is reduced to ensuring the comparability of procedures and obtaining sufficiently large numbers of participants. Previous research has shown that this can be done (see Naef et al. 2007 comparing Germany and the US). The challenges of course increase with the number of societies compared. Here, some type of collaboration, for example among different household panels, in running these experiments would be essential.

5. Conclusions and recommendations

The gold standard of any experimental science is having control over the environment and replicability of results. This holds true for experimental economics. The laboratory offers a high degree of control and many useful and replicable insights have been gained in that context. Experimental economics is an established tool that has become part of mainstream economics.

Most previous experiments have been conducted using undergraduates as subjects. The question of how these results generalize to other social groups is an interesting one. Running experiments in the field, via the Internet or as part of representative surveys, therefore, are all exciting and fruitful new tools for research that can help to answer this question.

With regard to availability of data, the situation is mixed and probably will remain so for some time. Some journals publish the data on their websites and some researchers do the same voluntarily on their individual websites. There is no "universally accepted" database or repository I am aware of where people post their data after results have been published, with the exception of the ExLab data repository described above. The question is, how desirable is such a data archive? A repository offers the advantage of creating one place where data can be found, so costs of searching are low. However, given the search machines and specialized mailing lists available today, it is also relatively cheap and easy to track down existing datasets, and most researchers are willing to send data upon request.[13] Those who are not willing to share information in this way would also likely be unwilling to submit their data to a repository as well. Maintaining a data archive and getting people to contribute to it is a very costly undertaking that probably would, due to its mainly administrative character, not be of great scientific advantage for those who maintain it.

Another issue concerns the quantitative comparison of research findings across studies (meta-analysis). This is not yet common in experimental economics, although some examples exist (Oosterbeek et al. 2004; Zelmer 2003). A meta-analysis looks at the means or medians of published findings and compares them. Even more scientifically interesting is where they merge all the data from a particular type of experiment into one database, and then perform the analysis on the combined observations (that is, all data points) of all the studies involved. Two types of analyses can be done: comparing the impact of different experimental rules on outcomes, and investigating the role of socio-demographics and other survey variables on decision making (that is, performing on a small scale what the representative experiments can do on the large scale). Being able to do this kind of research requires much more than a mere data repository can deliver. It requires building up a data archive (using database tools) that keeps track of all the dimensions and variables of the original studies (data and paradata).[14] The main problem is the nature of experimental data, which are multidimensional and very specific to a particular research question. Thus, in practice even experiments of one type (for example, trust games or public goods games) differ across multiple dimensions. Merging data from different experiments into one database and also thereby ensuring comparability is a very laborious and scientifically challenging task.

I am particularly aware of these challenges because, working together with my PhD student Eva Poen, I am currently constructing a database of all the public goods experiments I have been involved over the last fifteen years. Simply developing this database took more than one year and it is now only tailor-made for the public goods experiments I have been involved in. This database contains experimental data as well as socio-demographic information and questionnaire responses from more than 6000 participants from

13 ESA Experimental Methods Discussion. http://groups.google.com/group/esa-discuss/about.

14 Paradata are “data about data,” that is, the details of (experimental) data generation.

(only) eighteen different studies. This database will not be publicly available until we have answered our main research questions ourselves.

In summary, from my own experience I think that merging data (drawn from one type of experiment) into one database would be scientifically desirable. However, I do not think it is feasible without substantial scientific input from interested parties who then also will have property rights to the use of the database. These problems become even more profound when there is a larger number of involved scientists. A one-size-fits-all, or top-down solution to these problems will probably not work.

As I have already mentioned several times, the integration of experiments into representative surveys is an exciting new development in the field of experimental economics. This procedure allows researchers to investigate the impact of socio-demographics on experimentally observed behavior. Some researchers, including myself, have always elicited socio-demographics and responses to psychological questionnaires (similar to personality questionnaires) from the participants in their experiments. However, these efforts have not been coordinated between researchers. Moreover, (experimental) economists were only marginally interested in socio-demographics and therefore eliciting these variables was more of a subsidiary interest, which sometimes led to inconsistencies in the questionnaire design and thereby compromised comparability. Providing the scientific community with a standard set of well-conceived questions that can be administered after any experiment (and that does not last longer than 10 minutes) would be very helpful. A useful step in that direction would be if survey experts and experimental economists would collaboratively propose such a questionnaire and argue for its usefulness in the relevant and appropriate scientific forum.

References:

Anderhub, V./Müller, R. and Schmidt, C. (2001): Design and Evaluation of an Economic Experiment Via the Internet. Journal of Economic Behavior and Organization (46), 227-47.

Andersen, S./Harrison, G./Lau, M.I. and Rutström, E.E. (2007): Behavioral Econometrics for Psychologists. Working Paper 07-04, Department of Economics, College of Business Administration, University of Central Florida.

Bainbridge, W.S. (2007): The Scientific Research Potential of Virtual Worlds. Science 317 (5837), 472-76.

Bellemare, Ch. and Kröger, S. (2007): On Representative Social Capital. European Economic Review 51 (1), 183-202.

Bellemare, Ch./Kröger, S. and Van Soest, A. (2008): Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. Econometrica 76 (4), 815-39.

Camerer, C./Loewenstein, G. and Rabin, M. (Eds.) (2004): Advances in Behavioral Economics. Princeton.

Camerer, C.F. (2003): Behavioral Game Theory. Princeton.

Carpenter, J.P./Burks, S. and Verhoogen, E. (2005): Comparing Students to Workers: The Effect of Stakes, Social Framing, and Demographics on Bargaining Outcomes. In: Carpenter, J.P./Harrison, G. and List, J. (Eds.): Field Experiments in Economics. Research in Experimental Economics. Amsterdam.

Castranova, E. (2008): A Test of the Law of Demand in a Virtual World: Exploring the Petri Dish Approach to Social Science. CESifo Working Paper No. 2355.

Charness, G./Haruvy, E. and Sonsino, D. (2007): Social Distance and Reciprocity: An Internet Experiment. Journal of Economic Behavior & Organization 63 (1), 88-103.

Chesney, Th./Chuah, S.-H. and Hoffmann, R. (2007): Virtual World Experimentation: An Exploratory Study. CeDEx Discussion Paper No. 2007 (14).

Davis, D.D. and Holt, Ch.A. (1993): Experimental Economics. Princeton.

Dohmen, Th./Falk, A./Huffman, D./Sunde, U./Schupp, J. and Wagner, G.G. (2005): Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey. IZA Discussion Paper Series No. 1730.

Drehmann, M./Oechssler, J. and Roider, A. (2005): Herding and Contrarian Behaviour in Financial Markets – an Internet Experiment. American Economic Review 95 (5), 1403-1426.

Duffy, J. (2008): Experimental Macroeconomics. In: Durlauf, S. and Blume, L.E. (Eds.): The New Palgrave Dictionary of Economics Online. Basingstoke.

Egas, M. and Riedl, A. (2008): The Economics of Altruistic Punishment and the Maintenance of Cooperation. Proceedings of the Royal Society B – Biological Sciences 275 (1637), 871-878.

Falk, A. and Gächter, S. (2008): Experimental Labour Economics. In: Durlauf, S. and Blume, L.E. (Eds.): The New Palgrave Dictionary of Economics Online. Basingstoke.

Fehr, E. and Camerer, C. (2007): Social Neuroeconomics: The Neural Circuitry of Social Preferences. Trends in Cognitive Sciences 11 (10), 419-427.

Fehr, E./Fischbacher, U./Rosenbladt, B. von/Schupp, J. and Wagner, G.G. (2002): A Nationwide Laboratory. Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys. Schmollers Jahrbuch 122 (4), 519-542.

Fischbacher, U. (2007): Z-Tree: Zurich Toolbox for Readymade Economic Experiments. Experimental Economics 10 (2), 171-178.

Friedman, D. and Sunder, Sh. (1994): Experimental Methods. A Primer for Economists. Cambridge.

Gächter, S./Herrmann, B. and Thöni, Ch. (2004): Trust, Voluntary Cooperation, and Socio-Economic Background: Survey and Experimental Evidence. Journal of Economic Behavior and Organization 55 (4), 505-531.

Greiner, B. (2004): An Online Recruitment System for Economic Experiments. In: Kremer, K. Macho, V. (Eds.): Forschung Und Wissenschaftliches Rechnen. Gwdg Bericht 63. Göttingen.

