

RatSWD Working Paper Series

www.ratswd.de

187

RatSWD

Rat für Sozial- und
WirtschaftsDaten

Forschungsdaten in der Psychologie:
Disziplinspezifische und
disziplinübergreifende Bedürfnisse

Zusammenfassung des Forums (2) der
5. Konferenz für Sozial- und Wirtschaftsdaten

Erich Weichselgartner

Dezember 2011

GEFÖRDERT VOM



Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienenen empirischen Forschungsarbeiten sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Forschungsdaten in der Psychologie: Disziplinspezifische und disziplinübergreifende Bedürfnisse

Zusammenfassung des Forums (2) der 5. Konferenz für
Sozial- und Wirtschaftsdaten

Erich Weichselgartner

Universität Trier
Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)
wga@zpid.de

Keywords: Psychologie, Forschungsdaten, Längsschnittstudien

Überblick

Erkenntnisgewinn in der Psychologie als empirischer Wissenschaft ist unmittelbar an die Produktion und Verarbeitung von Daten gebunden, welche das Erleben und Verhalten von Menschen abbilden. Im Gegensatz zu den Forschungsergebnissen erhalten die Forschungsdaten bislang keine wissenschaftliche Aufmerksamkeit und Anerkennung und sind häufig schwer verfügbar. Die umfassende Dokumentation, Aufbereitung und Vorhaltung von Daten für die Nachnutzung ist in der Psychologie noch nicht selbstverständlicher Bestandteil des Forschungsprozesses. Allerdings weisen Entwicklungen in den letzten Jahren auf ein Umdenken in der Fachkultur hin.

Das Forum *Forschungsdaten in der Psychologie: Disziplinspezifische und disziplinübergreifende Bedürfnisse* im Rahmen der 5. Konferenz für Sozial- und Wirtschaftsdaten (KSWD) in Wiesbaden hatte zum Ziel, existierende Initiativen zum Forschungsdatenmanagement in der Psychologie vorzustellen sowie Anforderungen von Datenprovidern und Datennutzern zu reflektieren. Da die Idee der Weitergabe von Forschungsdaten (*data sharing*) und das damit zusammenhängende Forschungsdatenmanagement in der Psychologie noch nicht zur Fachkultur gehört, wurden dessen Vorteile herausgearbeitet, mögliche Anreize (*incentives*) vorgestellt und dessen Bedeutung für die disziplinübergreifende Zusammenarbeit thematisiert.

So wollte Frau Dr. Jule Specht von der Wilhelms-Universität Münster referieren, wie mit Hilfe großer Panelbefragungen aus dem Kontext der Wirtschafts- und Sozialwissenschaften Fragestellungen untersucht werden können, die sonst nicht analysierbar wären. Dies sollte anhand zweier Studien, welche die Daten des Sozio-ökonomischen Panels (SOEP) nutzen, exemplarisch aufgezeigt werden. Leider musste Frau Specht ihre Teilnahme am Forum kurzfristig absagen; die Ergebnisse ihrer Analysen können aber in zwei diesbezüglichen Publikationen nachgelesen werden (Specht, Egloff & Schmukle, 2011a,b).

Prof. Dr. Armin Günther vom Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) in Trier gab im ersten Forumsbeitrag einen Überblick über das Forschungsdatenmanagement in der Psychologie. Er zeigte auf, welche Ansätze, psychologische Forschungsdaten zugänglich zu machen, es trotz Widerstand aus dem Fach bereits gibt und versuchte einen Ausblick auf mögliche Entwicklungserspektiven des Forschungsdatenmanagements in der Psychologie zu geben. Mit PsychData, dem Forschungsdatenzentrum des ZPID für die Psychologie, steht eine Infrastruktur zur Verfügung, die speziell auf die Bedürfnisse des Faches abgestimmt ist. Primärforscher erhalten die Möglichkeit, ihre quantitativen Daten in einer standardisierten, den Besonderheiten der psychologischen Forschung angepassten Form zu

archivieren und bereitstellen zu lassen; andererseits werden Sekundärforscher auf ihrer Suche nach Datengrundlagen fündig.

