German Council for Social
and Economic Data (RatSWD)

www.ratswd.de

RatSWD
Working Paper Series

Working Paper No. 88

Review of Spatial Data Needs and Resources

David Owen, Anne Green and Peter Elias

July 2009

Federal Ministry
of Education
and Research

Working Paper Series of the Council for Social and Economic Data (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Contact: Council for Social and Economic Data (RatSWD) | Mohrenstraße 58 | 10117 Berlin | office@ratswd.de

Review of Spatial Data Needs and Resources

David Owen, Anne Green and Peter Elias
Institute for Employment Research, University of Warwick

Summary

Introduction

- This review is intended to inform the ESRC Research Resources Board (RRB) by addressing the question of how the social science research community can take advantage of the emerging plethora of spatial data sources in order to improve the quality of research and more effectively address key research questions.

Methodology

- Such a review is a challenging undertaking given the difficulty of determining where to 'draw the boundaries' around geospatial data and associated research and skills issues, the range of academic disciplines involved, the speed of technological developments and changes in the policy environment. It has involved consultation with those who provide/operate spatial data research resources, as well as users and potential users of spatial data.

The changing environment

- There have been rapid developments in computing in recent years, which have led to enhancements in the ability to handle large amounts of geospatial data.
- The private sector is playing an increasing role in the capture of geospatial data and in developing products and services based on such data.
- There is widespread acceptance that the effective use of geospatial data requires the establishment of a geospatial data framework, which both catalogues the data sets available and provides a means of accessing data.
- The EU Inspire Directive and UK Location Strategy are important developments in this vein.
- Academics will progressively have to adopt the standards applying to the public sector, and will have to update software, data management practices and training.

Availability of geospatial data

- There has been a growth in the availability of geospatial data, but at the same time there has been a growth in demand for such data.
- The rise of neogeography and volunteered geographical information has extended the use of geospatial data to new users, some of whom had no previous interest in or expertise in

handling and interpreting geographical data.

- There is a need for comprehensive and high quality metadata in order to enable users to identify, and make appropriate use of, different data sources.

Knowledge about and use of geospatial data

- Interest in, and use of, geospatial data is spread across a range of social science disciplines.
- Some suppliers of geospatial data believe that detailed knowledge of some key spatially-referenced socio-economic data sources is diminishing over time.
- There is a need for ongoing reinvestment in knowledge and skills if appropriate use of geospatial data is to be maximised.
- There is a broad distinction, and arguably an increasing divide, between those researchers with considerable expertise in terms of knowledge and analytical techniques for using geospatial data on the one hand, and on the other those with a much more limited knowledge and expertise in using geospatial data and applications.
- There is a clear need for a range of geo-spatial data services expressed by respondents to the online enquiry, particularly for advice and guidance on the use of, access to and linking with geospatial data.

Confidentiality and security

- Users face constraints in accessing and making full use of geospatial data due to licensing issues, confidentiality concerns and statistical disclosure control requirements.
- More experienced users are particularly frustrated by lack of detailed spatial codes being made available for use in data analyses.
- Some researchers are likely to find such obstacles off-putting, particularly if they lack guidance in dealing with them.
- There are fewer constraints in dealing with data from 'non-traditional' sources.

Geospatial data quality

- The quality of geospatial data varies in accordance with the source, data collection standards and the form in which it is made available to researchers.
- There are many examples of users 'making do' with data of dubious quality without necessarily realising the impact of this on their analyses and interpretation of results.
- It is important that users are encouraged to have a greater awareness of data quality issues.

Skills

- Technical and substantive skills are needed for effective use of geospatial data.
- The main aim of skills training must be to promote the effective and appropriate use of geospatial data to address substantive research questions.
- A variety of skills and knowledge transfer mechanisms need to be considered, varying from short courses to the provision of some form of centralised or networked service provision.
- Demonstration projects may be a valuable means of illustrating the potential for use of geospatial data.

Recommendations

- The full potential of geospatial data is not being realised by economic and social scientists is not being realised by economic and social scientists at the present time.
- It is our view that, alongside the provision of more training in the use of geospatial resources for research purposes, in the UK the ESRC (Economic and Social Research Council) should take the lead in seeking to establish, in collaboration with other funding bodies, a Geospatial Resources Advisory Service.

1. Introduction and aims

1.1 Background to the review

This review is intended to inform the ESRC (Economic and Social Research Council) Research Resources Board (RRB) by addressing the question of how the social science research community can take advantage of the emerging plethora of spatial data sources in order to improve the quality of research and more effectively address key research questions. It sets out to identify existing and potential applications of spatial data resources for geographers and non-geographers, and examines how the ESRC and its partners can make such data available and support user training needs.

The review is designed to inform the UK Strategy for Data Resources for Economic and Social Research (hereafter referred to as the National Data Strategy, NDS), which *'aims to ensure that the national data infrastructure meets the demands which will be placed upon it to address both current and future research needs'*. The Strategy seeks to ensure that the wide variety of different types of data now collected should take into account the potential to support future research. It is important to consider how spatial data can be used to inform and illuminate research questions.

As the following sections of this review illustrate, there is currently an ever-increasing range of geo-referenced data, but the ways in which it can be used to support social and economic research and the tools for handling and analysing merit further attention.

1.2 Aim

This project was commissioned by the ESRC's Strategic Advisor for Data Resources with the aim of:

> *Assisting the ESRC Research Resources Board (RRB) by addressing the question* ***of how the social science research community can take advantage of the emerging plethora of spatial data sources in the UK and internationally in order to improve the quality of research and more effectively address the key research questions.***

To support this aim, a review was commissioned which would:

- identify spatial data needs underlying current and likely future research questions for social scientists (and research partners in other disciplines);
- conduct a selective review of current facilities, infrastructure and capacity in place to support researchers who are using, or hoping to use, spatial data;

- identify the challenges faced by researchers in obtaining access to and using spatial data; and
- report findings and make recommendations to the RRB.

1.3 Structure of the report

Section 2 of the report briefly outlines the methodology adopted for this review. Section 3 outlines key features of the changing environment – in terms of trends in computing, developments in the private sector and the policy environment. An overview of the current state of availability of geospatial data is provided in Section 4, with particular emphasis on the rise of 'neogeography' and 'volunteered geographic information'. Section 5 describes the current state of knowledge about and use of geospatial data and highlights the need to share this knowledge. In Section 6 the important questions of confidentiality and security are addressed, while Section 7 is concerned with issues of data quality and fitness for purpose. Section 8 is concerned with skills and knowledge transfer. Recommendations are presented in Section 9.

2. Methodology

This review has been a challenging one given the difficulty of determining where to 'draw the boundaries' around geospatial data and associated research and skills issues, the range of academic disciplines involved, the speed of technological developments and changes in the policy environment. It has involved consultation with those who provide/operate spatial data research resources, as well as users and potential users of spatial data.

The stages of the research process were as follows:

- the identification of spatial data needs underlying current and likely future research questions for social scientists (and research partners in other disciplines) – via a study of recent literature produced by both academic specialists in GIS, geosciences and the GIS industry, supplemented by interviews with key organisations in academia and industry, in order to identify new developments and the issues involved in accessing and making use of the data;
- identification of the potential of spatial data for addressing existing and new research questions and the development of new analysis techniques spatial data needs underlying current and likely future research questions for social scientists (and research partners in other disciplines) through a review of the recent spatial analysis literature and

consultations with leading academic research centres working to develop new methods of spatial analysis;

- consultation (mainly via telephone interviews) with key data providers providing support for research (such as EDINA[1], ESDS[2] and CCSR[3]) and experts in using spatial data and developing training methods for users of spatial data, to derive information on current resources/facilities for access to spatial data, the skills users need to access and manipulate spatial data and current and emerging gaps in expertise;
- a short Web consultation (see Annex 1) of 512 researchers[4] to derive information on awareness and use of geospatial resources by researchers and their opinions about possible new areas for development via ESRC investment;[5]
- review and assessment of key findings from the various elements of the research; and
- formulation of recommendations concerning the steps that the ESRC and other relevant agencies could consider taking to facilitate and improve the nature of, and access to, spatial data resources.

3. The changing environment

Key messages

This section of the report provides a brief review of trends in computing hardware and software and the evolution of geographical information and the GIS industry.
There have been rapid developments in computing in recent years, which have led to enhancements in the ability to handle large amounts of geospatial data.
The private sector is playing an increasing role in the capture of geospatial data and in developing products and services based on such data.
There is widespread acceptance that the effective use of geospatial data requires the establishment of a geospatial data framework, which both catalogues the data sets available and provides a means of accessing data. The EU Inspire Directive and UK Location Strategy are important developments in this vein.
Academics will progressively have to adopt the standards applying to the public sector, and will have to update software, data management practices and training.

3.1 Developments in computing

Throughout the history of computing, the processing power of computers, the amount of RAM memory and disk storage, graphical display capabilities and the speed of networks have

---

1 EDINA is the Joint Information Systems Committee's national academic data centre based at the University of Edinburgh.
2 Economic and Social Data Service of the ESRC, University of Essex.
3 The centre for Census and Survey Research, University of Manchester.
4 The web consultation was managed by the UK Data Archive at the University of Essex. Members of mailing lists connected with the Economic and Social Data Service were invited to complete a short questionnaire (hosted by surveymonkey.com). The online consultation was 'live' from late December 2008 to February 2009.
5 Results from the Web consultation are presented in Sections 5 and 9.

continuously increased while the cost of processing power and data storage has decreased. This has led to an enormous increase in the ability to handle the vast data sets generated by geographic data capture (though the accelerating rate of increase in the volume of data captured still challenges improving hardware capabilities).

Geospatial data can be mapped and depicted visually in many ways, but can also be interrogated by applying database techniques to create new information. The capabilities of geographical software have evolved over time from simple cartographic data capture and display to integration of features from computer-aided design with database technology to create Geographical Information Systems (GIS) which can manipulate geographical data and the multidimensional attributes of geographical objects to create new information. While the development of GIS systems was initially driven by private sector organisations such as the Environmental Systems Research Institute (ESRI) which promoted proprietary systems such as ArcInfo/ArcGis,[6] the development of common standards (*e.g.* the OpenGIS Standards developed by the Open Geospatial Consortium) has accelerated the pace of development and adoption of interoperable solutions making geographical information more widely available and accessible to a wider range of applications. Open source GIS software (e.g. that created by the Open Source Geospatial Foundation[7]) has also recently expanded greatly, drawing upon these standards and creating new types of GIS applications, such as web-based mapping software and the tools to incorporate interactive maps on websites (*e.g.* the Mapguide project[8]). These software projects build upon general open source software tools such as Apache, PHP, SQL, PostgresSQL and Postgis, which are also being adopted by commercial GIS software.

Geographical Information Systems now incorporate analytical tools which make it possible to conduct analysis of geospatial data, such as estimating the population living within a given commuting time of a facility or conducting multivariate statistical analysis. There is increasing convergence with statistical software, with *S-Plus*, *SAS* and other packages adding GIS and geostatistical capabilities. The growth of open source software has had a particular impact on access to the capability to analyse geospatial data with public domain software such as *GeoDA*[9] and the open source *R* language offering sophisticated spatial analysis capabilities.