Guala, F. (2005): The Methodology of Experimental Economics. Cambridge.

Güth, W./Schmidt, C. and Sutter, M. (2003): Fairness in the Mail and Opportunism in the Internet – a Newspaper Experiment on Ultimatum Bargaining. German Economic Review 4 (2), 243-265.

Harrison, G.W. and List, J.A. (2004): Field Experiments. Journal of Economic Literature 42 (4), 1009-1055.

Henrich, J./Boyd, R./Bowles, S./Camerer, C.F./Fehr, E./Gintis, H./McElreath, R./Alvard, M./Barr, A./Ensminger, J./Henrich, N./Hill, K./Gil-White, F./Gurven, M./Marlowe, F.W./Patton, J.Q. and Tracer, D. (2005): Economic Man in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies. Behavioral and Brain Sciences 28 (6), 795-855.

Henrich, J./McElreath, R./Barr, A./Ensminger, J./Barrett, C./Bolyanatz, A./Cardenas, J.-C./Gurven, M./Gwako, E./Henrich, N./Lesorogol, C./Marlowe, F./David, T. and Ziker, J. (2006): Costly Punishment across Human Societies. Science 312 (5781), 1767-1770.

Herrmann, B./Thöni, Ch. and Gächter, S. (2008): Antisocial Punishment across Societies. Science 319 (5868), 1362-1367.

Hertwig, R. and Ortmann, A. (2001): Experimental Practices in Economics: A Methodological Challenge for Psychologists? Behavioral and Brain Sciences 24 (3), 383ff.

Kagel, J. and Roth, A.E. (1995): The Handbook of Experimental Economics. Princeton.

Naef, M./Fehr, E./Fischbacher, U./Schupp, J. and Wagner, G.G. (2007): Decomposing Trust: Explaining National and Ethnical Trust Differences. [Unpublished manuscript].

Oosterbeek, H./Sloof, R. and van de Kuilen, G. (2004): Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. Experimental Economics 7 (2), 171-188.

Plott, Ch.R. and Smith, V.L. (Eds.) (2008): Handbook of Experimental Economic Results, Vol. 1. Amsterdam.

Roth, A.E. (2002): The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics. Econometrica 70 (4), 1341-1378.

Sanfey, A.G./Loewenstein, G./McClure, S.M. and Cohen, J.D. (2006): Neuroeconomics: Cross-Currents in Research on Decision-Making. Trends in Cognitive Sciences 10 (3), 108-116.

Shogren, J.F. (2008): Experimental Methods in Environmental Economics. In: Durlauf, S. and Blume, L.E. (Eds.): The New Palgrave Dictionary of Economics Online. Basingstoke.

Siedler, T./Schupp, J./Spiess, C.K. and Wagner, G.G. (2008): The German Socio-Economic Panel as Reference Data Set. RatSWD Working Paper No. 48.

Smith, V.L. (1994): Economics in the Laboratory. Journal of Economic Perspectives 8 (1), 113-131.

Zelmer, J. (2003): Linear Public Goods Experiments: A Meta-Analysis. Experimental Economics 6 (3), 299-310.

8. Experience Sampling

Michaela Riediger

Contact:

Michaela Riediger
Max Planck Institute for Human Development
Max Planck Research Group “Affect Across the Lifespan”
Lentzeallee 94
14195 Berlin
Germany
e-mail: riediger[at]mpib-berlin.mpg.de

Abstract

Experience sampling refers to the repeated sampling of momentary experiences in the individual's natural environment. The methodological advantages of this approach include the minimization of retrospective response biases and the maximization of the validity of the assessment. The conceptual benefits it offers include insights into short-term processes and into the daily-life contexts of the phenomena under study. Making use of the benefits of experience sampling while taking its methodological challenges into consideration allows researchers to address important research questions in the social and behavioral sciences with great precision and clarity. Despite this, experience sampling information is rarely found in the data infrastructure publicly available to researchers. This situation is in stark contrast to the way this methodology is thriving today in research-producing datasets that are not publicly available, for instance, in many psychological investigations. After a discussion of the benefits and challenges of experience sampling, this report outlines its potential uses in social science and economic research and characterizes the status quo in experience-sampling applications in the currently available datasets, focusing primarily on household surveys conducted after 2001. Recommendations are offered for an intensified use of experience sampling in large-scale data collections and how this might be facilitated in the future.

Keywords: experience sampling in the social and behavioural sciences

1. What is experience sampling?

Experience sampling refers to the capturing of experiences – such as events, behaviors, feelings, or thoughts – at the moment of, or close to, their occurrence, and within the context of a person's everyday life. The distinctive characteristic that sets this methodology apart from other assessment approaches is the *repeated* sampling of *momentary* experiences in the individual's *natural* environment (as opposed to, for example, single-time retrospective reconstructions of past experiences in questionnaires or interviews). Many labels, such as event sampling, real-time data capture, time-situated method, ambulatory assessment, diary method, or ecological-momentary assessment, have been used to refer to this methodology. In this report, I use the term *experience sampling* coined by Mihaly Csikszentmihalyi and colleagues in the 1970s, which has since been widely adopted.

The core method in experience sampling, and hence the primary emphasis of this report, is the acquisition of repeated self-reports of momentary experiences or of experiences that occurred during short preceding time intervals (typically covering no more than 24 hours). Assessment schedules in experience sampling research include (a) *interval-contingent sampling* (assessments at fixed points in time, such as before going to bed at night), (b) *signal-contingent sampling* (assessments triggered by signals that typically occur at varying time intervals throughout the day and that are given by electronic assessment devices, such as handheld computers), (c) *event-contingent sampling* (assessments triggered by the occurrence of pre-specified events, such as expenditures), and (d) any combinations of the above. Which assessment schedule is most appropriate in a given study context depends on the specific research question at hand, the prevalence of the particular experience under study, and on feasibility considerations.

Although self-report is the core assessment method in experience sampling and the primary focus of this report, it should be noted that other assessment techniques originating from diverse scientific disciplines can be used as complementary assessment strategies to capture the multiple facets of naturally unfolding experiences and their contexts. These techniques include the ambulatory monitoring of physiological processes or physical activities (see the advisory report on bio-markers in this publication), the recording of behavioral information (e.g., performance in cognitive tasks), the recording of ambient environmental parameters (e.g., sound recordings, photographs of the environment), or the recording of the individual's geographical locations (e.g., geo-tracking, see report on geographical data).

This report opens with a discussion of the benefits and challenges of experience sampling, followed by an outline of its potential uses in social science and economic research. I will then characterize the current situation by looking at experience sampling applications in available datasets, focusing primarily on household surveys conducted after 2001. Based on this assessment, I will draw some conclusions about the future development of experience sampling and its contribution to the data infrastructure and offer some suggestions for how this methodology can address present and future research needs in the social and behavioral sciences.

2. The benefits and challenges of experience sampling methodology

When compared to retrospective self-report – the most widely used assessment approach in social and economic data surveys – experience sampling offers compelling benefits, both from a methodological and a conceptual perspective. At the same time, it is accompanied by some significant challenges, including being a more resource-intensive methodology. Hence, careful consideration of both its benefits and challenges is necessary in order to take full advantage of this powerful methodology.

There are important *methodological* advantages in experience sampling that are brought about by the immediacy of the measurement and the fact that it takes place in the participants' natural environments. It is well known that human memory imposes limits on the validity of what people can report retrospectively. In most questionnaires or interviews, respondents have to rely on partial recall and inference strategies when asked to report on their past behavior or experiences. There is ample empirical evidence that this results in retrospective memory biases and aggregation effects that impair the validity of the information assessed, sometimes profoundly. Experience sampling provides a promising alternative by obtaining reports of experiences at the moment of, or close to, their occurrence. Furthermore, the fact that this information is collected within the natural context of the participants' day-to-day lives further enhances the validity of the assessment, offering unique opportunities to understand experiences and behaviors in their ecological context (Schwarz 2007). Today, experience sampling assessment is typically implemented with the help of electronic assessment devices such as handheld computers, which provide the added methodological benefit of allowing close monitoring of participant response adherence to the measurement scheme.