PD Dr. Beatrice Rammstedt vom Leibniz Institut für Sozialwissenschaften (GESIS) in Mannheim stellte die Vorteile der Datenveröffentlichung sowohl von *large-scale surveys* als auch von kleineren Datensätzen aus der Sicht des Datenproviders wie auch aus der des Datennutzers dar und ging auf begleitende Fragen der Forschungsethik ein.

Prof. Dr. Rainer Riemann von der Universität Bielefeld arbeitete in seinem Beitrag zunächst am Beispiel der Bielefelder Zwillingsdaten einige Gründe für die Komplexität psychologischer Datensätze heraus, um zu verdeutlichen, vor welchen Schwierigkeiten Psychologinnen und Psychologen stehen, wenn sie derartige Datensätze nachnutzen wollen. Im zweiten Teil seines Vortrages skizzierte er, welche Anforderungen an Datenverwaltungswerzeuge zu stellen sind, die die Nutzung, Qualitätssicherung und Nachnutzung psychologischer Daten gleichermaßen unterstützen könnten.

Prof. Dr. Sabine Walper von der Ludwig-Maximilians-Universität München referierte zum Beziehungs- und Familienpanel *pairfam* und illustrierte dessen reichhaltige Analysemöglichkeiten für psychologische Fragestellungen am Beispiel der Partnerschaftsforschung. Obwohl aus Gründen der Erhebungskosten durchgängig auf sehr kurze Skalen zurückgegriffen werden musste, lassen sich attributions- und bindungstheoretisch fundierte Hypothesen zu Einflüssen auf die Partnerschaftsqualität überprüfen.

Beiträge und Diskussion

PD Dr. Erich Weichselgartner vom Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) in Trier wies in seinen einleitenden Worten darauf hin, dass die Psychologie erstmals mit einer eigenen Veranstaltung auf einer Konferenz für Sozial- und Wirtschaftsdaten vertreten sei. Der Rat für Sozial- und Wirtschaftsdaten (RatSWD) erhofft sich von der Psychologie eine Befruchtung anderer Disziplinen. Professor Dr. Gert Wagner hatte als Vorsitzender des RatSWD in seinem KSWD-Eröffnungsvortrag explizit auf diesen Aspekt hingewiesen. Weichselgartner merkte an, dass die Psychologie in Bezug auf die Methodenentwicklung weit fortgeschritten sei und messtheoretisch fundierte Skalierungsverfahren entwickelt habe, von denen andere Disziplinen profitieren könnten. Ob aber die Forschungsdaten der Psychologie für die Nachnutzung durch Dritte geeignet seien, sei eines der Diskussionsthemen des Forums.

Die ersten beiden Forumsbeiträge widmeten sich der Ist-Situation in der Psychologie und beleuchteten deren strukturelle und historische Ursachen. Bei den beiden nachfolgenden Beiträgen standen konkrete Aspekte des Forschungsdatenmanagements im Vordergrund.

1. Prof. Dr. Armin Günther: Forschungsdatenmanagement in der Psychologie: Rahmenbedingungen, Ansätze, Perspektiven.

Günther sieht neuere technische Entwicklungen und sich daraus ergebende Praktiken in anderen Disziplinen als impulsgebend für die Psychologie. Seiner Auffassung nach haben Digitalisierung und Vernetzung in den letzten zwei Jahrzehnten völlig neue technische Möglichkeiten geschaffen, Daten nicht nur zu generieren, zu verarbeiten und zu speichern, sondern auch weiterzugeben und in unterschiedlichen Kontexten zu nutzen. Dies gilt grundsätzlich auch für Forschungsdaten. In vielen wissenschaftlichen Disziplinen entwickeln sich daher neue Praktiken des Datenmanagements, die unter anderem darauf ausgerichtet sind, der *scientific community* Forschungsdaten zugänglich und nutzbar zu machen (*data sharing*). Die Psychologie steht hierbei allerdings noch weitgehend am Anfang. Als prototypisch für die anfänglichen Reaktionen auf eine Anfrage nach der Veröffentlichung von Forschungsdaten zitierte Günther einen konkreten Fall aus der Praxis:

„In unserem Projekt stecken mehr als 15 Jahre Arbeit und mehr als zwei Millionen DM. Das Projekt läuft noch. Darf ich Ihre Anfrage so verstehen, dass wir die Daten erheben und andere dann darauf zugreifen, keine Arbeit hatten, sie auswerten und dann publizieren? Das kann doch wohl nicht ernsthaft gemeint sein, oder?“

Im weiteren Vortrag ging Günther der Frage nach, welche allgemeinen und psychologiespezifischen Rahmenbedingungen die Entwicklung entsprechender Praktiken und Strukturen in der Psychologie prägen und, wie mit dem Zitat dargelegt, erschweren. Am Beispiel des Forschungsdatenzentrums PsychData zeigte er auf, welche Ansätze, psychologische Forschungsdaten zugänglich zu machen, es dennoch bereits gibt und versuchte einen Ausblick auf mögliche Entwicklungsperspektiven des Forschungsdatenmanagements in der Psychologie zu geben.

In der *Diskussion* wurde die Frage aufgeworfen, ob die Psychologie von ihrem Forschungsansatz her an Nachnutzung von Daten bzw. an großen Datenmengen oder an Datenaggregierungen überhaupt interessiert sei. Günther antwortete, dass je nach Forschungsthema, einzelne Datenpunkte (differentielle Psychologie) oder große repräsentative Stichproben (psychologische Diagnostik) wichtig sein können. In der Entwicklungspsychologie seien großangelegte Längsschnittstudien schon jeher von Bedeutung. Generell sei die Psychologie daran interessiert, mit vielen Untersuchungsdesigns und mit vielen unterschiedlichen Herangehensweisen Effekte belastbar zu replizieren.

Riemann wies darauf hin, dass auch in der Psychologie quantitative Metaanalysen verbreitet seien. Dem hielt Günther entgegen, dass gerade das vergebliche Bemühen von Wicherts (Wicherts et al, 2006), Daten für eine Metaanalyse zu erhalten, das Thema *data sharing* in der Psychologie und darüber hinaus ins Blickfeld der Aufmerksamkeit gerückt habe.

Rammstedt hakte beim Thema „Repräsentativität“ ein. Sie fragte, inwieweit die Psychologie dieses Thema ausblende, weil sie Schwierigkeiten mit der Erhebung entsprechender Datensätze habe. So müsse man doch sehr in Frage stellen, ob sich an nationalen Stichproben gefundene allgemeinpsychologische Gesetzmäßigkeiten auf internationale Verhältnisse übertragen ließen. Günther antwortete, dass dies die Frage der Forschungslogik aufwerfe, die man aus Zeitmangel nicht diskutieren könne, aber dass sich die Datenlage bei einem zunehmenden *data sharing* diesbezüglich verbessern könnte.

Aus dem Publikum wurde die Frage gestellt, ob Fachzeitschriften Vorgaben zum *data sharing* machen. Laut Günther sei das in der Psychologie nicht verbreitet – im Gegenteil, dort, wo es praktiziert wurde, sei es wieder abgeschafft worden. Nicht zuletzt seien auch die Gutachter überfordert gewesen. Deshalb gäbe es Modelle, bei denen Zeitschriften das Hinterlegen der Daten bei Forschungsdatenzentren verlangen, die auf diese Aufgabe spezialisiert sind. Weichselgartner wies noch darauf hin, dass auch die deutschen Forschungsförderer keine Verpflichtung auf Datenweitergabe vorsehen, im Unterschied zu einigen in Großbritannien und den USA.