Commercial GIS software has traditionally been extremely expensive, placed considerable demands on the hardware on which it runs and until PC hardware achieved a sufficient degree

6 There was much development work on GIS systems in the 1979s and 1980s by the public sector in the USA, but most of these initiatives proved unsustainable.
7 http://www.osgeo.org/
8 http://mapguide.osgeo.org/
9 http://geodacenter.asu.edu/

of processing power (and a relatively stable operating system) it was mainly oriented towards mainframe systems or workstations running Unix. The development of genuinely capable desktop and web GIS software is still very recent, and still makes considerable demands upon the user, in terms of learning the software, GIS techniques and data manipulation methods, as well as the appropriate use of cartographic tools. Thus, until the recent explosion of web-based mapping applications, GIS has been largely the domain of specialist staff.

3.2 Developments in the private sector

Though much innovative work was undertaken in the US public sector (*e.g.* US Geological Survey and Bureau of Land Management) in the 1970s and 1980s, the private sector has dominated the GIS industry. The demands of facilities management, surveying and geodemographics/marketing in local/regional government and private industry have largely driven the evolution of the industry, with academia lagging behind but developing analytical techniques with bespoke systems which have gradually been incorporated into industry standard software. There have been two broad strands of development which have gradually converged over time – raster-based systems derived from remote sensing (*e.g.* Earth Resources Data Analysis System (ERDAS)) and vector-based GIS systems derived from the integration of CAD and database systems (AutoCad and ArcInfo). GIS systems have traditionally been mainframe or workstation-based but have moved towards desktop and web-based systems.

The amount of geospatial data has multiplied enormously during the life of the industry aided by increasing speed of data capture. Initially, map data had to be converted to digital form by manual tracing on a digitiser table, but this has been superceded by remote sensing from satellites and aircraft (aerial survey), scanning of paper maps and recording of co-ordinates by GPS-enabled surveying equipment. Now vehicles (*e.g.* distribution lorries, public service vehicles, emergency services), mobile phones and other devices incorporate GPS receivers and can record co-ordinates and mobile phone calls can often be traced to a location by triangulation. Other business systems automatically generate locational attribute data, from the postal code of customers or recipients of public services.

Other companies have developed products which derive new information from geospatial data. For example, the geodemographics industry has developed classifications of the population for small areas derived from combining Census data with customer information. These have proved very powerful for targeted advertising of products and services. Other products have been developed around processing address lists and postcodes, or developing route planning software from digital road network data. These products make use of GIS techniques and GIS software to display their results – *e.g.* to draw a route on screen or produce a summary of the characteristics of the population within a given distance of a particular facility (*e.g.* a hospital or supermarket). These companies are also increasingly making use of web-based systems to present their results or services.

GIS systems have been important for facilities management in local government for many years and local government research units were some of the earliest users of computer cartography and spatial analysis tools (*e.g.* in planning). However, local and central government have also recently started using web mapping in a major way to communicate with citizens during the last 5 years (facilitated by the free availability of 2001 Census data and digital boundaries of Census areas) and local and regional information systems and 'observatories' have proliferated, using software such as InstantAtlas to provide interactive map-based profiles of localities and regions, which allow users to undertake limited interrogation of local data sources through a map interface. This has been paralleled in the private sector with consultancy companies developing web-based mapping systems to depict local characteristics (*e.g.* Local Futures). Geography is an increasingly important component of governmental data bases, with the unit postcode and street address of users of public services, benefit claimants and tax payers recorded.

The private sector is playing an increasing role in the capture of geospatial data. An important source of this is remote sensing data, mainly generated by aerial surveys (*e.g.* getmapping.com), which has increasingly moved towards the collection of 3-d data through LIDAR (radar) surveys (*e.g.* the Geoinformation Group 'Cities Revealed' product) and oblique photography (*e.g.* BLOM pictometry). There is a growing trend towards the photorealistic display of 3-d objects which enable an urban environment to be explored using 'fly-throughs' using computer gaming technology. A significant problem with remote sensing data is that much is out of date, but the frequency of data capture and updating of aerial photography imaging is increasing. The Geoinformation Group has just launched UkMap, a competitive product to Ordnance Survey Mastermap generated through aerial and GPS survey and updated every 3 years.

Private companies such as Navteq and Teleatlas have become dominant in the supply of digital road network data used in satellite navigation systems and others are now supplying 'point of interest' (POI) data for integration with these systems derived from commercial data. An example of this would be the location of petrol stations derived from the records of petrol delivery companies, which can be mapped on satnav or used by 'location based services' to identify the nearest petrol stations along a planned route.

3.3 The changing policy environment

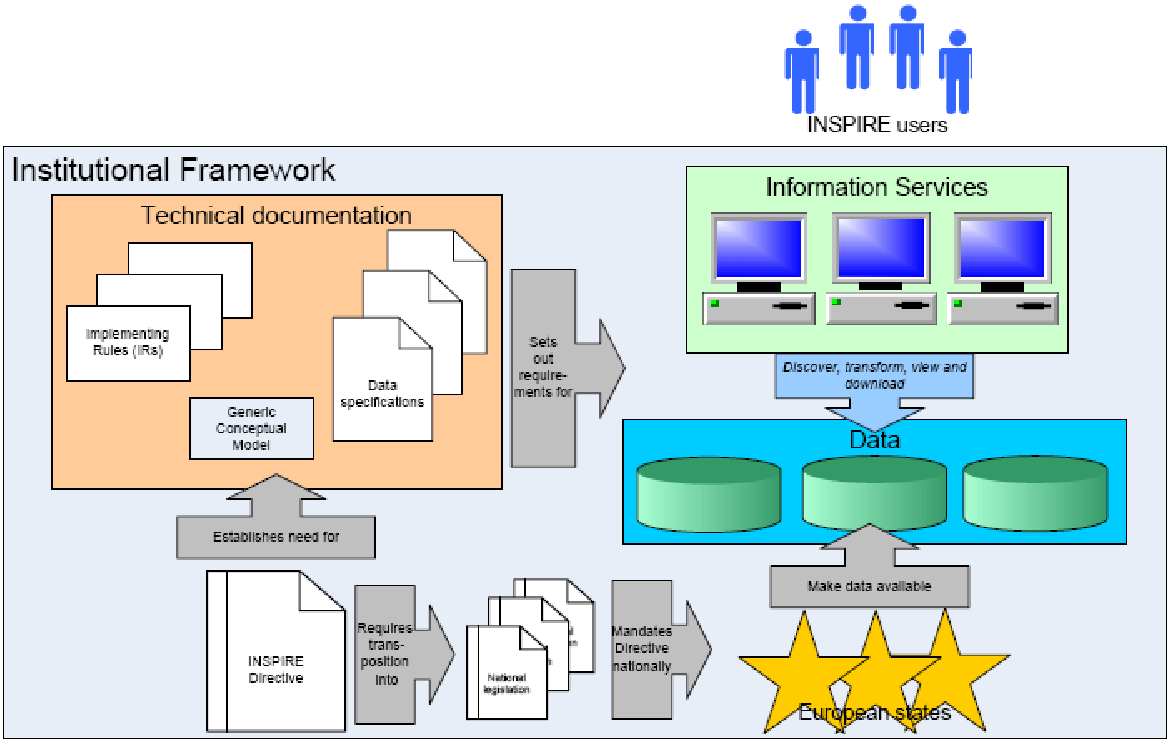
There is widespread acceptance that the effective use of geospatial data requires the establishment of a geospatial data framework, which both catalogues the data sets available and provides a means of accessing data. There have been a number of attempts in the UK to establish a geospatial data framework, the most ambitious of which was the National Geospatial Data Framework created in the late 1990s. This failed to achieve widespread acceptance, but its main achievement was the creation of the GIGateway,[10] hosted by the Association for Geographical Information and EDINA, which attempted to catalogue the available geospatial data resources in the UK and provided metadata describing data sets and access procedures. Independently, some parts of the private sector initiated another attempt to codify access to geospatial data, in the form of the Digital National Framework[11]. The rapid pace of development of data resources, the important role of commercial interests and the 'commodification' of public sector geospatial resources all pose considerable challenges to maintaining and developing such initiatives.

The most important recent development in terms of access to geospatial data has been the creation of ***the EU INSPIRE Directive***, which came into force on 15th May 2007.[12] The initial motivation for this was to facilitate the creation of a European spatial data infrastructure (SDI) which would be better able to formulate, implement and monitor environmental policies. It will deliver integrated spatial information services which will allow users to identify and access spatial or geographic information from a wide range of sources from local to European-wide. All public bodies will be required to create metadata and to provide access to data that they have collected or created, in standard form through the network services, but does not require the creation of new datasets. The Directive provides five sets of Implementing Rules which set out how the various elements of the system (metadata, data sharing, data specification, network services, monitoring and reporting) will operate. The main components of INSPIRE are shown in Figure 3.1. Each Member State is expected to create their own spatial data infrastructures to promote data sharing and interoperability within their own borders.

---

10 http://www.gigateway.org.uk/
11 See: www.dnf.org
12 The full text of the Directive (2007/2/EC) can be found at: http://www.ecgis.org/inspire/directive/l_10820070425en00010014.pdf

Figure 3.1: The main components of INSPIRE (taken from Digital National Framework (DNF), 2008)



Source: Digital National Framework, 2008.[13]

The key principles of INSPIRE are that:

- data should be collected once and maintained at the level where this can be done most effectively;
- it must be possible to combine seamlessly spatial data from different sources across the EU and share it between many users and applications;
- it must be possible for spatial data collected at one level of government to be shared between all levels of government;

The Directive is to be transposed into national legislation in each EU country by the end of May 2009. A technical infrastructure will be established at European level for Data and Information services. The UK will have to build a spatial data infrastructure which can at least make available spatial datasets for each of the themes mandated in the Directive and these datasets and their metadata will then have to be maintained. The themes will have to be made available either directly to a European INSPIRE portal or via a UK portal supporting discovery, transformation, viewing and download.

The ***UK Location Strategy*** was published by the Department for Communities and Local Government on 25th November 2008, partly prompted by the EU INSPIRE Directive, which

13 Digital National Framework (2008) Implications of the INSPIRE Directive, DNF White Paper, DNF0051.

is concerned with establishing an infrastructure for spatial information (and which will have to be transposed into UK law by May 2009). This aims to *'maximise the value to the public, government, UK business and industry of geographic information. It will provide a consistent framework to assist national, regional and local initiatives and service delivery'*. It argued the need for a coherent strategy because current users of geographical information spend 80 per cent of their time collating and managing it and only 20 per cent of their time analysing it, and commented that there is 'currently little understanding of the data collected, stored or maintained by either the public or private sector' (p.13). The strategy proposes that a UK Location Council establish and maintain an inventory of public sector geospatial datasets, common geographical reference data and establish an infrastructure for sharing location-related information. It thus aims to ensure that information about UK's land, sea and air is:

- fit for purpose;
- collected once to universally accepted standards;
- appropriately maintained and used many times by the public and private sector.

Data should be easy to discover (and with clear terms for its use), simple to access and easy to share and integrate and understood sufficiently to maximise its application. Public sector bodies should identify and document their geospatial data holdings to a common standard and a small number of Core Reference Geographies for which specific public sector bodies should be responsible should be established. Dataset owners should simplify licensing in order to facilitate data sharing, and the Location Council should lead the improvement in skills necessary to make better use of geospatial information. Implementation of the INSPIRE Directive is the responsibility of DEFRA. A consultation document was produced in March 2009, with a deadline of the end of May for comments. Therefore the timetable for transposition of the Directive into UK law has already slipped.