The prevailing emphasis in most available data collections in the social and economic sciences to date is on differences *between* individuals at given points in time. A fundamental dimension of many aspects in human life – their inherently fluctuating nature as reflected in short-term *within*-person variations – has not yet received much attention, even though the importance of within-person processes for understanding many social and behavioral phenomena has been acknowledged in theory. Hence, a compelling *conceptual* benefit of experience sampling results from the fact that assessments are repeated with short time intervals between them. This makes short-term processes and fluctuations – which cannot be studied with the traditional fixed annual assessment schedules – accessible to scientific investigation. Another conceptual benefit of experience sampling is that it provides insight into the role of everyday contexts for the target phenomena under study, such as the respective roles played by the individual's educational, work, or social environments.

Despite these methodological and conceptual benefits, there are significant challenges that need to be considered when implementing the experience sampling method. Of these challenges, three stand out as particularly critical. First, experience sampling is resource-intensive. Because motivation plays such a significant role in determining whether a participant will successfully complete an experience sampling study or not, close contact with the participants throughout the entire study and adequate remuneration are indispensable. Second, the burden for the participants (e.g., the necessary time commitment) is comparatively large. This creates difficulties in terms of both representativeness and attrition of the sample. The demanding nature of experience sampling studies could lead certain types of individuals to be over- or underrepresented in the sample from the beginning, or to drop out during the study interval. Finally, repeated measuring of a given phenomenon can cause reactivity effects. That is, it is possible that the phenomenon under study may change as a result of measurement or reporting. Although reactivity is a challenge for all social and behavioral research, it can be even more relevant in experience sampling research because the repeated assessments may lead people to pay unusual attention to their experiences and behaviors.

In short, experience sampling carries immense methodological and conceptual advantages. Nonetheless, it also presents a number of challenges that need to be considered, which I will discuss in detail in the concluding section of this report. When adequately applied, however, experience sampling indisputably represents a powerful tool with which to tackle new questions and investigate research questions in greater depth. In the following section I will describe the ways that experience sampling can be applied to social science and economic research.

3. Potential uses of experience sampling in social science and economic research

Generally speaking, experience sampling can provide fine-grained and ecologically valid information on

- the *Who*, *What*, *Where*, *When*, or *How* of experiences and behaviors as they occur in daily life and in natural environments,
- the naturally occurring *variation* and *co-variation* of experiences, behaviors, events, and contextual characteristics over time (both *within* and *between* individuals), and
- the within-person *variability* of experiences and behaviors (i.e., short-term fluctuations or changes) that, depending on the research domain under study, can be indicative either of people's flexibility or adaptability, or of their instability and vulnerability.

Obviously, these are questions that are of immense relevance and importance for a large variety of domains in social, behavioral, and economic research. There are a vast number of potential applications that could provide new insight into diverse phenomena. These include the investigation of life transitions (e.g., divorce, unemployment, childbirth, entering the workforce, or retirement), social interactions, investment or buying behaviors, health behaviors and health-care use and effectiveness, well-being and life satisfaction, family life, work life, availability, use and effectiveness of the educational system, major life events and stressors, as well as investigations of many other research domains. Despite the wide spectrum of potential applications, experience sampling information is still rare in the data infrastructure that is publicly available to researchers in the social and behavioral sciences. This stands in stark contrast to the growing application of this methodology in research activities which produce datasets that are not publicly available, as is the case in many psychological investigations. The following section provides an analysis of the current state of experience sampling applications in the social and behavioral sciences.

4. Status quo of experience sampling in the data infrastructure

The purpose of the following analysis is to characterize the status of experience sampling information in the available data infrastructure. The first part of this analysis addresses the present use of experience sampling in household surveys. It illustrates the scarcity of experience sampling information in the datasets that are accessible to the public and interested researchers. The second part of this analysis addresses the status of experience sampling in psychological research. The purpose of this section is to illustrate how the methodology is actively involved in the production of datasets, but these are available only to a small number of scientists connected to the original research. The concluding section of this report will build on this analysis of the status quo to formulate some recommendations for future research needs and challenges.

4.1 Experience sampling in household surveys with ongoing data collection since 2001

To identify contemporary household surveys employing experience sampling methodology, I conducted a search using the keywords "experience sampling," "diary/diaries," and "ambulatory assessment" in the following databases:

- Data Catalogue of the GESIS Data Archive[1]

1 http://www.gesis.org/Datenservice/Suche/Daten/index.htm

- Survey Databank of the German Youth Institute (*Surveydatenbank des Deutschen Jugendinstituts*)[2]
- National Statistics' Database of Longitudinal Studies[3]
- Data Catalogue of the Economic and Social Data Service[4]

Table 1 lists household surveys that apply experience sampling based on the results of this search strategy and that also demonstrate ongoing data collection since 2001 (up until June 20, 2008). The table shows that only a few household panels currently integrate experience sampling. All of the identified applications of this methodology in household surveys used experience sampling in the form of diaries; that is, in the form of interval-contingent, short-term retrospective assessments. Table 1 also shows that the methodology is applicable in large-scale data collections and well-suited for the investigation of a wide array of phenomena. This is further demonstrated by the fact that the German Federal Statistical Office in collaboration with the Statistical Offices of the *Länder* successfully obtains household expenditure diaries in the German Income and Consumption Survey (EVS, *Einkommens- und Verbrauchsstichprobe*).

None of the most prominent international prospective household panels – the US Panel Study of Income Dynamics (PSID), the German Socio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*), the British Household Panel Study (BHPS, to be succeeded by the UK HLS), and the Multidisciplinary Facility for Measurement and Experimentation in the Social Sciences (MESS, Netherlands) – have yet employed experience sampling methodology. Nonetheless there are clear signs of a growing awareness of, and interest in the powerful potential of this methodology. The study proposal of the Dutch household panel MESS, for example, highlights experience sampling as a potential method for future assessment waves. Furthermore, the German Socio-Economic Panel has recently developed a mobile-phone based experience sampling technology in cooperation with Max Planck Institute for Human Development (Berlin) that makes the application of signal-contingent experience sampling possible in heterogeneous and widely distributed samples. The feasibility of this technology has already been demonstrated in a first model study involving a sample of $N = 378$ participants ranging in age from 14 to 83 years. Participants were provided with mobile phones that they carried with them while pursuing their daily routines. Testing software was installed on the mobile phones that caused the phones to ring at certain points throughout the day and signaled the participant to complete an assessment instrument that referred to his or her mo-

2 http://db.dji.de/surveys/index.php?m=msa,0
3 http://www.iser.essex.ac.uk/ulsc/keeptrack/index.php
4 http://www.esds.ac.uk/search/searchStart.asp

mentary experiences. Participant responses were then immediately uploaded via the Internet to a central server. The server interface was also used to set up the study design, to manage the data collection, and to monitor participant response compliance.

Table 1. Experience Sampling in Household Panels with Ongoing Data Collection since 2001

Country	Panel	Experience sampling	Data accessibility
UK	Expenditure and Food Survey Start: 2001–2002 Most recent data: 2005–2006 Sample size: 6,164 households in Great Britain, and 527 in Northern Ireland Design: repeated cross-sectional	Diaries of personal expenditures, homegrown and wild food brought into the home. Kept by each adult for two weeks; simplified diaries kept by children aged 7 to 15 years for two weeks.	Derived variables from the diary are included in the dataset, as the raw diary data are not released to the public for confidentiality reasons (access contingent upon registration).
UK	Home On-Line Survey (HoL) 1998–2001 (finished) Sample size: 999 households, all household members older than 9	Seven end-of day diaries (comprehensive activity diaries).	Access contingent upon registration.
UK	Scottish Household Survey Start: 1999 Most recent data: 2007 Sample size: 27,000 in 2003–2004 (diaries) Design: repeated cross-sectional	One travel diary provided on day prior to interview by one randomly selected adult of the household.	Access contingent upon registration.
Denmark	Time Use of Households: A Scheduling of Danes Daily Use of Time Started: 1987 Most recent data: 2001 Sample size: 4,000 Design: longitudinal (2 occasions)	Diaries kept by respondents and their partners for two days, one randomly selected weekday, and one randomly selected weekend day (activities, social partners).	Application to Danish National Institute of Social Research.
Ireland	Household Budget Survey Started: 1951 Most recent data: 2004–2005 Sample size: 6,884 households in 2004–2005 Design: repeated cross-sectional	Detailed diary of household expenditure over a two-week period.	From 1987 on request to Irish Social Science Data Archive.