2. PD Dr. Beatrice Rammstedt: Möglichkeiten der Sekundärnutzung von Datensätzen.

In den empirischen Sozialwissenschaften hat die Archivierung und Bereitstellung von Primärdatensätzen eine lange Tradition. So wird von den meisten *large-scale surveys* erwartet, dass diese umfassend archiviert und dokumentiert werden. Die Nutzung dieser Datensätze erfolgt in erster Linie durch Sekundärnutzer; die Primärforscher sind zugleich die Datenprovider. Dieser Trend zur Veröffentlichung der eigenen Datensätze erreicht inzwischen auch kleinere Datensätze, die primär von dem Provider/Forscher genutzt werden, anschließend aber im Archiv der wissenschaftlichen Öffentlichkeit zur Verfügung gestellt werden. Somit wirkt sich der Trend auch auf benachbarte Disziplinen wie z.B. die Psychologie aus, die in der Regel kleinere Datensätze produzieren und analysieren. Allerdings geht die psychologische Forschung beim Erzeugen von Datensätzen mitunter recht eigenwillige Wege. Rammstedt verwies auf Allports (1958) klassische Kritik an der Persönlichkeitspsychologie:

„*Each assessor has his own pet units and uses a pet battery of diagnostic devices.*“

Als Antidot zu dieser individualistischen Vorgehensweise stellte Rammstedt die Vorteile einer Datenveröffentlichung sowohl aus der Sicht des Datenproviders als auch aus der des Datennutzers dar und ging auf begleitende Fragen der Forschungsethik ein.

In der *Diskussion* kam sofort die Frage auf, was denn in der Psychologie im Vergleich zu anderen Fächern bereitgestellt werden soll. Gibt es über-

haupt Daten mit hohem Nachnutzungspotential wie in den Sozial- und Wirtschaftswissenschaften? Rammstedt antwortete, dass man die Psychologie differenziert betrachten müsse und es durchaus positive Beispiele gebe, nicht zuletzt das in diesem Forum vorgestellte Familienpanel *pairfam*. Riemann verwies auf ein besonders faszinierendes Beispiel, bei dem man ursprünglich die Nachnutzung nicht im Auge hatte: Es gab in Schottland im Jahr 1931 eine Gesamterhebung des IQs aller Schülerinnen und Schüler, die an einem Tag in der Schule waren. Dieser Datensatz habe im Bereich der Untersuchung des Zusammenhangs von Gesundheit und Intelligenz eine enorme Bedeutung erlangt. Es gebe also in der Psychologie viele hochinteressante Fragestellungen, aber auch solche, die nur einen spezifischen Forscher interessieren und wo man sagen müsse, die Archivierung der zugehörigen Datensätze erscheine nicht angemessen.

Walper merkte im Zusammenhang mit der Skalenkonstruktion an, dass sie an vielen Stellen sehr dankbar gewesen wäre, wenn sie Zugang zu Primärdatensätzen gehabt hätte, weil das die Entwicklung von Kurzformen erleichtert hätte. Nur mit Hilfe eines reichhaltigen Datenpools erhalte man in aller Regel ausreichende Informationen zur Selektion von Items, zum Beispiel, ob sie in speziellen Subpopulationen funktionieren. Für sie war das SOEP eine hervorragende Referenz.

Rammstedts Ausführungen, dass diejenigen Forscher, die die Daten erzeugen, gar nicht die Hauptnutzer sein müssen, wurde dahingehend diskutiert, ob man zwischen Datenprovidern ohne wissenschaftlichem Interesse und Datennutzern mit wissenschaftlichem Interesse unterscheiden könne. Datenprovider könnten das als Beruf machen und vielleicht sogar Entgelt für ihre Dienstleistung verlangen, ohne sich wissenschaftliche Meriten verdienen zu wollen. Die Datennutzer wären dann die eigentlichen Forscher, die die Daten wissenschaftlich ausschlachten. Als positive Gegenbeispiele wurden großangelegte Studien wie NEPS, *pairfam* oder PISA angeführt, wo Provider- und Nutzer-Rollen vereint sind.

Walper wies auf einen weiteren Vorteil der Bereitstellung von Forschungsdaten in der Psychologie hin: Qualifikationsarbeiten könnten so schneller abgeschlossen werden.