A geospatial data infrastructure which is fit for purpose has the potential to improve the efficiency and effectiveness of services provided by government, while improving the quality of analysis by academics and policy makers. However, it is not just a one-off investment – substantial ongoing investment will be necessary in making effective procedures in using geospatial data, updating it and making it available for re-use and in training to handle, analyse and interpret the data.

3.4 Implications for academia and the ESRC

INSPIRE and the UK Location Strategy are primarily directed towards the public sector, and it is not clear how far major data and software providers in the commercial world will be directly included. However, there will be indirect effects upon the commercial sector because their customers in central and local government will demand that the software supplied will work to the standards required by the Strategy and will also require that data is supplied in compliant formats and metadata and other documentation adopts the relevant standards.

The academic sector is primarily a user of geospatial data supplied by UK and international government and official statistics bodies, with a more limited role for private sector data providers.[14] As a user, it will be mainly affected by the changing formats of geospatial information and new standardised metadata standards. There will be a need to update software to use these new data sets and associated training needs for using new software.

However, much academic analysis of geospatial data involves the manipulation of data and creation of new datasets. Where these are to be shared or archived for use by researchers from outside the academic sector (*e.g.* via the deposit of derived data set with the Economic and Social Data Service as a requirement of ESRC-funded research projects) there will inevitably be an expectation that these data sets are created to the standards laid down by the UK Spatial Data Infrastructure (SDI) and that the necessary metadata documenting the data set is created. This will generate a need for training and support. Students trained in GIS techniques will also have to be proficient in working to the standards laid down by the UK SDI. Hence it is likely that even if the academic sector is not explicitly included, in reality academics will progressively have to adopt the standards applying to the public sector, and will have to update software, data management practices and relevant training accordingly.

4. Availability of geospatial data

Key messages

- There has been a growth in the availability of geospatial data, but at the same time there has been a growth in demand for such data.
- The rise of neogeography and volunteered geographical information has extended the use of geospatial data to new users, some of whom had no previous interest in or expertise in handling and interpreting geographical data.
- There is a need for comprehensive and high quality metadata in order to enable users to identify, and make appropriate use of, different data sources.

14 This is the case for researchers, teachers and students, though the administrative and support services do make use of commercial services such as profiling of postcoded information and surveying services.

4.1 Introduction

A fairly comprehensive definition of the term 'geospatial data' is provided by the US Environmental Protection Agency:

> geospatial data identifies, depicts or describes geographic locations, boundaries or characteristics of Earth's inhabitants or natural or human-constructed features. Geospatial data include geographic coordinates (e.g., latitude and longitude) that identify a specific location on the Earth; and data that are linked to geographic locations or have a geospatial component (e.g. socio-economic data, land use records and analyses, land surveys, homeland security information, environmental analyses). Geospatial data may be obtained using a variety of approaches and technologies, including things such as surveys, satellite remote sensing, Global Position System (GPS) hand-held devices, and airborne imagery and detection devices.[15]

A large proportion of data generated by the government and commercial sectors can be classified as being geospatial because it includes some kind of explicit or implicit geographical reference. The European Joint Research Centre (JRC) has estimated that 90 per cent of European environmental data is spatial. At one extreme lies data which records the properties of spatial entities – *e.g.* Ordnance Survey maps, geological or hydrological survey maps or survey data recording roads and street furniture. At the other extreme is statistical data which contains a code that links it to a geographical location or geographical entity (*e.g.* an administrative area or a point in space identified by a national co-ordinate system and/or latitude and longitude).

The benefits of geospatial referencing have not been exploited to date nearly as much as they might have been to address academic and policy-related questions. Yet it is clear that utilisation of geospatial data can provide additional information and insights into key questions, offer possibilities for visualisation and for integrating different data sources and enable the inclusion of *context* in analyses via concepts of proximity, containment, overlap, adjacency and connectedness.[16]

4.2 Increase in availability

Technological advances and commercial developments are resulting in a wide range of new types of geospatial data becoming available. Examples are:

- Credit/debit/store card transactions recording location of purchase and home location
- Mobile phones recording location of phone call
- Increasingly detailed satellite imagery
- Aerial survey data – in 2 and 3 dimensions

15 http://www.epa.gov/Networkg/glossary.html#G

16 Lakes T. (2008) Improvements and Future Challenges for the Research Infrastructure in the field of Geodata, (RatSWD Working Paper No. 89), Geomatics Department, Humboldt University of Berlin.

- Pictometry – aerial survey data photographed from an oblique angle, which incorporates physical measurements of 3-dimensional objects
- Tracking data of vehicles fitted with GPS
- Photographic/video data tagged to a geographical location (including Google Streetview images and geographically coded digital photographs from phones/cameras fitted with GPs and uploaded to Google Maps/Earth)
- An increasing amount of statistical data generated via administrative systems is being made available containing detailed spatial references.

A key challenge is to make this vast range of data available for researchers. Recent years have seen the growth if Internet-based access points for geospatial data (i.e. geoportals) and other Internet-based information systems. In the UK the amount of micro-area level socio-economic data available has increased markedly with the development of Neighbourhood Statistics[17] (for a wide range of demographic and socio-economic statistics) and the DWP Tabulation Tool[18] for access to selected DWP data sets on claimants, etc. Increasingly, Web-based systems enable direct access to data in a variety of formats (*e.g.* via Excel spreadsheets, in map form or for input into other software systems).

In addition, new sources of 'traditional' digital map data are becoming available – *e.g.* the *TeleAtlas* and *Navmap* topographic and road network data incorporated in Google Earth/Maps and Microsoft Virtual Earth and made available 'free' to users of these services. The increasing ubiquity of Google Earth/Maps is beginning to threaten the commercial models underpinning national mapping agencies such as Ordnance Survey, who are trying to sell data superficially similar to that made freely available by Google and Microsoft.

4.3 Neogeography and volunteered geographic information

The emergence of 'Web 2.0'[19] has enabled the phenomenon of '***neogeography***' (or 'geography for the general public' – although the definition is the subject of debate[20]) to become established.[21] This term is used to represent the application of a diverse set of practices which are 'geographical' in character in order to present information in a visual manner or which manipulate spatial objects for a social or cultural purpose, but which do not

17 http://www.neighbourhood.statistics.gov.uk/dissemination/
18 http://www.dwp.gov.uk/asd/tabtool.asp
19 The term 'Web 2.0' summarises the emergence of services from web-based communities focusing on technologies of social networking, social bookmarking, blogging, Wiki's and RSS/XML feeds. This includes sites such as YouTube, Facebook, Blogger, Flickr etc., common to all of which is ease of use and the ability to communicate via freely available tools. They can be learnt quickly and effectively without the need for formal training and professional qualifications.
20 For a history and overview of developments in neogeography see: http://www.d-log.info/on-neogeography.pdf
21 http://www.casa.ucl.ac.uk/working_papers/paper142.pdf

conform to the standards of professional geography or cartography, and which while presenting detailed information in an innovative way rarely make use of spatial analysis. There are a growing number of Web 2.0 applications which provide access to geospatial data and (fairly simple) transformation/analysis capabilities via Google application programming interfaces. Individuals can now plot their data on a Google (or MS Virtual Earth) backdrop via keyhole markup language files exported from GIS systems, while a large number of 'mashup' websites (websites that weave together data from different sources to create a new integrated user service) have been developed which enable services and information such as pictures to be displayed and searched via a map interface through linking web directory services to Google Maps or Virtual Earth.[22]. These developments have been seen as bringing about a 'geospatial web',[23] in which the merger of geographical with abstract information creates an environment which enabled search for things based on location instead of by keyword only.

Some academics have taken considerable interest in neogeography, as the extract in Box 4.1 from the extended abstract of a call for papers for a session at the 2009 Annual Conference of the Association of American Geographers (AAG) shows.

Box 4.1: Extract from a call for papers on neogeography at the 2009 AAG conference

Concepts, Tools and Applications: The Rise of Neogeography
The world of Geographic Information (GI) Science has changed. It has experienced expeditious growth over the last few years leading to fundamental changes to the field. Web 2.0, specifically The Cloud, GeoWeb and Crowd Sourcing are revolutionising the way in which we gather, present, share and analyse geographic data. This renaissance in the importance of geography in the Web 2.0 world is becoming known as 'Neogeography'.

Neogeography is geography for the general public using Web 2.0 techniques to create and overlay their own locational and related information on and into systems that mirror the real world. Location and space now represents a key part of the Web 2.0 revolution. Tagging not only the type of information but where such information is produced, who uses it and at what time, is fast becoming the killer application that roots information about interactivity generated across the web to systems that users can easily access and use in their own communication with others.

The aim of this session is twofold; first to bring together practitioners to discuss concepts and challenges that the field of Neogeography faces. Secondly, to provide an opportunity for researchers and developers to present recent tools and applications for collecting, sharing and communicating spatial data for the Neogeographer. We are actively seeking topics ranging across the entire spectrum of Neogeography, from Crowdsourcing, Digital Earths, Neogeography, Web Mashups, Volunteered Geographic Information, Virtual Worlds (*e.g.* Second Life) and associated Web 2.0 technologies.

This changed environment has facilitated the growth of ***volunteered geographical information (VGI)*** or 'crowd sourcing' – which has been described as the harnessing of tools

22 An example is www.platial.com
23 www.geowebguru.com

to create, assemble, and disseminate geographic data provided voluntarily by individuals.[24] It harnesses the power of having geoinformation authors across the globe, with the Internet aiding and enabling public sharing of this information and provides a way of collecting geographical information which challenge the products of national mapping agencies and commercial GIS companies. The most remarkable example of this is Openstreetmap[25], an on-line digital map with street networks contributed by volunteers. This is complemented by Openrouteservice[26] which provides a route planner and 'drivetime' calculator based on Openstreetmap data for countries in Europe and which gathers GPS 'tracklogs' contributed by volunteers. Another example is the 'geotagging' of photographs using a GPS-enabled mobile phone (or by recoding the location of a photograph using a GPS receiver), which can be used to generate digital mapping data. The flickr service has experimented with this[27], and is inviting volunteers to improve the quality of the digital data created by contributing further images or correcting the geotagging where the modelled data exhibits problems.

The ***significance of neogeography and VGI*** is that it is both a way of enthusing students about geospatial data and it also brings individuals with no previous background or expertise in geospatial data handling into use of geospatial data. There are also important ***data quality issues*** raised by VGI. Whereas traditional mapping agencies have elaborate standards and specifications to govern the production of geographic information and their products have a reputation for quality, VGI can make no such claims – indeed, it may be thought of as '*asserted* geographic information',[28] in that its content is asserted by its creator without citation, reference, or other authority. Nevertheless, users often take a rather cavalier approach to data quality, taking little or no account of quality limitations – perhaps on the basis that meta data sheets that computerisation carries authority per se. This approach to data quality may then be transferred to other information sources.