4.2 Experience sampling in psychological research

The relatively rare use of experience sampling in large-scale data collections such as household surveys – surveys that are designed to contribute to a broadly accessible data infrastructure – stands in stark contrast to the way the methodology has been taken up in research activities designed to produce smaller datasets and available to a limited number of researchers. One example, which is discussed in this section, can be found in the field of psychological research. Other examples of fields where experience sampling is frequently used – in time use studies and transportation research – are the focus of other advisory reports in this publication so they are not addressed here.

The methodological and conceptual strengths of experience sampling are well-recognized in psychological research. This has led to a recent upsurge in the use of experience sampling methodology for psychological investigations. Hundreds of papers on experience sampling investigations have been published since 2001. As of 20 June 2008, for example, and taking into account only publications that have appeared between 2001 and 2008, the database PsycINFO yielded 355 hits for the keyword "experience sampling," 175 hits for the keyword "diary method," and 188 hits for the keyword "ambulatory assessment." Other indications of the dynamic growth of experience sampling methodology in this area is the recent publication of several monographs on experience sampling methodology and special issues dedicated to this theme in international psychology journals (e.g., Ebner-Priemer et al. in press; Hektner et al. 2007; Stone et al. 2007; Westmeyer 2007); and the recent foundation of the "Society of Ambulatory Assessment" in 2008[5]).

Although experience sampling in psychological research is most often applied in small samples (i.e., N < 200) that are queried only once, experience sampling has also been successfully included as an assessment method in comparatively larger and longitudinal research projects, particularly those conducted in the US. Examples of these include:

- the "National Survey of Midlife Development in the USA" (MIDUS, N = 7,189) in which experience sampling in the form of eight subsequent telephone interviews on daily life was administered in a subproject entitled, "National Study of Daily Experiences" (NSDE, N = 1,483);
- the "Normative Aging Study" (NAS, N = 2,280) in which experience sampling in the form of eight consecutive daily diaries on stressful events, memory failures, etc. was administered in a subsample of N = 333 participants; and,

5 http://www.ambulatory-assessment.org/

- the “Alfred P. Sloan Study of Youth and Social Development” in which signal-contingent sampling of momentary experiences was repeatedly administered in a sample of $N = 877$ adolescents.

Taken together, the recent increase in the use of experience sampling methodology in psychological research underscores the methodological and conceptual strengths of this approach and demonstrates its applicability to a variety of populations. However, these uses in psychological research have not yet contributed to an enrichment of a wider data infrastructure available to a community of interested researchers at large. Rather, access to experience sampling datasets in psychology typically remains limited to a narrow group of researchers within the network of those involved in the conceptualization of the study and the collection of the data. Release of those data to the research community is not yet common practice in psychological research.

5. Recommendations for future developments and challenges

To summarize, experience sampling is a promising research tool that has profound methodological and conceptual benefits compared to standard survey methodologies of retrospective or general self-reports. It has the potential to provide important and ecologically valid insights into a large array of research domains in the social and behavioral sciences. Although experience sampling currently occupies a lively position in psychological research, only a few applications of experience sampling are available in data collections that feed into the publicly available data infrastructure. There are, however, indications of a growing awareness of the potential of experience sampling in the international research landscape.

A broad conclusion that can be drawn from these analyses is that making use of the benefits of experience sampling, while taking its methodological challenges into consideration, will contribute to the creation of a data infrastructure that makes it possible to address current and future research questions with greater precision and clarity. In the following section I offer six concluding recommendations focused on facilitating the intensified use of experience sampling in large-scale data collections now and into the future.

(1) *Strengthen multi-method approach in large-scale surveys.* Experience sampling is a potent methodology that can supplement standard survey methodology such as global or long-term retrospective self-reports. Its methodological advantages (e.g., minimization of response biases and

maximization of ecological validity) allow for the investigation of existing research questions in great depth. Its conceptual advantages (e.g., accessibility of short-term fluctuations and change within and between individuals, the respective role of contextual characteristics) generate opportunities for tackling new research questions.

(2) *Consider a 'study within a study' solution in large-scale data collections.* Experience sampling is resource-intensive. Theory-driven applications in selected subsamples of participants will therefore increase the feasibility of experience sampling in large-scale data collections.

(3) *Make use of technological advances in experience sampling applications.* Technological advances can be used to increase the feasibility of experience sampling in large-scale and heterogeneous samples and also to decrease the burden of experience sampling for the participants. Particularly promising for large-scale data collections is the use of mobile technology. Among its advantages are (a) the potential to use the participants' own mobile phones as assessment devices, (b) the central control of study content and assessment schedules via web-interfaces in server-client systems, (c) the immediate upload of data to central servers allows the monitoring of participant response compliance, (d) the relative unobtrusiveness and feasibility of measurement completion in daily life contexts (provided assessment instruments are of adequate length), and (e) the easy combination with follow-up interviews or other assessment strategies stemming from diverse scientific areas (e.g., for ambulatory bio-monitoring see the advisory report on bio-markers in this publication; for location-tracking, see the report on geographical data).

(4) *Address the methodological challenges of experience sampling.* Study designs should adopt appropriate measures to address the methodological challenges of experience sampling. Control group designs are necessary to assess potential reactivity effects, to note possible changes in the phenomenon under study caused by its measurement. Careful sample recruitment strategies are needed to minimize potential self-selection biases that would result in limited sample representativeness. Sample attrition, or participant drop-out, can be minimized by maintaining close contact to the participants during the study interval and by implementing reasonable study characteristics, such as those pertaining to the number of measurement occasions and the length of the assessment instruments.

(5) *Increase the accessibility of experience sampling datasets.* To increase the availability of experience sampling datasets in the data infrastructure of the social and behavioral sciences, it is essential to foster the release of datasets to the larger research community. One possible form this

could take is to make research funding grants contingent upon the researcher consenting to release the obtained dataset to the research community after a reasonable amount of time (e.g., after 7–10 years).

(6) *Advance research on experience sampling methodology.* Methodological research will support the greater implementation of experience sampling methodology in survey designs. One way to promote research on experience sampling methodology is to include it as a research topic in the Priority Programme on Survey Methodology of the German Research Foundation (DFG, *Deutsche Forschungsgemeinschaft*).

References:

Ebner-Priemer, U./Pawlik, K. and Kubiak, Th.E. (Eds.) (in press): Ambulatory Assessment [Special issue]. European Psychologist.

Hektner, J.M./Schmidt, J.A. and Csikszentmihalyi, M. (2007): Experience sampling Method: Measuring the quality of everyday life. Thousand Oaks.

Schwarz, N. (2007): Retrospective and concurrent self-reports: The rationale for real-time data capture. In: Stone, A.A./Shiffman, S./Atienza, A.A. and Nebeling, L. (Eds.): The science of real-time data capture. New York.

Stone, A.A./Shiffman, S./Atienza, A.A. and Nebeling, L. (Eds.) (2007): The science of real-time data capture. New York.

Westmeyer, H.E. (Ed.) (2007): Advances in the methodology of ambulatory assessment [Special issue]. European Journal of Psychological Assessment 23 (4).

9. Virtual Worlds as Petri Dishes for the Social and Behavioral Sciences

Edward Castronova and Matthew Falk

Contact:

Edward Castronova
1229 East 7th Street
Bloomington, IN 47405
US
e-mail: castro[at]indiana.edu

Abstract

The next tool developed for social science experimentation should allow for macro level, generalizable scientific research. In the past, devices such as rat mazes, petri dishes and supercolliders were developed when scientists needed new tools to do research. We believe that virtual worlds are the modern equivalent to supercolliders for social scientists, and that they should be the next area to receive significant attention and funding. The advantages provided by virtual worlds research outweigh the costs. Virtual worlds allow for societal level research with no harm to humans, large numbers of experiments and participants, and make long term and panel studies possible. Virtual worlds do have some drawbacks in that they are expensive and time-consuming to build. These obstacles can be overcome, however, by adopting the models of revenue and maintenance practiced by the current game industry. The returns from using virtual worlds as scientific tools could reach levels that would self fund future research for decades to come. At the outset, however, an initial investment from funding agencies appears to be necessary.