Rammstedt sieht die Bereitschaft zum Bereitstellen von Forschungsdaten mit dem Vorhandensein von Anreizsystemen verknüpft. Dazu gehörten im akademischen Betrieb die Zitationen. Diese könne man durch Vergabe von persistenten Identifikatoren auch für Datensätze ermöglichen (Stichworte: DOI, URN).

3. Prof. Dr. Rainer Riemann: Datennachnutzung in der Psychologie: Bielefelder Zwillingssdaten.

Die Nachnutzung psychologischer Daten bereitet interessierten Wissenschaftlern große Schwierigkeiten, wenn der herangezogene Datensatz sehr komplex ist. Riemann stellte am Beispiel der Bielefelder Zwillingssdaten einige Gründe für die Komplexität psychologischer Datensätze heraus, um zu verdeutlichen, vor welchen Schwierigkeiten Forscher stehen, wenn sie derartige Datensätze nachnutzen wollen. Da Zwillinge vergleichsweise schwer zu rekrutieren sind, werden pro Person viele Variablen aus unterschiedlichen Datenquellen erhoben: Selbstberichtsdaten (Fragebogen), Testdaten, Bekanntenbeurteilungen, physiologische Daten etc. Des Weiteren ist nicht-textuelles oder nicht-numerisches Material zu integrieren (z.B. Videoaufzeichnungen). Solche Datensätze werden schnell unübersichtlich und sind von Außenstehenden nur nach längerer Einarbeitung zu nutzen.

Darüber hinaus gibt es in der psychologischen Forschung gelegentlich die Schwierigkeit, dass sensible personenbezogene Daten erhoben werden. Hier müssen Vorgehensweisen entwickelt werden, um diese Daten hinreichend zu anonymisieren. Schon bei der Planung müssen Möglichkeiten und Grenzen der Weitergabe von Daten geprüft werden (Ethikgutachten). Es wird Fälle geben, in denen eine Nachnutzung von Daten gar nicht vertretbar ist.

Im zweiten Teil des Vortrages skizzierte Riemann, welche Anforderungen an Datenverwaltungswerzeuge zu stellen sind, die die Nutzung, Qualitäts-sicherung und Nachnutzung psychologischer Daten gleichermaßen unterstützen können. Stichworte waren hier: (a) Speicherung von Daten und Datenaufbereitungsprozeduren, (b) hierarchischer Zugriff auf Daten, der die Datenstruktur entsprechend den Anforderungen von Wissenschaftlern abbildet und (c) definierte Vorgehensweisen, um die Ergebnisse von Analysen (auch von Nachnutzern) in den Datensatz einzupflegen. Insgesamt müsse es das Ziel der Entwicklung von Werkzeugen sein, die Nutzer von aufwändigen Programmieraufgaben zu entlasten und Daten in der gewünschten Struktur und in wählbaren Formaten automatisch zur Verfügung zu stellen.

In der *Diskussion* war man sich einig, dass Datenarchivierung während des Forschungsprozesses (forschungsbegleitend) bedacht werden müsse. Die retrospektive Archivierung bereite unvergleichlich mehr Mühe.

Auf die Frage, ob bei Bild- und Videodaten ein Datenschutz- bzw. Anonymisierungsproblem bestehe, gab Riemann zu, dies vorab unterschätzt zu haben. Man hätte vor Untersuchungsbeginn darauf achten müssen, sich vertraglich klar und juristisch einwandfrei abzusichern. Es sollte unbedingt der Rat von Experten, z.B. Datenschützern und Ethikbeauftragten, eingeholt werden. Speziell die Weiterverwendung der Videodaten sei im Vorfeld klar zu regeln; die Einschränkung für wissenschaftliche Zwecke biete sich an. Solche Daten, bei denen eine De-Anonymisierung in Frage kommen könnte, sollten gar nicht weitergegeben, sondern auf Rechnern vor Ort ausgewertet werden.

Weitere Faktoren, die die Datennachnutzung von Videomaterial behindern, seien technischer Art. Dies habe man am Anfang deutlich unterschätzt. Technische Probleme des mit analoger Technik erstellten Videomaterials behinderten nicht nur eine Datennachnutzung durch die eigene Arbeitsgruppe, sondern erschwerten auch Qualifikationsarbeiten. Der Zugriff auf die Videoaufzeichnungen sollte deutlich einfacher sein.

Damit weder die Komplexität des Materials noch die Komplexität der Datenstrukturen in der Praxis zu einer unüberwindlichen Hürde für die Datennutzung wird, stimmte die Diskussion Riemanns Forderung zu, dass Tools entwickelt werden müssen, die den Zugriff auch auf komplexe Forschungsdaten erleichtern. Die Wortmeldungen zeigten, dass hier ein erheblicher und akuter Entwicklungsbedarf gesehen wird, ließ aber offen, wer dies leisten und finanzieren soll. Riemann sieht Wissenschaftsorganisationen in der Pflicht, Know-how und Best Practices zu bündeln und an Wissenschaftler zu transferieren.

4. Prof. Dr. Sabine Walper: Das *pairfam*-Projekt als Chance für die psychologische Beziehungs- und Familienforschung.

Seit Ende 2008 finden im Rahmen des DFG-finanzierten *pairfam*-Projekts (Panel Analysis of Intimate Relationships and Family Dynamics) jährliche Längsschnitterhebungen statt, deren Daten reichhaltige Analysemöglichkeiten zu Beziehungen in der Jugend und im (frühen bis mittleren) Erwachsenenalter bieten. Ziel von *pairfam* ist die Stärkung der interdisziplinären Beziehungs- und Familienforschung in Deutschland durch Bereitstellung umfangreicher Daten.

Inhaltliche Schwerpunkte des interdisziplinär entwickelten Befragungsprogramms sind (1) Partnerschaftsbeziehungen, (2) Kinderwunsch und Familienplanung, (3) Intergenerationenbeziehungen zur Herkunftsfamilie und (4) Erziehung der Kinder und deren Entwicklung. Teilnehmer dieser längsschnittlich angelegten Befragung waren in der ersten Welle jeweils rund 4.000 Befragte der drei Geburtskohorten 1991-1993, 1981-1983 und 1971-1973 sowie ggf. deren Partner (Teilnahmequote 50%). Ab der zweiten Erhebungswelle (2009/2010) werden auch die Eltern der Ankerperson sowie ein Kind im Haushalt (im Alter von 8 bis 15 Jahren) befragt.

Walper illustrierte in ihrem Beitrag exemplarisch die Nutzungsmöglichkeiten von *pairfam* anhand von Fragestellungen der Partnerschaftsforschung. Obwohl aus Gründen der Erhebungsökonomie durchgängig auf sehr kurze Skalen zurückgegriffen werden musste, ließen sich beispielsweise attributions- und bindungstheoretisch fundierte Hypothesen zu Einflüssen auf die Partnerschaftsqualität testen. Darüber hinaus gab Walper einen Ausblick auf das Erhebungsprogramm der nächsten Wellen, die weiterhin im Jahresabstand durchgeführt werden, und erläuterte die Zugangsmodalitäten zu den Daten.

Diese sind ausschließlich für wissenschaftliche Analysen bestimmt und werden nach Abschluss eines Nutzungsvertrags vom *pairfam user support* an der TU Chemnitz distribuiert. Die *scientific use files* sind als CD-ROM verfügbar; sollen kontextuelle Rahmenbedingungen hinzugespielt werden (z.B. Regionalindikatoren der Firma microm Consumer Marketing), müssen entsprechende Analysen an speziell gesicherten Arbeitsplätzen in Bremen, Chemnitz, Mannheim und München durchgeführt werden.

In der *Diskussion* wurde akzentuiert, dass man im Unterschied zur verbreiteten Praxis, wo man sich nachträglich für die Bereitstellung der Forschungsdaten entscheidet, bei *pairfam* von vornherein an die Veröffentlichung gedacht habe. Günther fragte, inwiefern dies die Konzeption, die Inhalte und die Methoden der Untersuchung beeinflusst habe. Walper antwortete, dass man ein bestimmtes Design aus Gründen der besseren Nachnutzbarkeit gewählt habe; es soll Anknüpfungspunkte für viele andere liefern, die auch spannende Fragestellungen haben mögen. Trotzdem würden bei den einzelnen Wellen unterschiedliche Schwerpunkte gesetzt, wobei bestimmte Dinge konstant blieben und andere rotierten. Es gelte, einen Kompromiss zwischen Übersichtlichkeit und Komplexität zu finden.

Aus dem Publikum wurde die Frage gestellt, ob es bei Paneluntersuchungen nicht ein methodisches Problem mit dem Rücklauf gebe, insbesondere was die Partner der Ankerpersonen angehe. Walper konzidierte, dass dies ein sehr wichtiger Punkt sei. In der Tat seien die Rückläufe bei älteren Kohorten deutlich höher als in der jüngeren Kohorte. Die Ursache könne sein, dass es sich vorwiegend um neue Partner handle im Vergleich zu denen, die in der Vorwelle da waren, oder dass die Partner nicht zusammen wohnten, was die Ausfüllwahrscheinlichkeit beziehungsweise die Rücksendewahrscheinlichkeit deutlich reduziere gegenüber älteren Kohorten, wo die Partner häufiger zusammen in einem Haushalt wohnten. Dies müsse noch genauer untersucht werden. Die Identitäten der Partner müssten auch exakt abgeglichen werden, damit nicht falsche Kontinuitäten angenommen werden.

Sehr wichtig ist im Zusammenhang mit dem Rücklauf die Panelpflege. Bei *pairfam* werden als Anreize (Incentives) Dankschreiben, Grußkarten, Mitteilungen über Untersuchungsergebnisse und Lotterielose eingesetzt. Begrenzender Faktor sind die Finanzen.

Weichselgartner fragte in Bezug auf die berichtete ausländische Nutzung nach, wie mit dem Sprachproblem umgegangen werde. Walper antwortete, dass man schon von vorneherein versuche, die internationale Nutzung zu ermöglichen. Die Datendokumentation erfolge auf Englisch und auch im Web seien englischsprachige Informationen verfügbar; auch die Erhebungsinstrumente würden zeitnah übersetzt. Für die Kosten sei kein gesondertes Budget eingestellt, was allerdings zur „Selbstausbeutung“ der *pairfam*-Mitarbeiter führe. Da man aber wisse, dass im Ausland die Nutzung von Forschungsdaten schon viel weiter verbreitet sei, erhoffe man sich quasi durch die Hintertür einen zusätzlichen Aufmerksamkeitsgewinn auch im Inland.

Aus dem Publikum kam die Frage, wie hoch der Betreuungsaufwand für die Wissenschaftler sei, die die Daten nutzen. Laut Walper ist dieser nicht sehr hoch, es lasse sich alles telefonisch oder per Email so abwickeln, dass die Nutzer damit klar kämen. Allerdings müssten letztere an einen der vier oben genannten *pairfam*-Standorte kommen, um mit Datensätzen zu arbeiten, bei denen die Gefahr der De-Anonymisierung existiere. Hier bestünden entsprechend intensivere Betreuungsmöglichkeiten.

Günther hakte nach, warum bei den *pairfam*-Nutzern Psychologen nicht so stark vertreten seien wie Forscher aus anderen Disziplinen. Walper meinte, dass dies an der Forschungstradition der Psychologie liegen könne, wo man mit wenigen ausführlichen Indikatoren an kleinen Stichproben arbeite. Mit vielen knappen Indikatoren bei großen Stichproben zu arbeiten erscheine methodisch unsauber. Es seien Vorbehalte da, ob man zum Beispiel mit einem 3-Item-Indikator für feindselige Attribution überhaupt etwas anfangen könne. Deshalb sei es wichtig zu zeigen, dass hier durchaus etwas Sinnvolles herauskommen könne. Mit dem richtigen Ansatz könne man auch psychologische Konstrukte in großen Surveys mit Hilfe kleiner „Päckchen“ erfassen.