Clearly, there is scope to make use of VGI. For example, such information may be particularly suited to exploring the ***geographies of everyday life*** (i.e. understanding how individuals move around between different activities on a day-to-day basis). On a more systematic and technical basis, in the USA design and implementation is underway of a prototypical *Geospatial Exploratory Data Mining Web Agent (GEDWMA)*, which reads webpage data and follows links to acquire knowledge in order to add geoinformation usable in a GIS. The agent creates a database of webpage text, mines it for location information, and

---

24 See Goodchild M.F. (2007) 'Citizens as sensors: the world of volunteered geography', Geojournal 69, 211-21.
25 www.openstreetmap.org
26 http://www.openrouteservice.org/
27 http://code.flickr.com/blog/2008/10/30/the-shape-of-alpha/
28 Goodchild M.F. (2007) op cit., 220.

then converts it to geospatial data format. The data may be visualised and analysed after it is converted into GIS and virtual globe formats. In so doing this provides diverse user communities with a tool that utilises a variety of distributed data sources to discover additional knowledge about their fields of interest.

4.4 Meta data

Meta data is ***'data about data'***. Such information about when and where data are collected, the data collection methodology used (including sampling frame, sample size, *etc.*, as appropriate), classification schema used, disaggregations available, conditions of use, is essential for documentation purposes and to aid users in assessing the suitability of data sets for various applications. As noted in section 3, the changing policy environment places greater emphasis on good quality meta data. Likewise interviewees emphasised the continuing need for high quality meta data. The metadata sheets that come with ARC GIS were noted as being helpful in improving 'good practice'.

4.5 Curation and preservation

Both public and private sector (and to a lesser extent the academic sector) tend to have a lower interest in historical data. Curation and preservation is an important issue and one that the academic sector is well placed to deal with. Several interviewees noted their interests in the use of historical spatial data and related metadata for research purposes.

5. Knowledge about and use of geospatial data

Key messages

- Interest in, and use of, geospatial data is spread across a range of social science disciplines.
- Some suppliers of geospatial data believe that detailed knowledge of some key spatially-referenced socio-economic data sources is diminishing over time.
- There is a need for ongoing reinvestment in knowledge and skills if appropriate use of geospatial data is to be maximised.
- There is a broad distinction, and arguably an increasing divide, between those researchers with considerable expertise in terms of knowledge and analytical techniques for using geospatial data on the one hand, and on the other those with a much more limited knowledge and expertise in using geospatial data and applications.
- There is a clear need for a range of geo-spatial data services expressed by respondents to the online enquiry, particularly for advice and guidance on the use of, access to and linking with geospatial data.

5.1 The state of current knowledge and use of geospatial data

It is difficult to make an authoritative assessment of the state of current knowledge and use of geospatial data. Some researchers may not be aware of geospatial data that could help them answer their research questions, while others might have an awareness of the availability of such data but lack the detailed knowledge of data sources, or proficiency in the analytical techniques necessary, in order to analyse them. Others may have knowledge of data sources and some competence in the application of analytical techniques, but might not interpret the outputs correctly or overlook important data quality issues.

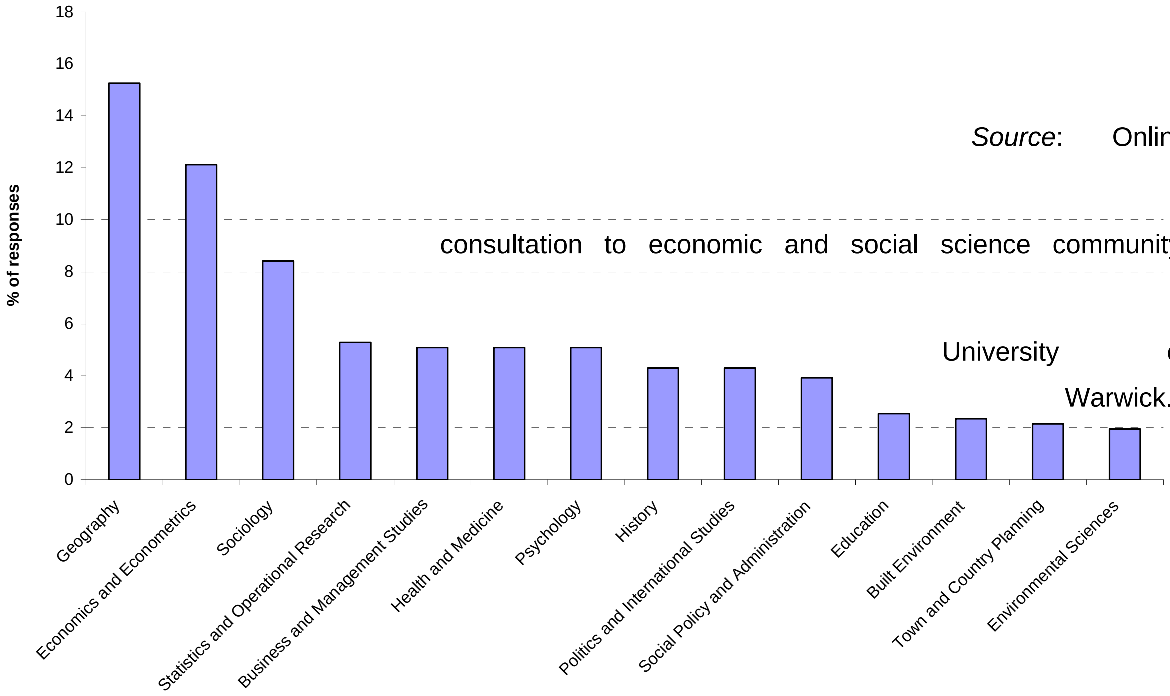
In order to gain some insights into the state of current knowledge and use of geospatial data interviews were conducted with some of the leading academics involved in the application of geospatial data sources and with individuals concerned with facilitating access to such sources. This was complemented by the online consultation outlined in section 2, which had 512 respondents. While together the information provided from these sources is not comprehensive, it does provide some useful insights into the state of current knowledge and use of geospatial data.

At face value, it might be expected that geographers would be the main users of geospatial data. However, from the interviews conducted with leading academics and data providers it was clear that ***knowledge, use and interest in geospatial data sources extends beyond geography to a range of other social science (and other disciplines)***.[29] Indeed, some respondents considered that the ***main increase in interest in and use of geospatial data amongst academics is from users in disciplines outside geography*** – reflecting 'spatial turns' in disciplines such as sociology and an upsurge of interest in geographical economics. One interviewee summed up this move as: 'the retreat of geographers from spatial data and the increasing interest of other disciplines in spatial data'. The disciplinary profile of respondents to the online consultation conducted for this project underlines an interest in geospatial data from a range of disciplines (see Figure 5.1). The largest single group of respondents reported that their academic discipline was Geography (15%), but Economics (12%) and Sociology (8%) were also strongly represented.

Academic users interested in the application of ***multi-level modelling techniques***, who wish to take account of the influence of geographical context (alongside individual and household level variations), have a particular interest in using geospatial data. Likewise there is an interest in ***adding spatial data or using geospatial data to link in other data to existing***

29 Which is strikingly similar to the reported profile of users of EDINA Digimap.

Figure 5.1: Respondents to online consultation on knowledge and use of geospatial data by discipline



Source: Online consultation to economic and social science community, University of Warwick.

survey data **–** often enabling analysis across policy or subject domains (*e.g.* the labour market and health). However, users might not know how to go about such linkage.

Another key growth area is amongst those ***outside academia in local authority settings (and other public agency settings)*** where there is a growing emphasis on linking and mapping geospatial data sources from different domains and making comparisons across areas. Such analyses may be used to inform and target policy interventions (particularly in the context of neighbourhood renewal and the development of Local Area Agreements (LAAs) – as outlined in the Supporting Evidence for Local Delivery (SELD) programme.[30]

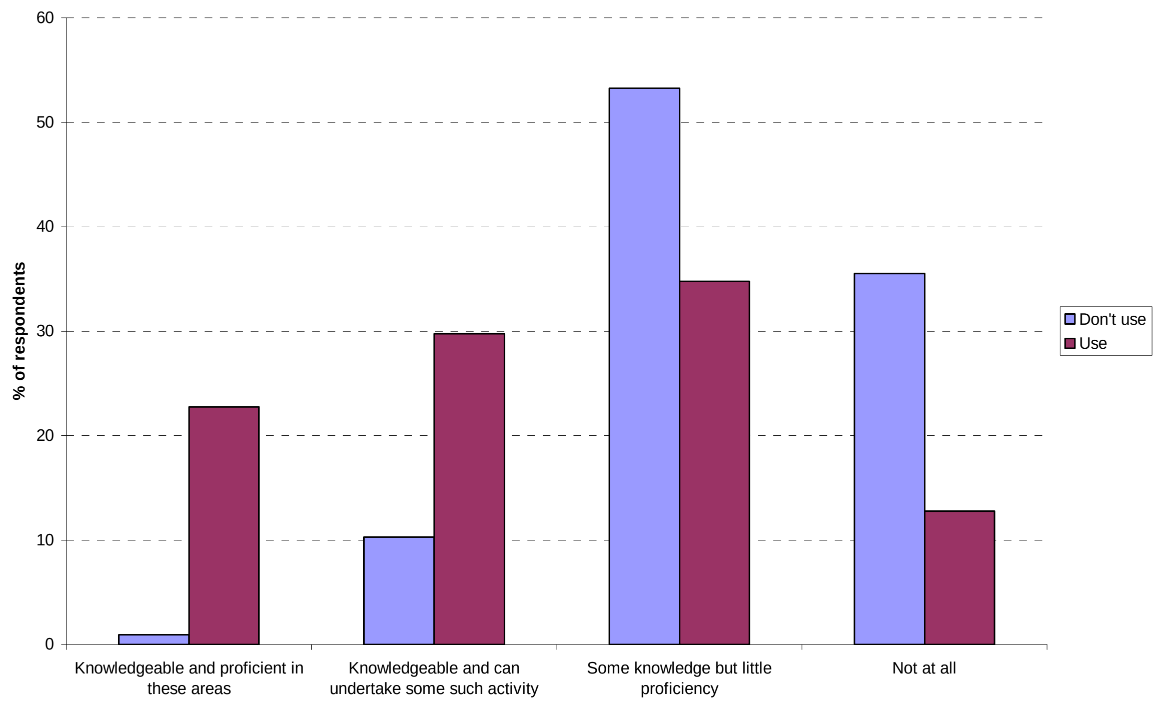
Some interviewees acknowledged that a great deal of specialist knowledge on key data sources had previously existed in the public sector. Some information providers interviewed felt that, in general, the 'knowledge base' of users of geospatial information is decreasing over time, citing more users (often with degree level qualifications) asking relatively simple questions about relatively simple statistical concepts and also there being ***less awareness of the strengths and weaknesses of key data sources and an understanding of appropriate uses***. It was considered that a great deal of specialist knowledge on key data sources had

30 Johnstone D., Johnstone S., Bell F., Garrard S., Knight B. and Peck F. (2008) Supporting Evidence for Local Delivery: National Research and Evaluation – Key Findings, Communities and Local Government, London. http://www.communities.gov.uk/publications/communities/localdelivery

previously existed in the public sector. The demise of Research & Intelligence (R&I) units in local authorities and analogous organisations was considered to have contributed to this, as had the practice of more junior staff 'doing data analysis' for one day per week alongside other roles, following examples of what had been done in the past, rather than devoting any time to exploring the potential of new analytical techniques and data sets. Despite the rise of Regional and Local Observatories, the current economic context of public sector cost savings may diminish the existing knowledge base still further.