Keywords: virtual worlds, macro-level experiments, research infrastructure
JEL Classification: C15, C59, C82, C99

1. Introduction

In the past, science developed new tools for research as the need arose. From petri dishes to rat mazes, and continuing on even to the construction of supercolliders, scientists require specific tools to answer the questions they ask. These devices all influence specific micro-level observations. But when it comes to social science and research questions on the societal level, tools for empirical research had not developed much beyond where they were two centuries ago until recently. Developments in the collection of survey data began to take place after "World War II." In the 1990s, "experimental economics" started to become a new and popular tool for empirical research (although, surprisingly, seldom applied by sociologists). Now, however, social scientists should be looking toward a new area: virtual worlds (VW).

To be considered a virtual world, a game or social networking site must: be computer generated, persistent (i.e., always there even if no one is currently logged into it), and have humans represented in the form of avatars (the embodiment of the user in the virtual space) capable of taking actions on behalf of their human counterpart. Only in virtual worlds do we find the proper tool set for large scale social science research, something previously unavailable to scientists. These defining features combine to allow scientists access to long-running persistent societies of users, all engaged in actions that resemble what we see in the real world (Castronova and Falk 2009).[1] Please note that what we are discussing in this paper are experiments on the macro- level of a society. We are not addressing the issue of using virtual worlds for conducting experiments on the micro-level of individual players (see, for example, Chesney et al. 2007). For an even simpler approach using virtual worlds like “Second Life” for social surveys, see Bell et al. (2008).

Because of the large-scale commercial success and now widespread use of VWs, it is possible to collect large amounts of data from large numbers of users. Instead of a few hundred people in one place for a short time, as in current experimental economics, for example, and other lab-based research, we can draw from populations that range from thousands to millions and take measurements over time. Because of the size of the populations involved, VWs let us look at causation at the macro or societal level.

Virtual worlds range in scope from small-scale, internet browser based games with perhaps a few hundred players to the massively successful game “World of Warcraft,” which has had ten million subscriptions purchased in its four years of existence and has an estimated consistent player base of eight million. The populations of these worlds span the globe, and it is just as possible to meet someone from thousands of miles away as it is to join your friends from down the street when exploring the virtual world (Castronova 2005; 2007).

2. Petri dishes, rat mazes, supercolliders

While the virtual world is not a sealed vacuum, it does resemble a petri dish in its functionality (Castronova and Falk 2009). Many users, millions at a time in fact, can exist in a game like “World of Warcraft.” These users are not, however, all interacting with each other in one space. Like petri dishes in the laboratory, individual servers – digital copies of the same world with unique users interacting – make it possible for technology to handle the demand. The servers, or individual petri dishes, contain the same ingredients in them. It is the bacteria – in this case the players – that differ based on the server they choose.

Because the servers all inherently begin as exactly the same world, it is possible to make one small change to the composition of the goo in the petri dish – a single variable on one server – to create experimental conditions. Server after server, side by side, can resemble rows of petri dishes in a lab. One group contains a set of control conditions, another group one experimental condition, and so on. The underlying code, or the thing that makes it

1 See also Giles (2007) and Miller (2007).

all work, does not change. The color of the sky, the names of the places, and the sizes of the oceans do not differ, unless of course that is what the scientist chooses to change. The only restriction to the number of servers and amount of players is monetary, something we will return to below.

Just because all this is possible does not make virtual worlds perfect for answering all questions. In fact, there are some types of questions that virtual worlds are poor at answering. Like all experimental tools, the tool must be designed to answer the types of questions that the researcher wants to ask.

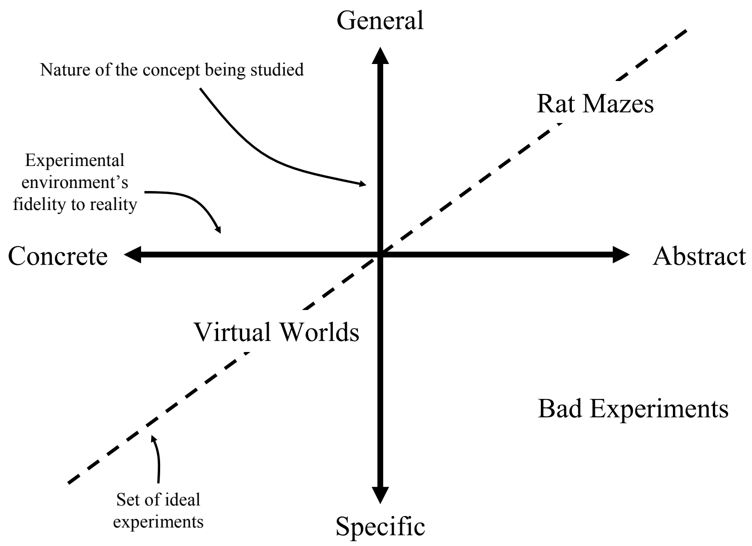
For instance, mammalian cognition is a frequently studied topic. Some scientists use rat mazes to test the cognitive habits of rats and others recreate the mediated environments humans encounter every day to examine what effects they have on the brain. Both of these cases provide good examples of how scientific tools are shaped to correspond to the questions at hand.

Figure 1 demonstrates the relationship between type of question asked and the tools used to study them. The horizontal axis in figure 1 arrays experimental environments according to their fidelity to reality. An environment that is very concrete replicates reality quite well. It is a simulation. A media effects environment that places a TV with current programming in an American living room – replicated right down to the six-pack of beer and the cat odor – is concrete. A media effects environment that attaches wires to individuals' heads and has them watch triangles on a small screen while holding a buzzer is abstract.

The dashed line running diagonally across figure 1 registers the set of ideal experiments – where the conceptual level of the question is well-matched by the concreteness of the experimental environment. If the research question is specific, the experimental environment must be concrete. If the question is general, the experimental environment must be abstract. The area labeled "Bad Experiments" in the figure refers to the attempt to study a specific question within an abstract research environment. You cannot conclude much about the reaction of typical American families to last night's newscast by wiring their heads and asking them to watch triangles on a screen. Bad experiments can go the other way too: you cannot learn much about the general rate of response times to visual stimulus just by watching people in their living rooms.

The point of the diagram is this: it is senseless to make claims about the validity of a research environment unless you know what sort of question is being studied. A rat maze is a terribly abstract environment, yet would anyone say, "You can't learn anything about anything in a rat maze. A rat maze is too unlike the real world." It is possible to learn a great deal about mammalian cognition from rat mazes. When research questions involve societies, or have macro-level implications, we must build a more concrete and specific environment.

Figure 1. Experimental Tools and Scientific Questions



Source: Castronova and Falk (2009).

3. Current state of the research field

While many researchers examine communications and media in the form of virtual worlds, only a very small number are using empirical methods to do so. Much of the research is concerned with theory creation, observational and ethnographic methods, including observational analyses by means of regression analysis. All of these methods are valid and collect pieces of information, but none of them are experimental and as such do not lead to the concrete, generalizable, macro-level information about human behavior that social science is seeking.

Social and behavioral scientists seek to understand how humans interact. Social scientists, specifically, want to explore the large questions of human interactions: war, disease, starvation, ecological disaster, economic stability, etc. The only way to solve macro-level problems is with macro-level science. *Controlled experimentation at the societal level* is not being conducted, and in fact would be impossible to conduct under normal circumstances. It is inadvisable and indeed impossible to remove real humans from society, place them in a vacuum for months or years at a time and then experiment on them. There are some attempts being made at this type of science, known to economists as "field experiments" (Harrison and List 2004; Hausman and Wise 1985; List 2008) but they are often derided for their inability to produce controlled and generalized results.

Field experiments take one of three forms: artificial field experiments, framed field experiments, and natural field experiments. Artificial field experiments tend to resemble laboratory experiments as closely as possible, but with a sample drawn from a specific population of interest. This eliminates generalizability unless that group is itself representative of the general population. Framed field experiments entail placing experimental differences in their natural habitat, such as providing different social programs for groups and determining which choice worked better overall. These again target a specific population (i.e., those participating in that particular choice at that location at that time). While less abstract than artificial field experiments, they do not hold up to the rigor of laboratory testing standards. The closest of the three to laboratory science are natural field experiments, which combine the anonymity on the part of the subject with the experimental manipulation of framed field experimentation. They fall short again, however. Due to the interaction of natural environment and the lack of available opportunities to produce works, researchers are limited by their reliance on the presence of naturally occurring phenomena that they can get approval to study.