Fazit

Insgesamt bewerteten die Teilnehmer des Forums den Status Quo in der Psychologie als nicht zufriedenstellend, da derzeit nur wenige Forscher in der Psychologie ihre Daten der Wissenschaftsgemeinschaft zur Verfügung stellen. Allerdings sind auch die Möglichkeiten, Forschungsdaten in der Psychologie zu publizieren, noch wenig entwickelt. Es wird begrüßt, dass mit *PsychData* der Psychologie erstmals ein disziplingebundenes Forschungsdatenzentrum zur Verfügung steht. Die Diskussion zeigte aber auch, dass die Nachfrage nach diesen Daten in der Psychologie weniger stark ausgeprägt ist als beispielsweise in der Soziologie. Forscher in der Psychologie haben noch Vorbehalte gegenüber der Qualität dieser Daten oder glauben diese für ihre eigenen Forschungsvorhaben nicht nutzen zu können. Die stärkere Beteiligung der Psychologie an großangelegten Panelstudien könnte helfen, die Vorbehalte und Defizite zu überwinden.

Die wesentlichen Punkte der Diskussion werden im Folgenden gebündelt in chronologischer Reihenfolge resümiert:

- Es besteht die Befürchtung, dass andere von eigener Arbeit profitieren.
- Ist die Psychologie an der Nachnutzung von Daten überhaupt interessiert?
- Wird in der Psychologie auf die Repräsentativität von Stichproben geachtet (ist diese überhaupt wichtig)?

- Machen Verlage oder Forschungsförderer Vorgaben zur Datenarchivierung?
- Die Datenarchivierung bei Verlagen wird nicht als sinnvoll erachtet (nicht nachhaltig, Material schlecht erschlossen).
- Gibt es in der Psychologie Daten mit hohem Nachnutzungspotential wie in den Sozial- und Wirtschaftswissenschaften?
- Gibt es Selektionskriterien für die Auswahl erhaltenswerter Datensätze?
- Kann es Datenprovider als Dienstleister geben, die kein eigenes wissenschaftliches Interesse haben?
- Ein großer Datenpool ist für die Skalenkonstruktion besonders wichtig.
- Qualifikationsarbeiten können bei Verfügbarkeit von Datenpools schneller abgeschlossen werden.
- Die Bereitschaft zum Bereitstellen von Forschungsdaten kann mit Anreizsystemen gesteigert werden (z.B. Zitierbarkeit von Datensätzen).
- Die retrospektive Archivierung bereitet unvergleichlich mehr Mühe als die forschungsbegleitende. Beim Untersuchungsdesign sollte die Nachnutzbarkeit im Auge behalten werden.
- Die Nachnutzung personenbezogener Daten ist vorab vertraglich klar und juristisch einwandfrei zu regeln. Datenschützer und Ethikbeauftragte sind zu konsultieren.
- Tools sollten entwickelt werden, die den Zugriff auch auf komplexe Forschungsdaten erleichtern.
- Es besteht ein methodisches Problem mit dem Rücklauf bei Paneluntersuchungen. Die Panelpflege ist sehr wichtig.
- Die ausländische Nutzung von Forschungsdaten ist potentiell sehr hoch, die Datenaufbereitung für die internationale Nutzung ist aber teuer.
- Die Datenbereitstellung erfordert immer auch einen Betreuungsaufwand.
- Es gibt in der Psychologie Vorbehalte bezüglich der Nutzung von Daten, die andere erhoben haben (Qualität, Nützlichkeit).

Literatur

- Allport, G. W. (1958). What units shall we employ? In G. Lindzey (Ed.), *Assessment of human motives* (pp. 238-260). New York: Rinehart.
- Specht, J., Egloff, B., & Schmukle, S. C. (2011a). The benefits of believing in chance or fate: External locus of control as a protective factor for coping with the death of a spouse. *Social Psychological and Personality Science*, 2, 132-137.
- Specht, J., Egloff, B., & Schmukle, S. C. (2011b). Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology*, 101, 862-882.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The Poor Availability of Psychological Research Data for Reanalysis. *American Psychologist*, 61, 726-728.