In a fast changing environment it is difficult to keep abreast of developments in geospatial data; hence those individuals who are ***knowledgable and proficient*** users of geospatial data at one point in time need not necessarily be so at a later date, unless they have the opportunity and inclination to reinvest in their knowledge and skills. Of respondents to the online consultation who used geospatial data, less than a quarter claimed to be 'knowledgeable and proficient', the most common category being 'some knowledge but little proficiency'. About a fifth of respondents (107) made no use of geospatial data, and the great majority of these had little or no knowledge of and proficiency with geospatial data analysis and mapping techniques (Figure 5.2).

Figure 5.2: Use of geospatial data and self-assessed proficiency

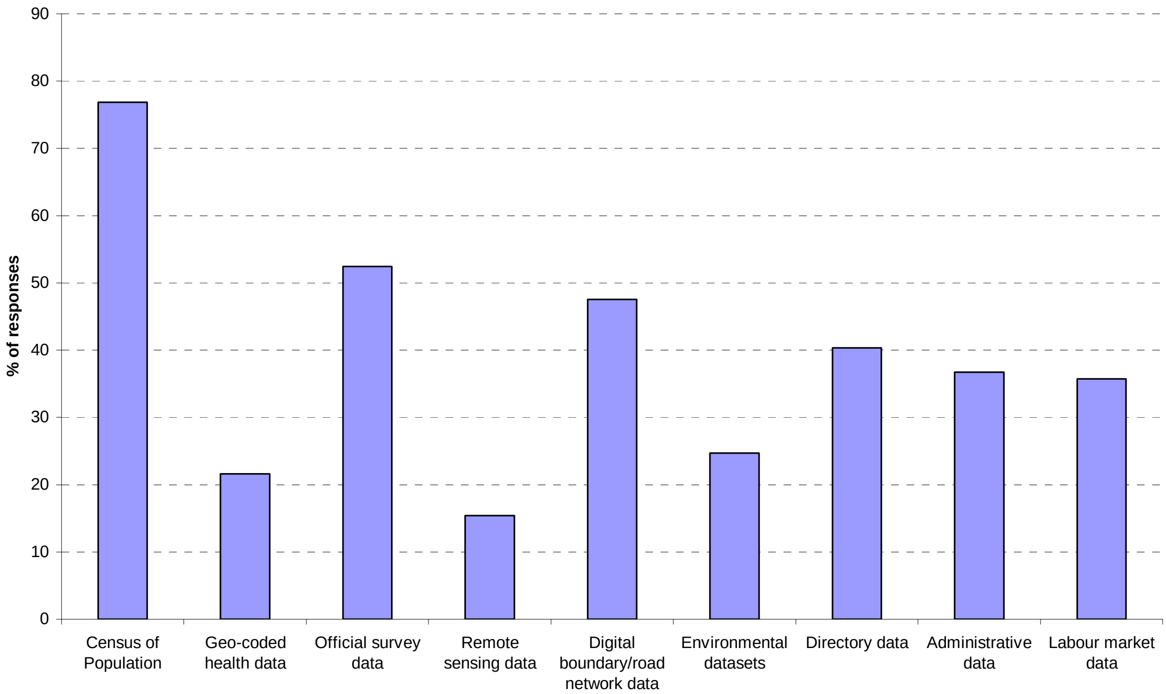


Source: Online consultation to economic and social science community, University of Warwick.

It is clear from the self-assessment of knowledge and proficiency of respondents that ***a lack of such knowledge and proficiency is one barrier to use of geospatial data***. Just under a quarter of survey respondents can be classified as 'expert' in terms of being 'knowledgeable and proficient' in the use of geospatial information.

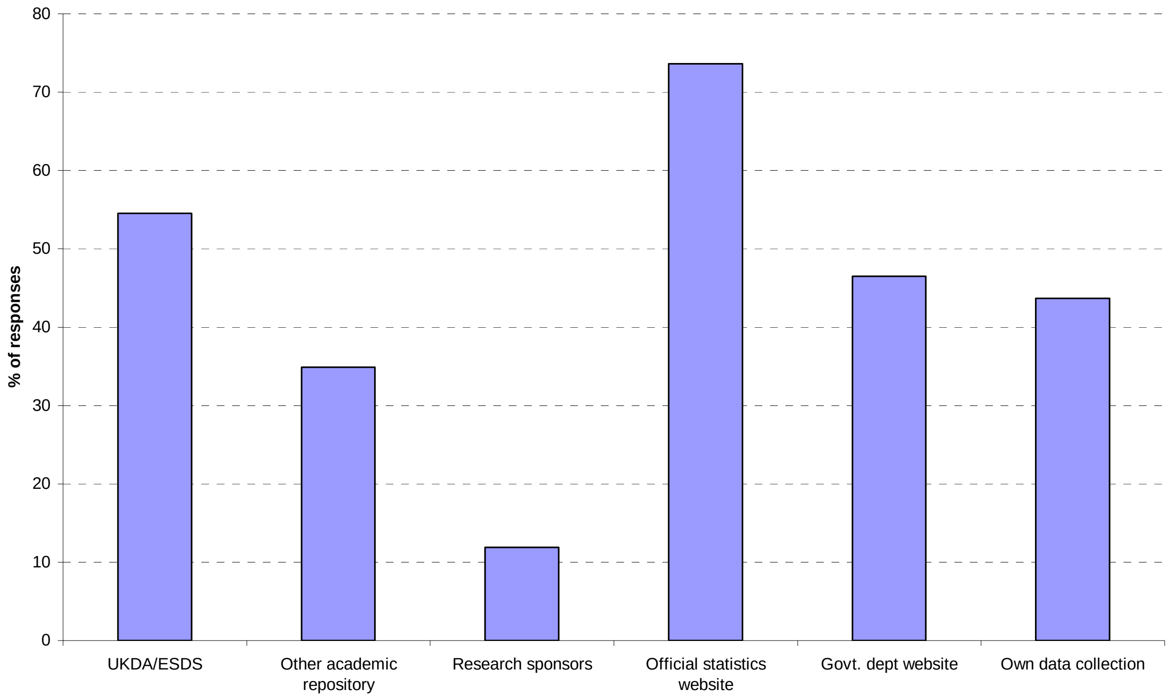
The main data source used by survey respondents was the ***Census of Population*** (mentioned by over three-quarters of respondents), while around half used official survey data and digital boundary data (Figure 5.3). It seems likely that the release of 2011 Census data in due course will lead to an upsurge in interest in using geospatial data and that there will be continuing interest in the use of official survey data. The most frequently used source used for accessing data was the ***Office for National Statistics (ONS) website***, with just over half of respondents using the UK Data Archive / ESDS (Figure 5.4).

Figure 5.3: Geospatial data sources used by respondents



Source: Online consultation to economic and social science community, University of Warwick.

Figure 5.4: Source from which data sources are accessed



Source: Online consultation to economic and social science community, University of Warwick.

5.2 The need to share knowledge

On the basis of a synthesis of information gleaned from interviews and from the online consultation conducted for this project, in simple and very broad terms there would appear to be a distinction between: (1) those researchers with considerable expertise in terms of knowledge and analytical techniques for using geospatial data; and (2) those with a much more limited knowledge and expertise in using geospatial data and applications. It seems that the first group may talk in a language that the second group often find difficult to comprehend. The use of visualisation techniques (which can bring geospatial data to those individuals who do not necessarily have high level skills in terms of handling geospatial data) and of neogeography and VGI (as highlighted in section 4) might help to promote an appreciation of the use and potential of geospatial data, but in a way that seems 'out of reach' (in terms of access to computing power, specialist knowledge of data sources and analytical and statistical techniques that seems out of reach to those in the second group. While the first group need support in terms of access to cyberinfrastructure,[31] access to data (including non-UK data),[32] lobbying about the balance between data suppression and maintaining the usefulness of data while preserving of confidentiality, and negotiation of licensing

31 This is where a lot of the emphasis in the USA is being put.
32 Difficulties in accessing European data were highlighted by several interviewees.

agreements, *etc*., in some ways they are better able to support themselves than the second group, for whom greater assistance is needed in terms of:

a) how to access geospatial data;
b) meta data about different data sources;
c) how to append geospatial codes to data and use geospatial identifiers to merge different sources; and
d) the identification of statistical techniques appropriate for analysing geospatial data and interpreting results.

One possible approach to this 'geospatial divide' would be for expert users to be encouraged by various mechanisms to provide relevant training and guidance for the non-expert users and potential users (i.e. those who are not both knowledgeable and proficient in the use of geospatial information). However, expert users are not necessarily in the best position to undertake cutting edge research using geospatial resources at the same time as providing training and advisory services on use of such data.

Evidence of the need for such services was apparent in the responses to the online enquiry. Respondents were asked which of four possible data services they would find useful:

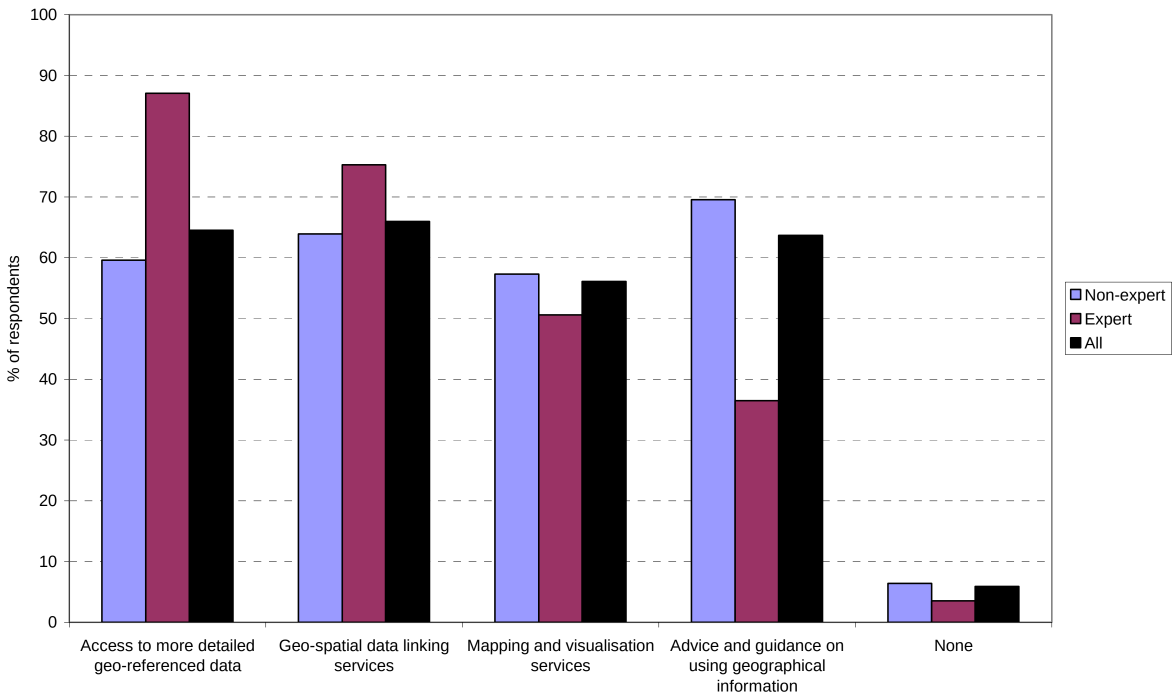
- access to more detailed geo-referenced data
- geo-spatial data linking services
- mapping and visualisation services
- advice and guidance on using geographical information

Figure 5.5 shows the distribution of responses to this question. All four services were regarded as useful, with approximately two thirds of all respondents selecting 'access to more detailed geo-referenced data', 'geospatial data linking services' and 'advice and guidance on using geographical information', and just under three-fifths identifying mapping and visualisation services as useful. A distinction is also made between 'expert' users (those who reported that they were knowledgeable about geospatial information and could undertake spatial analysis) and the remainder (referred to here as 'non-expert'). A larger proportion of expert than of non-expert users identified 'access to more detailed geo-referenced data' and 'geo-spatial linking services' as useful. Conversely, the proportion of non-expert users identifying 'advice and guidance on using geographical information' and 'mapping and visualisation services' was greater than for expert users.

Figure 5.6 shows which of the services specified respondents regarded as most useful. Experts identified 'access to more detailed geo-referenced data' as being most important, with

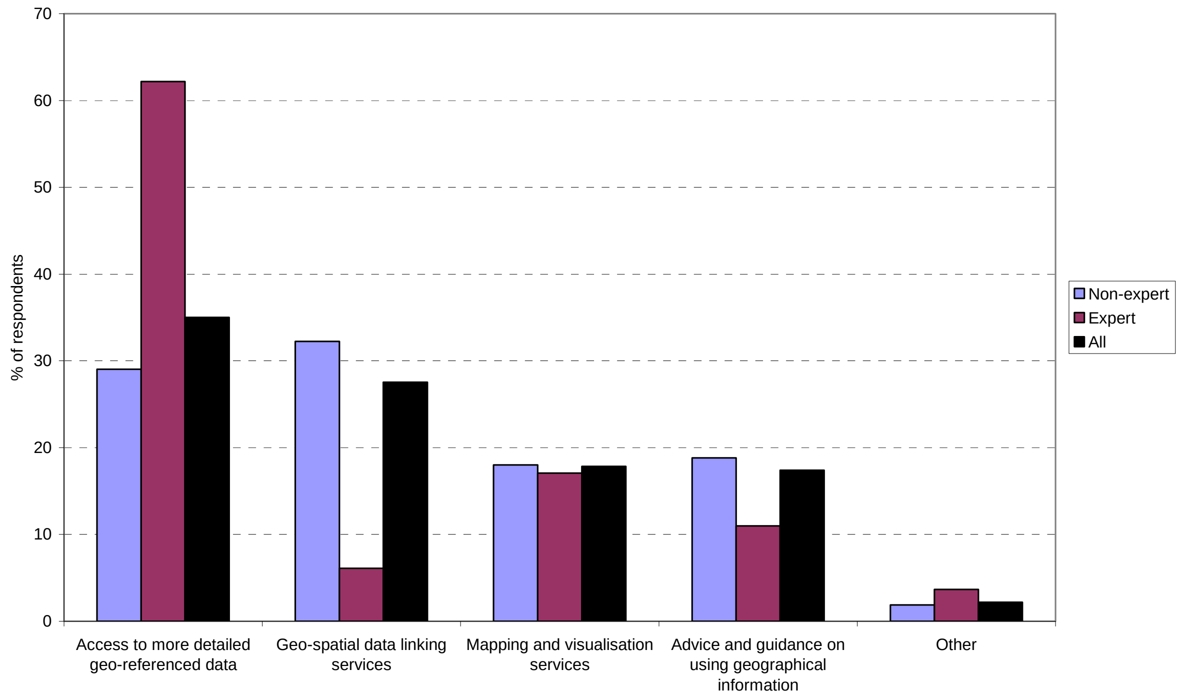
62% of respondents in this category identifying this option. Access to such information was identified as most important by 29% of non-expert users. For non-expert users ‘geo-spatial data linking services’ emerged as the single most important service, identified by 32% of such users. However, it is clear that non-expert users have a somewhat wider spread of needs than non-expert users.

Figure 5.5 Which of the following services would you find useful?



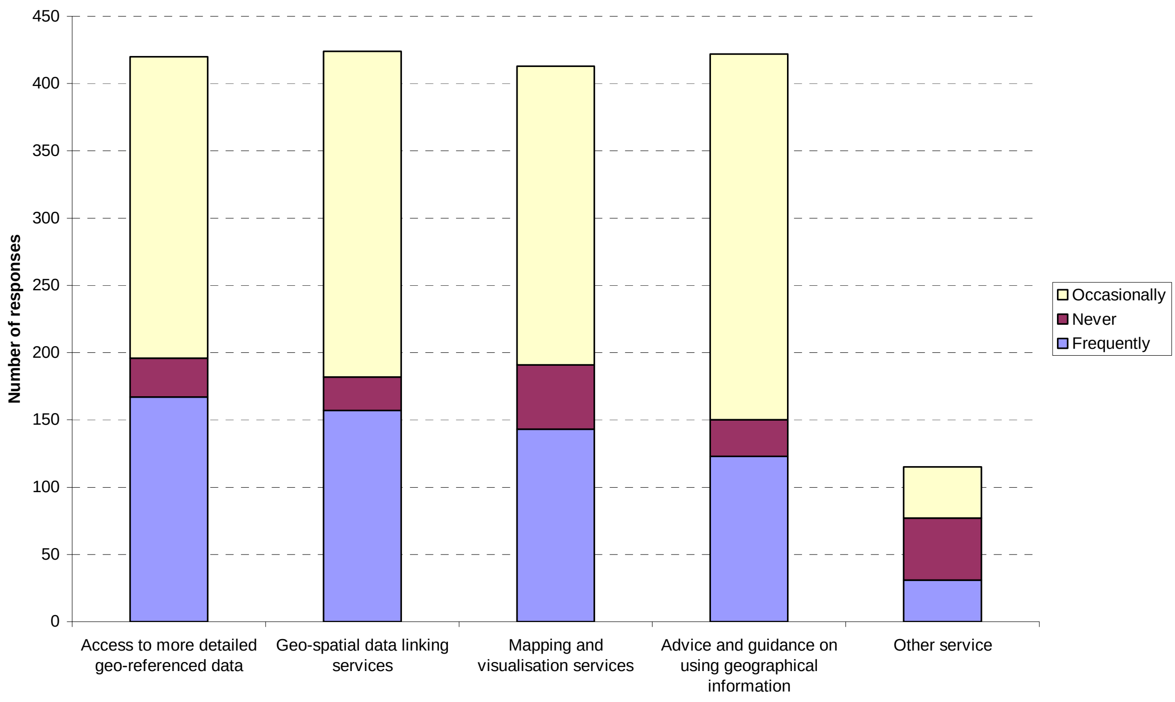
Source: Online consultation to economic and social science community, University of Warwick.

Figure 5.6 Which service would be most useful?



Source: Online consultation to economic and social science community, University of Warwick.

Figure 5.7 How frequently would services be used?



Source: Online consultation to economic and social science community, University of Warwick.

Figure 5.7 shows how frequently services would be used. The most frequently used would be ‘access to more detailed geo-referenced data’ and ‘geo-spatial data linking services’. A disaggregation by user type (not shown in the chart) indicates that expert users would expect to make use of all services, with the exception of ‘advice and guidance on using geographical information’, more frequently than non-expert users. The majority of non-expert users would expect to use all of the services specified ‘occasionally’.

The online enquiry also gave respondents the opportunity to express more general views on the need for such services via an open ended question. Typical of the comments made were the following, highlighting ***key needs to enable (potential) users to make better use of geospatial information***:

A need for better meta data and advice

> Access to geospatial data has improved tremendously in the last 5 years... but more metadata is needed and useful advice on the development / structure of the datasets... I have had difficulty with some EDINA datasets and had I known more about their design/purpose etc it would saved a lot of time. Still too much Black Arts knowledge being kept close to the chest by some suppliers
>
> (competent computer user, knowledgeable about geospatial information, Earth Sciences academic)

A desire for advice in non-specialist language

> Guidance in layman’s terms how to use specific data would be really useful.
>
> (routine computer user, limited geospatial information knowledge, local/regional government officer)

Mapping, visualisation and facilitation of access to data

> It would be useful to get training on mapping and visualisation techniques and facilitate access to ONS data at local authority level.
>
> (competent computer user, limited GI knowledge independent research unit)

An academic wishing to help persuade colleagues to make use of geographical information in their research

> Thank goodness! I have been howling into the wind about the use of such data for as long as I can remember for Psychology but none of my colleagues seem to realise the use and value of it! Taster days are essential in this area in order to really stimulate exciting moves in research paradigms!’ (competent computer user, some GI knowledge, Psychology academic)

The relevance of geospatial information across disciplines – hence the need for a joined-up approach

> There is a critical need for a joined-up approach to geospatial resource provision. Location is pervasive and pretty much every discipline has an interest, such that few data are the sole province of one discipline or discipline grouping. For example, flood risk can have a major impact on demography and deprivation, yet the data to investigate such issues comes from two separate research councils (ESRC and NERC), JISC services and National Statistics. The case for expenditure on almost any specific geospatial resource will almost always be better made by an

academic consortium than any one 'special' interest group, as represented, for example, by a research council. The need to work together is amply illustrated by the issues arising from the UK Location Strategy: all interested parties need to work together.

(advanced computer user, knowledgeable about geospatial information, earth sciences academic/support)

Benefits of a 'one stop shop'

The creation of a single dedicated 'one stop shop' - a Geoservice - in conjunction with existing ESDS services would be of great value to current and future researchers in this fast developing field. It would provide them with easily to obtain quality advice and resources including, if situated or linked to the UKDA, secure access to archived research projects and data.

(compenent computer user, knowledable about geospatial information, archivist)

6. Confidentiality and security

Key messages

- Users face constraints in accessing and making full use of geospatial data due to licensing issues, confidentiality concerns and statistical disclosure control requirements.
- More experienced users are particularly frustrated by lack of detailed spatial codes being made available for use in data analyses.
- Some researchers are likely to find such obstacles off-putting, and particularly if they lack guidance in dealing with them.
- There are fewer constraints in dealing with data from 'non-traditional' sources.

6.1 Key issues

There are a number of issues under this heading which lead to constraints on the user of geospatial data. These might be summarised as:

- licensing issues;
- confidentiality;
- monitoring of data use; and
- statistical disclosure control.

The use and re-use of geospatial data is heavily constrained by copyright and licensing issues. Data and software produced by commercial companies is supplied subject to copyright and license agreements which restrict the uses which can be made of data and its redistribution. The Ordnance Survey also has a restrictive licensing policy aimed at protecting revenue from sale of map data. On the other hand, most data produced by the Office for National Statistics is generally covered by the Stationery Office 'click-use' licence which makes the data available free for most users but restricts its use for commercial purposes. Use of the 2001 Census data was encouraged by the distribution of Ordnance Survey administrative boundary data, but the free distribution of Census data was initially threatened because this conflicted with the much more restrictive nature of OS licences.

ESRC, the Joint Information Systems Committee and CHEST[33] play an extremely important role in negotiating licence arrangements on behalf of the academic sector which limit the constraints imposed by licensing issues and which permit access to commercial data to the academic sector on very advantageous terms. However, there is still a major problem of access to data for academics in institutions where there is a small number of users of a data set, and the institution does not feel justified in paying the licence fee (an example would be EDINA Digimap data). The devising of access arrangements which maximise the use of data will always conflict with the need of suppliers to protect revenue and to prevent the inappropriate use of data.

While digital boundary data do not often have problems of disclosure control because they describe physical objects, geospatial data containing the attributes of particular areas can be confidential or can potentially disclose data on individuals. Following a series of well publicised breaches of data security, and with heightened public concerns over the safe-keeping of data they entrust to public bodies, the statistical offices of the UK and all government departments are adopting a much more active policy towards data security, disclosure control and the monitoring of data usage. More stringent standards are being applied to the publication of data from surveys to ensure that published tables cannot inadvertently disclose data on individuals or that permits attempts to identify the individuals described by statistical tables. This has led to a reduction in the detail published and the suppression from individual-level datasets of those variables most likely to enable individuals to be identified[34]. This problem is most severe for sample data sets, but the Census of Population is also subject to measures to protect the confidentiality of respondents because data for small areas may only refer to a few people resident in the area. The response of the Census offices has been to adopt various approaches to increasing the level of uncertainty in the data. Unfortunately, this has often had the effect of making the data less useful for analysis at the most detailed spatial scale.[35]

Another response has been to limit the amount of geographical information provided in individual-level data sets. Thus the Samples of Anonymised Records from the 2001 Census and copies of the Labour Force Survey made available through ESDS only include information on the region of residence of an individual. This greatly limits the potential for

33 See http://www.eduserv.org.uk/licence-negotiation/agreements
34 An example of this is the reduction in detail of the age and occupation variables in the individual data files from the Labour Force Survey.
35 Examples of this are the 1981 Census Special Migration Statistics and the 2001 Census journey-to-work data for Output Areas.

using these sources in geographical analysis, which conflicts with the increasing recognition of the need to take detailed individual and locational circumstances into account in socio-economic data analyses and the growing capability to analyse data at the individual level. Legislation has recently been introduced which allows the Office for National Statistics to share between departments individual-level databases containing postcode and address details, thus facilitating potentially powerful longitudinal analysis of the circumstances of individuals to be undertaken (*e.g.* the DWP Longitudinal Study). However, the perceived high risk of disclosure of confidential data on individuals has resulted in little use being made of these geo-referenced data sources within government and has effectively prevented their development as geo-spatial research resources.

In contrast, as was indicated in section 3 and 4, there is the prospect of a wide variety of different type of geospatial data becoming available from non-traditional sources (*e.g.* GPS track logs, user-provided georeferenced data) and being made widely available via web-based services, often with little thought given to the potential confidentiality implications.

6.2 Addressing key issues

The issues discussed above are of key importance if the potential of developments in computer processing and software capabilities are to bear fruit in improved analysis of socio-economic data, spatial analysis and the analysis of policy effectiveness.

One response to reconciling the need for access to detailed data with the need to preserve confidentiality has been to create secure environments for data analysis. An early example of this is the Census Longitudinal Study for England and Wales. This is a very large and complex data set, with highly sensitive information on individuals and families with a high risk of disclosure. Researchers have no direct contact with the data and produce a specification for data extraction which is undertaken by a dedicated team of specialists using data in a secure setting. The published outputs produced are checked to ensure confidentiality is preserved. The same approach has been adopted by the longitudinal Census linkage studies for Scotland and Northern Ireland. This offers also limited potential for geographical analysis, because spatially detailed information is likely to have too high a disclosure risk to be released.

The ESRC has recently set up an Administrative Data Liaison Service, based at the Universities of St Andrews, Oxford and Manchester to provide access for the academic sector to some of the detailed data being generated by the government. This Service will not hold data, but will be a repository of knowledge and skills about a variety of administrative data sources and will facilitate their use by researchers.

Another example of the secure data analysis environment is the ONS Virtual Microdata Laboratory (VML) in which authorised researchers can access sensitive data sets in a secure setting in a government building, but cannot remove data and all outputs are monitored by staff running the service. The ONS has a VML for access to individual and organisation-based data from surveys and the Controlled Access Microdata Survey. GIS software (ArcView) is available, but the limited ability to merge other data sets and create new derived data sets mean its usefulness is probably limited to being a mapping facility (*e.g.* for mapping locations of enterprises from the Annual Business Inquiry). In the United States, the Bureau of the Census has set up a series of regional Census data centres for access to microdata held by the Bureau.

The ESRC is establishing a Secure Data Service for academic users run by the University of Essex, but intended to be available via remote access from the researcher's own desktop (at their institution), which will make using such a service more convenient. The speed of response of the service should be much better than the ONS VML because it will run over the SuperJANET network rather than the slower ONS network. Spatial data analysis could be effectively undertaken in a secure data environment if the network was rapid, other data sets could be imported and disk space was available for derived data sets to be saved. The checking of data generated from GIS analysis for confidentiality would probably prove challenging for those in charge of the secure environment.

The implication of licensing arrangements, confidentiality and security requirements creates another set of obstacles that the potential researcher must overcome if they wish to gain access to and utilise spatially-referenced data sets that are placed within these environments and/or subject to more stringent security controls.

7. Geospatial data quality

Key messages

- The quality of geospatial data varies in accordance with the source, data collection standards and the form in which it is made available to researchers.
- There are many examples of users 'making do' with data of dubious quality without necessarily realising the impact of this on their analyses and interpretation of results.
- It is important that users are encouraged to have a greater awareness of data quality issues.

7.1 Key issues

The most salient issues for geospatial data quality are ***the source of data*** and the ***standards to which it has been collected and/or is made available to researchers***. The established sources of data are the national statistics and mapping agencies, local and central government departments, private sector companies and academic bodies, who apply professional standards to data collection. As noted in Section 4, recently new providers of geographical data have emerged whose data collection standards are less clear and who document the data they provide less comprehensively (*e.g.* Google). Emerging developments for data capture such as 'crowd sourcing' pose further challenges for ensuring the quality of the data available for analysis.

7.2 Data quality and fitness for purpose

Though national statistical agencies and other bodies tasked to provide data for research and policy purposes work to the highest professional standards in collection, processing and presentation of data outputs, these products are still subject to sampling error and other sources of uncertainty. Some of this is inevitable – for example any boundary or coastline cannot be recorded with absolute accuracy and even the most detailed digital or physical map will generalise physical reality to some degree.

Because it is difficult and costly to undertake surveys which are comprehensive in terms of topic and population coverage, many social and economic phenomena are measured using data derived from administrative processes. However, the processes through which individual and household data are collected for administrative systems are often designed without reference to the analysis of such data. A case in point is the monitoring of east European migration trends. The main statistical survey of migrants – the International Passenger Survey – is not designed for this purpose and has only a small sample size. Thus policy-making has drawn upon analysis of the Workers Registration Scheme which does not use ONS standards for recording occupation and industry, is ambiguous about where a registered worker lives and works, cannot identify repeat registrations and cannot trace workers within the country, let alone identify when they have left. Similarly, analysis of unemployment trends using the data published from the official count of people registered for unemployment benefit has to contend with non-coverage of the unemployed without benefits and the effects of numerous changes in entitlement to benefit. Data from the commercial sector is similarly biased towards customers of the organisation from which the data is sourced (*e.g.* a particular credit or store card) or self-selected respondents to surveys (*e.g.* shopping surveys).

In the mapping arena, large commercial providers are able to undertake their own survey work or outsource it to specialist working to a specified standard. Local government maintain their own property registers and utilities companies maintain records of the location of the physical services they provide. However, all of these are subject to error, inconsistent recording and duplication. There are errors in the National Statistics Postcode Directory and the Postal Address File which is fundamental to much geographical marketing activity and survey sample design, compounded by the occasional local creation of new postcodes to reflect changes in physical development. The 2001 Census of Population seriously undercounted the population of Manchester and Westminster because the property registers used were very out-of-date and in the case of Manchester missed much recent new development. There are now two competing national property/street address registers, which have differences in coverage. ONS is creating its own property register through on the ground surveying by its own field staff for the 2011 Census because of its lack of confidence in the products available. Satellite navigation systems are notorious for inaccurate recoding of the names of features, the topology of road networks and for introducing a degree of error into the location of roads.

The emergence of the new data sources and new mapping possibilities discussed in Section 4 raises a fundamental contradiction. People drawn in to mapping by 'neogeography' often expect that data of tremendous detail and capability is available free. Some data providers have challenged the need for the continued existence of national mapping agencies who are constrained to impose high charges for their data as a result of national governments expecting them to raise revenue. However, the latter produce data of very high quality, fully documented and regularly updated, while 'free' mapping data is subject to considerable uncertainty in terms of the time period it relates to, the standards to which it was collected and the degree of error incorporated. The examples mentioned above of user contributed data (*e.g.* for OpenStreetmap) depend upon the diligence and competence of the person contributing the information. That said, communities of users (especially when highly motivated) can be extremely effective in checking and correcting data.

For many purposes absolute accuracy of geospatial data is not required. National-level mapping and most spatial statistical analysis can be undertaken with data at a high level of generalisation. Indeed, it is necessary to generalise digital map data when presenting it in cartographic fashion. H owever, errors in the location of geographic objects can disrupt GIS analysis, and when overlaying one data set on another (*e.g.* when vector data is overlain on raster data) small differences in co-ordinate systems or bias in digitising can have a substantial impact on the results of analysis or the appearance of maps. The degree of error in one data set is compounded when it is combined with another subject to a different type of error.

The implication of the issues concerning data quality outlined above is that users of geospatial data need to have an awareness of ‘quality’ issues and use data appropriately.

8. Skills

Key messages

- Technical and substantive skills are needed for effective use of geospatial data.
- The main aim of skills training must be to promote the effective and appropriate use of geospatial data to address substantive research questions.
- A variety of skills and knowledge transfer mechanisms need to be considered, varying from short courses to the provision of some form of centralised or networked service provision.
- Demonstration projects may be a valuable means of illustrating the potential for use of geospatial data.

8.1 Skills for whom?

The effective use of geospatial data is dependent on possession of adequate technical and substantive skills, but these will differ among those playing different roles in the production of knowledge based on geospatial data. As the range of data available and the possibilities for analysis expand, the range of skills necessary also expand and make collaboration between specialists more of a necessity.

The specialists involved will include at one extreme those concerned with data capture, through the designers of database systems, GIS software and web-based delivery systems, archivists of digital data, GIS analysts, and statisticians. Researchers, teachers, students and users of analyses based on geospatial data will be variously represented within these categories.

There is considerable variation between researchers in terms of their level of expertise in terms of knowledge and analytical techniques for using geospatial data. The primary concern for skills training will be with ‘novices’ and those whose knowledge of geospatial data and its analysis is outdated. Those who are ‘expert’ would be seeking to improve their skills in specialist areas. The main aim of skills training must be to promote the effective and appropriate use of geospatial data to address substantive research questions.

8.2 *Types of skills needed*

The list of potential skills required is long and diverse, including:

- Understanding of co-ordinate systems and capture of geospatial data;
- Survey design and analysis;
- Computer programming;
- Awareness of geospatial data and methods of accessing such data;
- Data manipulation skills;
- Web programming (*e.g.* creating 'mashups');
- Proficiency with GIS systems;
- Statistical analysis skills;
- Understanding of the limits of statistical data;
- Understanding of the properties of spatial data, *e.g.* the implications of different spatial levels of data capture and generalisation for data analysis and the implications of 'the modifiable areal unit problem' for the interpretation of spatial patterns;
- Principles of cartography and design;
- Aptitude for documenting each stage of the analysis.

An individual user is unlikely to have to encompass the full range of these skills, mostly focusing on bringing data on spatial entities together with attribute data in a GIS or similar system and undertaking analysis which is then presented in the form of a map. While only a minority need be proficient in programming, everyone using maps should have some appreciation of how to avoid producing 'bad' or misleading maps. This would include avoiding emphasising large ratios/changes in small populations or large sparsely populated areas.

8.3 *Skills and knowledge transfer*

The types of skills required by an individual involved in geospatial data analysis are strongly influenced by their position in the continuum of roles outlined in 8.1. Individuals will tend to specialise towards data capture, data manipulation and GIS presentation, spatial analysis and descriptive analysis/mapping. These skill sets will tend to be reflected in distinct user communities, which will have their own conferences, journals and information circulation networks. Training courses would need to be directed towards these communities of interest.

It is difficult to identify routes/mechanisms through which 'generalists' would acquire the skills necessary. GIS training courses would cover some of the ground listed above, but would probably involve too much investment of effort and funds for the majority of nonspecialists. A variety of short courses could be made available through existing organisations such as the National Centre for Research Methods. Alternatively, the approach adopted by EDINA, appointing ‘site representatives’ for subscribing institutions, could be used to assist potential users to acquire the skills and knowledge needed to engage in data capture, data manipulation, spatial analysis and descriptive mapping. This would, in turn, require a significant contribution from participating Higher Education Institutions (HEIs) and, without some stimulus from the Funding Councils, is likely to result in an inequitable distribution of provision across HEIs

A different approach would be to set up a regional network of centres in which the full array of skills mentioned in 8.2 are represented and where researchers would have the opportunity to gain hands-on experience with geospatial data analysis techniques and build up their skills by bringing their own data to analyse under the guidance of experts. While this would help transfer the skills and knowledge required by potential users of geospatial resources, the costs would be high, with little guarantee that potential users would participate.

Demonstration projects, especially those which result in online tools and guides (e.g. ConvertGrid[36]), could also be useful, particularly if they could be presented in a manner in which they could be adapted to the problem which a researcher is interested in. This means that the methods for achieving each part of a project should be presented explicitly in such a way that the trainee can understand how to implement each part of the project in terms of their specific research aims.

9. Recommendations

Key messages

- The full potential of geospatial data is not being realised by economic and social scientists is not being realised by economic and social scientists at the present time.
- It is our view that, alongside the provision of more training in the use of geospatial resources for research purposes, the ESRC should take the lead in seeking to establish, in collaboration with other funding bodies, a ***Geospatial Resources Advisory Service***.

On the basis of the preceding review it is evident that ***the full potential of geospatial data is not being realised by economic and social scientists*** at present. This is no surprise, given the pace of change in technology and in the policy environment, as well as issues surrounding access to and use of geospatial data for different purposes. It is clear that there is an expressed

36 See http://www.ncess.ac.uk/research/pilot_projects/convert/20060629_cole_ConvertGRID.pdf and http://pascal.rcs.manchester.ac.uk:9080/convert

need from non-expert users to navigate their way through a seeming maze of rules and protocols in order to gain access to certain spatial data, even before assessing how best to analyse them and interpret the results. Yet many research questions – particularly at the boundaries of social, environmental and medical sciences – would benefit from utilisation of geospatial data, not only to provide perspectives on particular places and patterns of spatial variation, but also to enable linkage of data across domains, such as:

- health and the environment;
- employment and health;
- variations in educational provision and outcomes;
- transport and economic deprivation;
- other environmental challenges (recycling, waste, energy use, shopping habits, etc.)
- access to services, deprivation and well-being; and
- changes in the natural environment and impacts on the built environment.

The problem rests with the non-expert users and their ability to take advantage of the potential of geospatial data to enhance the range, quality and sophistication of their research. Two different approaches are considered. The first would be to extend the range, type, availability of skills and knowledge transfer activities across the UK, encouraging researchers to identify the training and skills development activities that would be useful to them. In the preceding section a set of options was identified, ranging from the provision of more short courses by existing providers to the establishment of dedicated centres for the development of geo-spatial research skills. While we would encourage training providers to establish more short courses in relevant areas, we are of the view that this would not resolve the problem simply because the 'non-expert user' is not in a position to identify the skills and expertise that are needed and probably has little knowledge of the variety of resources that could be utilised to enhance their research.

It is our view that, alongside the provision of more training in the use of geospatial resources for research purposes, the ESRC should take the lead in seeking to establish, in collaboration with other funding bodies, a ***Geospatial Resources Advisory Service***. This service would add value to existing and potential research across the social and economic sciences and at the boundary with other scientific disciplines. The aim would be to provide a 'one stop shop' at which researchers would discuss the nature of the research issue that they were addressing and would receive the best advice on relevant data sources, geography, analytical tools, mapping and data linking. By having a single centralised location and a virtual presence via the web, the service could efficiently provide:

- access to more detailed geo-referenced data and metadata;
- advice and guidance on using geographical information;
- geo-spatial data linking services;
- tools to access actual or simulated data on individuals (*e.g.* daily activity data which might originally be derived from credit card or mobile phone data) without the need for advanced programming skills;
- assistance/tools to help researchers create 'mashups' and location-based services as a way of presenting their findings, or visualising research data;
- assistance with making use of more advanced mapping and visualisation facilities; *e.g.* 3-d visualisation, 'virtual worlds', simulations.

The service would work collaboratively with other training providers in both the academic and commercial sectors, hopefully negotiating preferential rates for provision from the latter. It could be charged with responsibility for academic liaison with the UK Location Council, by establishing and maintaining an inventory of geospatial datasets created via the recipients of UK research awards, (providing common geographical reference data and by sharing location-related information). It would also act as a focal point for knowledge about such courses and promote participation from those who seek its services. It could serve as a centre for 'showcasing' good practice in applications of geospatial data in a research context, developing online demonstrator tools that would have general application.

Recent experience with the establishment of the ESRC Secure Data Service and the Administrative Data Liaison Service suggests that it would be appropriate to set up the new Geospatial Resources Advisory Service on a three year basis in the first instance, extending this to five years on the basis of a review of its first two years of operation. To provide the range of services listed above, a three year budget of at least £750,000 would be required.

ANNEX: QUESTIONNAIRE USED IN WEB CONSULTATION OF CURRENT USE OF AND NEEDS FOR GEOSPATIAL RESOURCES AND SERVICES

1. Introduction

In support of the National Strategy for Data Resources for Research in the Social Sciences, the Economic and Social Research Council has commissioned a review of the current availability of, access to, and future needs for geo-spatial data services by the research community. These are broadly defined and include not only the data necessary to map and analyse data with a spatial dimension, but also resources which broaden the range of spatial research undertaken and enhance the ability of researchers to undertake spatial analysis.

To assist us with this task, please spare a few minutes to answer the questions that follow – **even if you think they are not immediately relevant to your situation**.

Any information you give will be used purely for statistical purposes in order to plan future services. This survey will remain confidential and any published results will not identify individuals.

The survey should take 5 to 10 minutes to complete and will remain open until 31 January 2009.

Thank you for your time.

2. Demographic information

*** Which category best describes your current position?**

*** How would you best describe your disciplinary background? (please pick the closest match)**

3. Use of computers and statistics

*** Which of these categories best describes your use of computers and statistics in your current role?**

- ◯ Not used
- ◯ Routine (e.g. using a computer for word processing and/or spreadsheets or communication with others by email)
- ◯ Complex (e.g. using a computer for gathering and analysing information or design, including use of statistical packages)
- ◯ Advanced (e.g. programming)

4. Spatial data analysis and mapping skills

*** How familiar are you with spatial data analysis and mapping techniques?**

- ◯ Not at all
- ◯ Some knowledge but little proficiency
- ◯ Knowledgeable and can undertake some such activity
- ◯ Knowledgeable and proficient in these areas

5. Use of data sources

*** Do you currently use any data sources containing spatial information? (for example, Census of Population data, directory data, OS data, geographical look-up tables, etc.)**

- ◯ Yes
- ◯ No

6. Data sources used

*** What data sources containing spatial information do you currently use?**

- ☐ Administrative data (e.g. benefit records)
- ☐ Census of Population
- ☐ Digital boundary data/Road network data
- ☐ Directory data (e.g.postcode directories)
- ☐ Environmental datasets
- ☐ Geo-coded health data
- ☐ Labour market data
- ☐ Official survey data
- ☐ Remote sensing data
- ☐ Other (please specify below)

*** Which of the following do you use to access these data?**

- [ ] UKDA/ESDS
- [ ] Other academic repository
- [ ] Research sponsors
- [ ] Official statistics web site (e.g. ONS, Neighbourhood Statistics, Nomis)
- [ ] Government department web site (e.g. Defra)
- [ ] Own data collection
- [ ] Other (please specify below)

7. Use of spatial data sources

*** Do you currently use any resources to enable you to carry out mapping or spatial analysis? (e.g. OS data, administrative boundary data, remote sensing data, etc.)**

- ( ) Yes
- ( ) No

8. Spatial data sources used

*** What resources do you use for mapping or spatial analysis?**

- [ ] Administrative boundary data
- [ ] Geographical look-up tables
- [ ] Grid square data
- [ ] OS data
- [ ] Other digital map data
- [ ] Remote sensing data
- [ ] Other (please specify below)

*** How do you access these resources?**

- [ ] UKDA/ESDS
- [ ] EDINA
- [ ] Mimas
- [ ] Other academic repository
- [ ] Research sponsors
- [ ] Official statistics web site (e.g. ONS, Neighbourhood Statistics, Nomis)
- [ ] Government department web site (e.g. Defra)
- [ ] Own data collection
- [ ] Other (please specify below)

9. Research interests

What issues or questions are you interested in researching or analysing using geo-spatial resources?

10. Spatial data services

The ESRC wishes to determine whether there is a demand for enhancing existing and developing new spatial data services which are currently lacking.

* Which spatial data services would you find useful?

- [ ] Access to more detailed geo-referenced data
- [ ] Geo-spatial data linking services (e.g. look-up tables between different geographical units, linking spatial indicators to existing data, etc.)
- [ ] Mapping and visualisation services/facilities
- [ ] Advice and guidance on using geographical information
- [ ] None
- [ ] Other (please specify below)

11. Frequency of use

* How often do you anticipate you would use them?

	Frequently	Occasionally	Never
Access to more detailed geo-referenced data	○	○	○
Geo-spatial data linking services (e.g. look-up tables between different geographical units, linking spatial indicators to existing data, etc.)	○	○	○
Mapping and visualisation services/facilities	○	○	○
Advice and guidance on using geographical information	○	○	○
Other (as specified in previous question)	○	○	○

* Which service would you find most useful?

- ○ Access to more detailed geo-referenced data
- ○ Geo-spatial data linking services
- ○ Mapping and visualisation services/facilities
- ○ Advice and guidance on using geographical information
- ○ Other (as specified in previous question)

12. Comments

If you have any other comments please enter them below

13. Thank you

Many thanks for helping us with this enquiry.

If you have any questions about the review, please contact us at the email addresses shown below.

A short anonymised report will be prepared from these and other responses. If you wish to receive a copy of this report, please insert your email address below. Your email address will not be used for any other purpose.

My email address:

Peter.Elias@warwick.ac.uk
Anne.Green@warwick.ac.uk
D.W.Owen@warwick.ac.uk

Please remember to click the 'Submit' button