Governmental interest in research generally falls into a category all of its own: simulation research. On the surface, simulation research looks like the virtual worlds research we propose here, but there is a fundamental difference between the two. Simulations are essentially computer-run models, in which the players (known as "agents") are also computerized. Each behaves in a manner that is simple and is predictive of how an individual would act, assuming that each individual will always make the most "rational" choice. This is problematic, however, since many believe that humans do not react rationally to many situations, if any at all, and therefore consider the interactions of simple, rational models to be incomplete. A fundamental improvement to this model would be to use real humans in place of the agents – which is exactly what virtual worlds offer.

Building what amounts to the social science equivalent to a super-collider, however, appears both necessary and expensive. Preliminary forays into this research field are already being conducted. Our group, the Synthetic Worlds Initiative of Indiana University, has already completed the construction of a small-scale virtual world and the subsequent experimentation process within.

Funded by a USD $250,000 grant from the MacArthur Foundation, "Arden" – a world based loosely on the works of William Shakespeare – took a student team almost two years to build. It then required another several months to run an experiment within the world and to compile those results for publication (Castronova et al. 2008). The experimental test run of Arden was an investigation of the economic theory of supply and demand. Having found that the law of supply and demand holds true in a virtual world, Arden was deemed a successful first step towards the creation of a virtual petri dish, or supercollider. But it was, in fact, only a first step.

The next logical step is to create another virtual world, capable of housing more users and answering bigger questions. Along these lines, we are currently developing a game called Greenland that will be used to test the emergence of currency in the form of a web browser based resource collection game. But even this project, which can expect somewhere between several hundred to several thousand subjects, is merely another small development. The ideal supercollider-level virtual world would be more like "World of Warcraft," and consequently cost much more.

4. Development costs, future research and recommendations

Developing a persistent A-list, or top quality, virtual world game requires not only a significant investment of time and personnel, but also involves large overhead costs for startup. This can be an insurmountable obstacle in terms of current social science research funding awards. Other areas of research and public service, however, provide models for the research and development of extremely expensive projects that get results and, in the end, generate profits to replenish those research and development costs.

The cost of developing virtual worlds are typically held in secrecy, since game design companies do not want to publicize exactly how much they've spent developing their projects. However, on the basis of the knowledge shared by former and current leaders in that industry, it is quite possible to infer how much one should cost. Game development costs come primarily from three areas: the game design team (development), game launch, and customer service during the years in which a world operates.

Game design teams are typically small at the beginning, possibly five or so people, but rapidly expand to include teams of 25–40 people, depending on the size of the project. This expansion occurs over a couple of years, and projects regularly take more than thirty months from the initial design meeting to the end of testing, or launch. For example, in 2005, the rule of thumb was that it cost approximately USD $10,000 per month per person on the team. This does not match up with the current size of research funding in the field – remember that Arden was created on a USD $250,000 grant over the course of two years by a team of approximately fifteen people. Professional game developers also work longer than forty-hour weeks, and are dedicated staff, whereas Arden and Greenland are being developed by graduate students working part time. Hiring professional staff would greatly speed up projects like this and allow for faster game development and more experimentation.

In addition to the cost of personnel, each of the servers (i.e., the petri dishes) discussed earlier is quite an expensive investment and requires a large amount of expensive bandwidth to run. Setting all this up and making sure it works before releasing it to the public is the next step in game development. In 2003, for example, the average cost for game "launch" (as it is known in the industry) was USD $7 million, with amounts in the USD $10–12 million range being more the rule than the exception. Current research is being performed on two small servers hosted on university campuses. These servers simply cannot handle the mass numbers of players as the servers game companies use. Therefore, this limits both the number of study participants and variations of the virtual world it is possible to have.

After the game goes public and research and play commence, there is still a significant amount of time and money required on the part of the game support staff. They must maintain player relations, collect subscription fees (if using one of the fee-based models we discuss below), and take care to maintain the software and hardware that allow users to access the world.

There are, however, two examples of ways through which it would be possible to fund large public projects, pay back the funding agencies, and create profits for further research and projects. These examples can be gleaned from parallels between nuclear power plants, the pharmaceutical research industry, and our vision of a virtual world as a supercollider.

Nuclear power plants, like new experimental drugs, are initially funded on public money. These infrastructures, once built, begin to sell their services to customers (in the form of power and pills, respectively). Through this revenue stream, the companies that undertook the burden of building and maintaining the facilities (in the case of nuclear power), or developing, researching, and testing (in the case of drugs) pay back the startup money they required to make those advances.

This is also the model used by the game industry around the world today. Games are launched to the public with both a "box fee" (the price the consumer pays in the store for the software) and then a monthly subscription fee. For example, upon release, "World of Warcraft" cost approximately USD $50 US dollars, and also costs users around USD $15 a month to play. These fees mean that Blizzard Entertainment, the parent company that financed the creation of the project, has seen its money back and more. Blizzard continues to use profits from "World of Warcraft" not only to pay the aforementioned support staff, but to fund new projects as well. It is important to remember though that it does take time to see this return on investment – typically twelve months at a minimum if a game is a large commercial success. If it is not, this process can take much longer. This does present a valid and established format for funding agencies to consider when making choices about funding large projects of this nature.

5. Conclusion

The next tool for social science experimentation should allow for macro level, generalizable scientific research. In the past, devices such as rat mazes, petri dishes, and supercolliders have been developed when scientists needed new tools to do research. We believe that virtual worlds are the modern equivalent to supercolliders for social scientists, and feel they should be the next area to receive significant attention and funding. The advantages provided by virtual worlds research outweigh the costs. Virtual worlds allow for societal level research with no harm to humans, incorporate large numbers of experiments and participants, and make long term and panel studies possible.

Virtual worlds do have some drawbacks; they are expensive and time consuming to build. These obstacles can be overcome, however, by adopting the models of revenue and maintenance practiced by the current game industry. The returns from virtual worlds being used as scientific tools could reach levels that would self fund future research for decades to come. At the outset, however, an initial investment from funding agencies appears to be necessary.

References:

Bell, M./Castronova, E. and Wagner, G.G. (2008): Virtual Assisted Self Interviewing (VASI): An Expansion of Survey Data Collection Methods to the Virtual Worlds by Means of VDCI. RatSWD Working Paper No. 42.

Castronova, E. et al. (2008): A Test of the Law of Demand in a Virtual World: Exploring the Petri Dish Approach to Social Science. International Journal of Gaming and Computer-Mediated Simulations (2), 1-14.

Castronova, E. (2007): Exodus to the Virtual World How Online Fun is Changing Reality. New York.

Castronova, E. (2005): Synthetic Worlds. Chicago.

Castronova, E and Falk, M. (2009): Virtual Worlds: Petri Dishes, Rat Mazes, Supercolliders. Games and Culture 4 (4), 396-407.

Chesney, Th./Chuah, S.-H. and Hoffmann, R. (2007): Virtual world experimentation: An exploratory study. Nottingham University Business School – Industrial Economics Division Occasional Paper Series No. 2007-21. Nottingham.

Giles, J. (2007): Life's a Game. Nature 445 (7123), 18-20.

Harrison, G.W. and List, J.A. (2004): Field experiments. Journal of Economic Literature 42 (4).

Hausman, J.A. and Wise, D.A. (1985): Social Experimentation. Chicago.

List, J.A. (2008): Homo experimentalis evolves. Science 321 (5886), 207-208.

Miller, G. (2007): The Promise of Parallel Universes. Science 317 (5843), 1341-1343.

10. Qualitative Interviewing of Respondents in Large Representative Surveys

Olaf Groh-Samberg and Ingrid Tucci

Contact:

Olaf Groh-Samberg
BIGSSS, University of Bremen
FVG-West, Wiener Straße
28359 Bremen
Germany
e-mail: ogrohsamberg[at]bigsss.uni-bremen.de

Ingrid Tucci
Sozio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin
Germany
e-mail: itucci[at]diw.de

Abstract

Large representative surveys are using mixed methods to an ever-increasing degree. Biomarkers, register data, and experiments, for example, provide different types of data that can be linked with survey data. The use of qualitative interviewing of participants in longitudinal surveys is, however, still rare in the social sciences. Yet qualitative methods have proven just as valuable as quantitative methods in providing insights into social reality by reflecting the multidimensionality of individual life courses and lived realities. Furthermore, in-depth interviews can provide a better understanding of individual decision-making processes and behavior resulting from more or less unconscious strategies. They also provide insights into decisive turning points in people's lives. Finally, by linking quantitative and qualitative data, the reliability of longitudinal information can be analyzed thoroughly in terms of accuracy as well as meaningfulness.

Keywords: mixed methods, qualitative data, longitudinal data, life course
JEL Code: C81, C83, Z13

1. Introduction

In the social and behavioral sciences, the use of mixed methods to address a particular research question typically involves a combination of quantitative and qualitative methodologies (Brannen 2005; Bryman 2006; Tashakkori and Teddlie 2003). As an increasing range of data becomes available for scientific research – as documented throughout this publication and in the Working Paper Series of the German Data Forum (RatSWD) – the possibilities for mixed method approaches are growing. However, the use of mixed methods to link data from large representative surveys to qualitative data is still rare. A recent trend in longitudinal surveys worldwide consists of the linkage of survey data with data from different sources using diverse methodologies. For example, birth cohort studies or household panels like BHPS,[1] HILDA,[2] PSID,[3] and SOEP,[4] are collecting biomarkers, objective health measures, data from experiments, daily experience sampling or register and institutional context data to survey respondents (see the respective chapters in this publication and, e.g., the new UK Household Longitudinal Study Understanding Society, or UKHLS). In this context of methodological innovations of longitudinal surveys, conducting in-depth qualitative inter-

1 British Household Panel Study.
2 Household, Income and Labour Dynamics in Australia.
3 US Panel Study of Income Dynamics.
4 German Sozio-Economic Panel (SOEP, *Sozio-oekonomisches Panel*).

views with sub-samples of respondents is one important and promising, yet only recently developing issue.

Up to now, qualitative methods have been used primarily with quantitative data to "embellish" analyses (Mason 2006a). However, mixed methods approaches in the sense of a triangulation of quantitative and qualitative data collected from the same respondents might help to understand the mechanisms underlying human behavior and individual life courses (e.g., Giele and Elder 1998). This is particularly true with respect to individual decision-making processes, coping strategies, and biographical "turning points," i.e., events or experiences that play a decisive role an individual's life course by correcting trajectories (Abbot 1997). The importance of decision-making is not only central to the so-called *rational actor model* that has become a common reference model in the economic and social sciences and is typically associated with the large-scale quantitative data analysis (Goldthorpe 2000); it is even considered a broader "unifying framework" for the behavioral sciences (Gintis 2007).

However, as quantitative research along these lines only observes the contexts, determinants, and outcomes of individual decisions – which are measured at least indirectly by means of proxy information – the decision-making process itself can only be modeled in a "reduced form" due to the lack of information on what is really going on in the individual's mind. This is exactly where qualitative in-depth interviews with sub-samples of survey respondents offer possibilities for new research prospects. Qualitative interviews may provide insights into how people select relevant information, what relevance they assign to them, and how their values, attitudes, perceptions, states of knowledge, and conscious as well as unconscious strategies are shaped by and shape their behavior.

Thus, qualitative methods can provide insights into something that still remains a "black box" for quantitative methods that aim to connect measured "inputs" with measured "outcomes" of human decisions and behavior and strive to establish a "causal link" by testing the theoretically derived hypotheses. From a qualitative perspective, this causal link appears to present itself as a dynamic and recursive system of "meanings." This does not mean, however, that the two methods are incompatible (Brannen 2005; Kelle 2001). Rather, by developing explanations of human behavior – for example, regarding educational decisions – the assumptions of quantitative research typically derived from the rational actor model, or any other theory, can be more directly tested, specified, and enriched or even rejected by means of qualitative methods that allow a deeper understanding of how choices come about.

2. The state of the art

Although still rather rare, the linkage of survey data with qualitative interviews seems to have reached scientific maturity, and is being discussed increasingly within the scientific community (Tashakkori and Teddlie 2003). Although there are still forces at work promoting the separation of quantitative and qualitative methods – separate training courses, academic journals, funding schemes, and university chairs – efforts are also underway seeking to actively press forward with mixed method research (e.g., Bryman 2006; Mason 2006b).

It has become apparent that mixed methods are not a third way, or even a third methodology in their own right, and that there exists a broad variety of means by which mixed method approaches can be rationalized and employed in empirical research. A meta-analysis by Bryman (2006) of more than 200 research projects employing mixed methods reveals that mixed methods are mainly employed in sociology, and that they combine self-administered questionnaire surveys with semi-structured interviews to address specific research questions. Mixed methods are typically used to produce "complementary" data or to "enhance" data, facilitating the examination of different perspectives or different aspects of a particular research question. However, there is no strict methodology that determines how different methods should be linked. Rather, there are good arguments for designing and linking mixed methods based on theoretical principles in order to produce non-redundant and non-trivial results (Kelle 2001).

Mixed methods approaches were formerly used primarily in larger-scale research projects aiming to explore new, uncharted research fields. The seminal work of Marie Jahoda, Paul Lazarsfeld et al. (1933) "Marienthal: The Sociography of an Unemployed Community" dealt with the challenges posed by the external economic shock of mass unemployment during the 1930s. The sociology and psychology of the time was entirely incapable of predicting how modern society might respond to such a shock, so the research team attempted to collect as wide a variety of data as possible, ranging from the observation of walking speed, conventional household interviews to content analysis of school essays. Once testable concepts had been produced – such as the concept of individual stages of unemployment experiences – they could then easily be tested using standard quantitative methods or more focused qualitative interviews from predefined samples. This gave rise to Lazarsfeld's idea that qualitative methods could be used to *develop hypotheses* and that quantitative methods could be used to *test hypotheses*. Following this idea, mixed methods research designs often use qualitative interviews and ethnographic research to develop a hypothesis, and survey questionnaires to test the hypothesis. However, the strict two-stage model of sequentially combining qualitative and quantitative methods has not become widespread (Leech and Onwuegbuzie 2009; Creswell et al. 2003). Rather, many larger mixed methods research projects use qualitative methods to supplement quantitative surveys in order to gain a fuller understanding of the "real lives" of the individuals and households surveyed (e.g., Portes and Fernández-Kelly 2008; Mayer and Schulze 2009a; 2009b).

3. The unique potential of qualitative projects based on longitudinal survey respondents

In some sense, longitudinal surveys such as household panel or birth cohort studies can be said to follow in the tradition of Jahoda et al. (1933) in establishing a large survey to analyze how households adapt to social and economic changes and in turn contribute to social change. Longitudinal surveys provide a constantly expanding body of diverse data and are therefore becoming multiple or mixed method enterprises to an increasing degree. Conducting qualitative interviews with long-term survey respondents provides a unique opportunity for a real triangulation of different types of data on people's life courses. In long-running longitudinal studies, it is possible to conduct biographical interviews with long-term respondents for whom more than a decade of prospectively collected panel data are available. In principle, the longitudinal data can also be linked with register data from employment or social insurance agencies.

Triangulations like these would make it possible to thoroughly analyze the validity, reliability, and meaningfulness of panel data. Biographical crises or "turning points" in the life course as reported in qualitative interviews can be checked against the standardized yearly measures collected in longitudinal surveys (e.g., life satisfaction). Is it possible to detect biographical crisis through quantitative longitudinal data? Are respondents able to remember negative events like unemployment or the timing of a divorce? Does the use of combined methodologies affect non-response behavior (item non-response as well as partial unit non-response or panel attrition)?

Mixed method research designs are often used for validation purposes: this is the case with qualitative interviews or experiments being used to validate and/or improve measures in survey questionnaires (e.g., Dohmen et al. 2010 for measuring risk aversion). Cognitive interviewing has been developed as a qualitative tool for this purpose (Willis 2005). Moreover, by drawing on a large ongoing survey, one can systematically select respondents who appeared to be particularly interesting in the quantitative analysis for qualitative interviews. A common feature of such designs is the construction of typologies by clustering survey data and then selecting "representative" respondents for each cluster, or by selecting extreme cases or even outliers for more in-depth analysis (see Portes and Fernández-Kelly 2008 for an outlier analysis).

Apart from investigating the methodological effects arising from the type of data, qualitative interviewing of respondents to longitudinal surveys allows insights in a wide range of particular research questions, such as school choice, educational and occupational aspirations, and family formation. Qualitative interviews can be carried out with entire households and address issues such as family relations within and across households, social networks, perceptions of neighborhoods, schools, employers, and how these shape life goals and individual behavior. However, these rich opportunities have only recently entered the research agenda of longitudinal surveys.

4. Review of qualitative projects based on longitudinal survey respondents

To date, very few projects have been carried out involving qualitative interviews with respondents to longitudinal surveys, but a growing number of such projects have started recently or are currently under planning:

- For the German context, about three dozen interviews were conducted with respondents from the 1971 birth cohort of the German Life History Study (GLHS). Using narrative interviews, Mayer and Schulze (2009a) used a "modest mixed-methods strategy" to analyze the life courses of this generation in West and East Germany and, in another study, to study parenthood processes in order to provide evidence of mechanisms resulting in delayed family formation (Mayer and Schulze 2009b: 12).
- In a project at the University of Manchester on interactions between and within generations, data from the English Longitudinal Study of Ageing (ELSA) were linked to qualitative interviews of between 25 and 30 respondents and approximately 20 of their descendants.[5] The goal of the study was to understand intergenerational transfers and communication and the role played by older people.
- In a project conducted at the Center for Longitudinal Studies at the University of Manchester, qualitative interviews are planned with about 180 respondents (aged 50) from the 1958 British Birth Cohort Study in order

5 More details on this project are available at http://www.socialsciences.manchester.ac.uk/realities/research/generations/

to understand the driving forces and the dynamics underlying voluntary social engagement.

- In the UK, qualitative interviews are planned for the new UKHLS.[6]
- In the US, Portes and Fernández-Kelly (2008) also used mixed methodologies to analyze data from the Children of Immigrants Longitudinal Study (CILS). They conducted narrative interviews with 50 second-generation youths and their families to understand how young respondents have coped with disadvantages during their childhood and teen years and to examine their educational success.
- Also in the US, researchers linked data from the Women's Employment Study (WES) with qualitative data gathered from a sub-sample of the survey's respondents (approximately 70) in order to analyze processes of union formation among low-income women and to formulate hypotheses that can be tested by the use of panel data (see Seefeldt 2008).[7]
- Researchers have used mixed methodologies on data from the South African KwaZulu-Natal Income Dynamics Study (KIDS) in order to understand the factors explaining transitions into or out of poverty (Adato et al. 2006). Qualitative data was collected on members of eight households selected from this first large-scale longitudinal study of household poverty in South Africa.

5. Challenges

Linking qualitative in-depth interviews to quantitative surveys poses new challenges. First of all, ethical and data protection issues have to be considered and resolved (Leahey 2007). For legal reasons, survey respondents have to declare their willingness to participate in the survey, and this declaration should explicitly include their agreement to participate in personal in-depth interviews. Moreover, respondents need to understand exactly how qualitative interviews – or the transcript, audio, or video file – will be linked with the quantitative microdata.

6 For more details, see http://www.understandingsociety.org.uk/design/features/qualitative.aspx as well as http://www.understandingsociety.org.uk/news/latest/ and http://www.esrc.ac.uk/ESRCInfoCentre/research/resources/UKHLS.aspx

7 More details on this project are available at http://cairo.pop.psu.edu/allen/Project.cfm?ProjectID=189.

For longitudinal survey respondents, time-consuming in-depth interviews may negatively affect survey participation, and requests to divulge intimate biographical details could impair the respondent's relationship to the interviewer. From what we know so far about the effects of introducing new and more demanding kinds of surveying in ongoing longitudinal studies, they seem to strengthen rather than weaken respondents' personal commitment to the survey.

An important challenge in developing the social science research infrastructure in the future relates to the rules of access to qualitative data on survey respondents. Those responsible for managing longitudinal surveys need to establish working models that can provide external researchers the opportunity to interview respondents.

6. Recommendations

The inclusion of qualitative in-depth interviewing in the repertoire of data collection methods used in sample surveys is a highly promising innovation in terms of both methodological and substantial research. However, there is still a long way to go in laying the foundations and exploring the possibilities and limits of such an approach.

- Theory & methodology: more extensive use of qualitative methods in surveys should be based on theoretical and methodological proposals that guide the triangulation of qualitative and quantitative methods.
- Ethics, data protection, and access: ethical and data protection issues need to be addressed. Rules for access to samples of respondents should be established.
- Exploration: the possibilities and problems of conducting semi-structured and biographical interviews should be explored with rather small test samples of long-term survey respondents, focusing on methodological issues of "triangulating" life courses.

References:

Abbott, A. (1997): On the Concept of Turning Point. Comparative Social Research 16/1997, 89-109.

Adato, M./Lund, F. and Mhlongo, P. (2007): Methodological Innovations in Research on the Dynamics of Poverty: A Longitudinal Study in KwaZulu-Natal, South Africa. World Development 35 (2), 247-263.

Brannen, J. (2005): Mixing methods: the entry of qualitative and quantitative approaches into the research process. International Journal of Social Research Methodology 8 (3), 173-84.

Bryman, A. (2006): Integrating quantitative and qualitative research: how is it done? Qualitative Research 6 (1), 97-113.

Creswell, J.W./Plano-Clark, V.L./Gutmann, M.L. and Hanson, W.E. (2003): Advanced mixed methods research designs. In: Tashakkori, A. and Teddlie, Ch.B. (Eds.): Handbook of mixed methods in social and behavioral research. Thousand Oaks.

Dohmen, Th./Falk, A./Huffman, D./Sunde, U./Schupp, J. and Wagner, G.G. (2010): Individual Risk Attitudes: New Evidence from a Large, Representative Experimentally-Validated Survey. Journal of the European Economic Association 7. [Forthcoming].

Giele, J.Z. and Elder, G.H. (Eds.) (1998): Methods of Life Course Research: Quantitative and Qualitative Approaches. Thousand Oakes/London/New Delhi.

Gintis, H. (2007): A framework for the unification of the behavioral sciences. Behavioral and Brain Sciences 30 (1), 1-61.

Goldthorpe, J.H. (2000): The Quantitative Analysis of Large-scale Data Sets and Rational Action Theory: For a Sociological Alliance. In: Goldthorpe, J.H.: On Sociology: Numbers, Narratives, and the Integration of Research and Theory. Oxford University Press.

Jahoda, M./Lazarsfeld, P.L. and Zeisel, H. (1933): Marienthal: The Sociography of an Unemployed Community – Translated by the authors with John Reginait and Thomas Elsaesser. Chicago 1971.

Kelle, U. (2001): Sociological Explanations Between Micro and Macro and the Integration of Qualitative and Quantitative Methods. Forum: Qualitative Social Research 2 (1). http://www.qualitative-research.net/index.php/fqs/article/view Article/966/2108. [Last visited: 5/19/2010].

Laurie, H. and Sullivan, O. (1991): Combining qualitative and quantitative methods in the longitudinal study of household allocations. Sociological Review 39 (1), 113-139.

Leahey, E. (2007): Convergence and confidentiality? Limits to the implementation of mixed methodology. Social Science Research 36 (1), 149-158.

Mason, J. (2006a): Six strategies for mixing methods and linking data in social science research. Real Life Methods Working Paper.

Mason, J. (2006b): Mixing methods in a qualitatively driven way. Qualitative Research 6 (1), 9-25.

Mayer, K.U. and Schulze, E. (2009a): Die Wendegeneration. Lebensverläufe des Jahrgangs 1971. Frankfurt am Main/New York.

Mayer, K.U. and Schulze, E. (2009b): Delaying Parenthood in East and West Germany. A Mixed-Methods Study of the Onset of Childbirth and the Vocabulary of Motives of Women of the Birth Cohort of 1971. In: Andersson, G./Bernardi, L./Kulu, H. and Neyer, G. (Eds.): The Demography of Europe: Trends and Perspectives. Berlin.
Leech, N.L. and Onwuegbuzie, A.J. (2009): A typology of mixed methods research designs. Quality and Quantity 43 (2), 265-75.
Portes, A. and Fernández-Kelly, P. (2008): No Margin for Error: Educational and Occupational Achievement among Disadvantaged Children of Immigrants. The Annals of the American Academy of Political and Social Science 620 (1), 12-36.
Seefeldt, K.S. (2008): Working after welfare: how women balance jobs and family in the wake of welfare reform. W.E. Upjohn Institute for Employment Research, Michigan.
Tashakkori A. and Teddlie Ch.B. (Eds.) (2003): Handbook of Mixed Methods in Social and Behavioural Research. Thousand Oaks.
Willis, G.B. (2005): Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks.