Geodata

Tobia Lakes

July 2009
Abstract

Empirical data can be characterized by a precise location in space and time. An estimated 80% of all data holds such a spatio-temporal reference and is termed geodata. This paper starts with the question: What is the additional benefit for socio-economic sciences using geodata and the spatial dimension respectively? In the following a multidimensional approach is chosen to outline the Status Quo of geodata and spatial techniques in Germany. It is particularly the continuously growing amount and the variety of available geodata which is stated. Data security is an issue of high importance when using geodata. Furthermore, the present developments in price and user concepts, accessibility, technical standards and institutionalisation are addressed. A number of challenges concerning the field of geodata are identified including the open access to geodata, data security issues and standardization. The main challenge however seems to be the exchange between the rather segregated fields of geoinformation and the information infrastructure. Furthermore, the census 2011 is identified as a major challenge for the acquisition and management of geodata. Geodata and spatial techniques are a rapidly developing field due to technology developments of data and methods as well as due to recently growing public interest. Their additional benefit for socioeconomic research should be exploited in the future.

Keywords: geodata, geoinformation, Web-GIS, geodata-infrastructure, spatial techniques
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 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)
1. Introduction

Recent research issues such as climate change impact on mankind, megacity development, disparities between the rich and the poor, environmental justice, security all have one in common: they benefit from empirical studies and hence, they are critically relying on empirical data (IPCC 2007; UN Habitat 2009, EC 2008). Empirical data about households, source and targets of migration, meteorological data, accessibility of education, range of environmental pollution is for example needed - data which can be characterized by a precise location in space and time. An estimated 80% of all data holds such a spatio-temporal reference and is termed geodata. Using geodata and spatially explicit techniques is well-established in geography or spatial planning as well as in specific subsdisciplines such as social geography or economical geography (Longley et al. 2005). However, until recently, the benefit of spatially explicit approaches by using geodata and Geoinformatics techniques has only rarely been exploited in socio-economic sciences and policy-related research questions (Goodchild & Janelle 2004). In addition to regional data approaches (see also the contribution of Matiaske & Grözinger in this publication) the explicit linkage of data to a location has now been of growing interest, e.g. in the context of the next German Census in 2011.

What is the additional benefit for socio-economic sciences using geodata and the spatial dimension respectively? First of all, geodata is data like every other data set, hence spatial data can provide additional information and therefore should be valued and included in empirical research. In Germany, a large pool of geodata already exists and is continuously enlarged, as will be described further on, waiting to be exploited by new users. Secondly, geodata can add fundamental benefits by offering visualisations in form of maps and database search algorithms based on location. Thirdly, the spatial information offers possibilities to integrate various datasets via the spatial location and examine possible interrelationships between datasets. In a recent study the life satisfaction approach is used to value air quality: individual-level panel and high-resolution SO2 data are combined to identify the effect of SO2 concentration on life satisfaction and housing rents and relate the total willingness-to-pay for improvements in air quality (Lüchinger 2008). Directly georeferenced data is also of particular interest to provide a comparability when repeatedly collected data is based on modified statistical units. Finally, perhaps the most important benefit: spatial analyses enable the inclusion of the context via concepts of proximity, range, containment, overlap, adjacency, or connectedness. Visualization and statistical analyses of
these properties is one way of detecting patterns, anomalies, outliers, and sometimes even causation to generate new insights. Of course, underlying processes cannot be detected, but sometimes approximated. In a recent study the influencing factors of graduate students’ choice on further education was modeled: the distance between the households and the Universities turns out to be of high significance (Spiess and Wrohlich 2008). These spatially explicit analyses can be extended to spatio-temporal modelling approaches which aim at modelling spatial processes in time including likely future developments, such as land use change at the interface between the human and environment system (Lakes and Müller 2008).

In order to do research across discipline boundaries and develop policy-relevant strategies access and analysis of geodata on a national, European, and global scale can therefore be an essential benefit. In Germany, “Geoinformation” is now regarded as one of the most important cross-sectional technologies of this century and a policy field with an outstanding future (Bundesregierung 2008).

2. **Status Quo: Geodata and spatial techniques**

The outlined research questions that may benefit from geodata show that not only the available geodata sets are of interest but also available techniques to handle and exploit the spatial dimension of geodata. Consequently, in this paper data and techniques in Germany within the international context are described. A multidimensional approach is chosen including data availability, influencing factors of data availability (accessibility, technical standards, price and user concepts, data security, and institutionalisation) and spatial techniques.

2.1 **Present situation of Geodata**

The amount and the variety of available geodata is continuously growing in Germany. Regarding the content, geodata can be categorized in spatial base data and spatial thematic data acquired and provided by administrative or private sources. The **spatial base data** contain general topographical and property information and hence offer the basis for most research studies.
Excursus:

Geodata can either hold a **direct spatial reference**, that means the information about the location is defined by two- or three-dimensional coordinates within a coordinate reference system, such as the most frequently used ones in Germany, the Gauss-Krüger and ETRS 89 systems. Data can also hold an **indirect spatial reference**, i.e. systems closer to everyday human experience, such as administrative areas, postal addresses, or place names. In order to digitally process the complexity of real world objects, they must be generalized and simplified: discrete objects and continuous fields are two approaches which go along with a respective data model, the vector and raster model. Points (e.g. trees, cities), lines (e.g. roads, rivers) and areas/polygons (e.g. city-parcels, administrative boundaries) are examples of the **vector model**. **Raster data** consists of cells within a rectangular grid, such as remote sensing data of airborne or satellite systems.

It is **predominately an administrative task to acquire and manage spatial base data** datasets, which are accessible at the Bundesamt für Kartographie und Geodäsie (BKG) and at the survey administrations of the Länder and communes due to the federal system in Germany. The two most important Germany-wide standardised spatial base data sets are (see AdV):

- **ALKIS** (Automatisiertes Liegenschaftskataster Informationssystem, Automated Real Estate and Cadastre Information System) comprises the Real Estate Map, Real Estate Book, and Official House Coordinates.

Objects within these standardized approaches are classified according to a specific hierarchical object catalogue, e.g. 20000 Siedlungsflächen, 2100 Baulich geprägte Flächen, 2111 Wohnbauflächen/2121 Sportanlage. In recent years, **administrative spatial base data have been increasingly substituted** by new approaches of data provision. On the one hand geodata acquired and provided by a world-wide user community via the Internet is of growing importance. The OpenStreetMap Project is an example of this Wikipedia-style information which can be used and updated in a collaborative way by everyone. Beyond question is on the other hand the influence of **private data providers** within the geodata market: up-to-date road networks data (e.g. Navteq, Tele-Atlas), household address data, aerial photos and satellite data are more and more offered by private companies. While aerial photos are still predominately acquired by German companies (e.g. Hansa-Luftbild), the market of satellite data is a global one, i.e. data from worldwide managed satellites are of interest for local studies. Within recent years **remote sensing data has gained the user's**
interest. Partly initiated by new internet-based technologies such as GoogleEarth it became obvious that aerial photos as well as satellite data provide a good data source on first sight, not to name the possibilities of sophisticated remote sensing data analyses also in the context of social sciences (Rindfuss and Stern 1998; Goodchild and Janelle 2004). The available variety of remote sensing data sets is growing, each offering specific advantages depending on the study aim and context: for example very high spatial resolution (Quickbird), very high temporal resolution (Rapid Eye) versus satellite data covering very large areas (Landsat TM). A few companies in Germany have specialized on providing remote sensing data, such as GAF or EuroMap. Remote sensing datasets are also available at the German Remote Sensing Data Center (DFD) of the German Aerospace Centre (DLR).

In addition to topographical and property information of geospatial base data the focus in research and application is predominately put on spatial thematic data, which can cover a wide variety of fields, such as environmental data, employment data or business data depending on the specific aim of a research study. On the administrative side they are collected and used by the Bund, the Länder as well as by communes. While some Federal administrations are experienced in working with a spatially explicit approach, such as the Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit or the Bundesamt für Bauwesen und Raumordnung, others traditionally provide data without or rather very aggregated spatial reference, e.g. the Arbeitsagentur or the official statistics. An important administrative spatial thematic data source for research is particularly the communal level, such as Planning, Forestry, Environment, Statistics, and Police. Of particular interest in this regard is the German Census 2011 which will provide macro-census information that can be precisely linked to location for further analysis. In addition to administrative data, a large amount of geospatial thematic data is collected by the research sector itself. Furthermore, organizations as well as commercial data providers hold and offer a large amount of spatial thematic data. It is particularly in commercial fields that data needs are not sufficiently covered by public data provision: branch specific information, as well as building, socio-demographic, market and consumer, communication, and lifestyle data (Fornefeld et al. 2003). Another important source of data is acquired within the field of geomarketing to name e.g. Pitney Bowes Inc. offering world-wide services in direct-marketing and post-services based on a GIS (MapInfo).
2.2 Present situation of influencing factors of data availability

The most decisive challenge of present usage of spatial base data and spatial thematic data (not only) in Germany is the accessibility of the large amount of geodata that is available, however distributed in several places, acquired and provided by different sources. This situation has been recognized by science, business, administration, and politics within the field of geoinformation and a number of measures to enhance the accessibility has been brought under way. **Geodata infrastructures and geodatabases have been initiated on different levels within administrations and institutions.** The aim of the Geodata infrastructures (GDI) is to improve the access and usage of available geodata. Geodata infrastructure projects are very often connected to the management of geodatabases and internet-based geoportals for user-friendly data provision. Germany-wide a national geodatabase has been set up to hold all geodata which is needed for legal purposes, administrative actions, economic development and research. The aim is to provide access to data of different sectors via standardised web services of the Bund, Länder and communes (in a first step geodata of the Bund is accessible on www.geoportal.bund.de). A second substantial instrument for enabling access to spatial thematic data is the Umweltportal Deutschland PortalU, which offers the possibility to search for environmental information of 120 administrations and organisations via thematic, spatial, and temporal criteria.

An important recent issue are **price and user concepts of geodata with the aim of a transparent and market oriented development** without putting ownership and responsibilities in question. The basic approach of the Federal Government is to assess fees for administrative geodata based on the costs for data provision (Bundesregierung 2008). Milestones in this context are available eGovernment-procedures, e.g. ePayment, as well as legal guidelines, e.g. Geodatenzugangsgesetz, Umweltinformationsgesetz and the AdV-Gebührenrichtlinie.

Technical challenges regarding different specifications and formats have continuously been dealt with by developing standards in geoinformation technology, also as a part of eGovernment. The system independent access on geodata of different levels requires the definition of standards based on the European (CEN) and international standards (ISO, OpenGeospatial Consortium) as well as on national regulations of eGovernment. In 2007 an architecture concept GDI-DE was put into force, which holds service oriented and technical information for the development of future GDI within Germany.

**Data security is of ongoing importance for all types of data** - apart from the fundamental Informationsfreiheits- und Informationsweiterverwendungsgesetz, geodata is a
specific case since the issue of personal rights is particularly sensitive. Up to now there is no consistent approach on how to balance between publishing and non-disclosure of geodata, it generally depends on in how far the personal right of persons concerned is attacked (Karg and Weichert 2007). Recent activities of Google to take photos of street panoramas to provide them in 3-D city models via the Internet have provoked new discussions about data security. Specifically relevant for geodata is the Umweltinformationsgesetz which is the national realization of the European guideline on the public access to information on the environment. Remote sensing data is a particular type of data which offers benefits due to area-wide, comparable, up-to-date data on different issues of the earth surface however it may also be critical concerning data security which are addressed in the recently published Satellitendatensicherheitsgesetz. The data security issue is of high importance but at the same time very complex considering the provision of social and economic data and cannot be fully exploited within this paper (see also the contribution of Schaar on data protection and on record linkage of Metschke in this publication).

The measures to assure the accessibility and an efficient use of geodata are strengthened by **major achievements in institutionalisation of the cooperation** between different levels and types of administrations in Germany, as well as with economy and research sectors using geodata. To name the most important: Interministerieller Ausschuss für Geoinformationswesen (IMAGI), Lenkungsgremium GDI-DE (LG GDI-DE), the commission for geoinformation business (GIWKommission), Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV, http://www.adv-online.de), and Arbeitskreis der Staatssekretäre für eGovernment des Bundes und der Länder (Deutschland Online).

Not only within Germany but also within the international context the cross-border exchange of geodata on a national and international level is of growing importance. The **International interoperability of geodata and geoinformation** has been particularly strengthened by the European INSPIRE-guideline, which is the fundamental guideline for interoperability regarding geodata management and provision and the development of a European geodata infrastructure.

Furthermore, three major **initiatives for newly available data** are of importance within an international context: Galileo, the European satellite navigation system will provide the basis for future referencing of geodata, the localisation and positioning of objects. In 2013, Galileo is expected to offer positioning data which is of interest for multiple user groups. A central platform for future usage of Galileo has been put up with the “Forum für
Satellitennavigation” by the Bundesministerium für Verkehr, Bau und Stadtentwicklung. A second initiative is the Global Monitoring for Environment and Security which is supposed to integrate terrestrial, satellite, airborne, maritime and further data sources for environment, climate measures and sustainable development, as well as for humanitarian, development and security relevant issues. It is a joint initiative of the European Commission and the European Space Agency. Thirdly, the Group on Earth Observation (GEO) should be mentioned, which was initiated in 2005 to build „Global Earth Observation System of Systems“ (GEOSS), to offer better access to earth observation information. A central access point (GEO-Portal) as well as a catalogue service (GEO-Clearinghouse) is envisaged.

2.3 Present situation in spatial techniques

Geodata can be used as any other data set in a statistical software to extract the thematic information it holds, however in addition to the above mentioned techniques for geodata access and distribution, specific spatial extensions are needed to exploit the spatio-temporal dimension of geodata. The specific type of professional software which offers the needed spatial techniques is called GIS (geographical information system). It is a computer system for capturing, management, analysing, and displaying Geodata. GIS includes hardware, software, networks, standards and protocols for data handling and analysis (Longley et al. 2005).

Apart from proprietary software, Open Source GIS and databases increasingly provide an interesting alternative (see e.g. PostGIS, PostgreSQL, GRASS). In addition spatial extensions for frequently used database systems are being used, such as Oracle Spatial and the new releases of SQL-Servers. It substantially depends on the user interests and requirements if a professional GIS is needed or basic tools are sufficient. Of particular interest are Web-Services, which offer basic spatial services without the need of an installed GIS-Software on the user's PC. While basic functions such as map visualisation of decentralized servers via Web-Services are well-established, more sophisticated techniques are still under development and need further research. Furthermore, freely available Internet tools are a growing sector; sponsored user-community-portals, such as Picasa, which offers a service to place photos in GoogleEarth or portals financed by advertisements, such as Map24.de offering navigation data and services. These go along with Navigation and mobile services which have reached operational application level. Accompanied by the development of GPS-sensors in mobile phones and widely spread mobile phone cameras open up new opportunities for location based services as well as for research.
3. Future Developments

Future developments based on the present status Quo regarding geodata and spatial techniques are manifold - only selected ones will be addressed here (see also Bundesregierung 2008).

The **amount of available data sets will continue to grow**, the variety concerning thematics, spatial, and temporal characteristics will increase. It is particularly the development of **new data acquisition technology** which will contribute to the growing amount of data such as more frequently used positioning systems and new remote sensing technologies, to name only the German developments of Terra-SAR-X, RapidEye, EnMAP. Geodata will more and more be acquired by both, public and private data providers; hence, new forms of **public-private partnerships and cooperation** for data acquisition including collaborative web-based initiatives need specific attention. A project of major relevance in terms of georeferenced data acquisition and provision within Germany is the next **German Census in 2011**.

The already initiated **development of internetbased access points for geodata - geoportals**, either administrative or business portals, will continue. The overall goal to build up a national geodata basis with the aim of a demand oriented geodata supply will be one major future task. A further extension of the www.geoportal.bund.de from data of the Federal State towards Länder and communes as a Geoportal for Germany is envisaged by the Geodatenzentrum of the Bundesamt für Kartographie und Geodäsie. Apart from the development of geoportals for administrative data, business geoportals will also grow in number (MICUS 2008).

The recent trend that portals will not be limited to data or meta data, but include **web services**, enabling the direct access to data and thematic map visualisations via the Internet without specialized software products installed will continue. This goes along with the principle of decentralized data in the specific administrations and centralized data provision for the user will continue. With the growing importance of the Internet, coordinated efforts with **eGovernment actions** will be of interest, such as ePayment for geodata. Technical standards need further attention not only within the field of Geoinformation but also beyond specialized science and as part of the eGovernment concept developed by LG GDI-DE.

In addition to new spatial Internet-based technologies spatial extensions of widely used database systems support the trend “**GIS goes mainstream**”. Hence, the user community is expected to grow constantly enhanced by available free and open source products. Furthermore, **new spatial techniques** in professional GIS software offer possibilities to
integrate different data sets or spatio-temporal modelling.

Cost and usage concepts are continuously important issue for the administrative data – a centralized tool for geodata fees for all administrative levels as well as private data providers will be the long-term aim. According to the Federal Government (Bundesregierung 2008) costs for data acquired by tax money will be limited to the indeed cost of provision, enlarging the geodata user group.

Discussions on data security and the need to legally and consistently define data access rights will further increase, such as e.g. in the context of the Census 2011. A consultancy rating of the most frequently used data and a categorization according to data security relevance will take place (Bundesregierung 2008). The draft of a Geodatenzugangsgesetz is under development. It aims at offering geodata and geodata services for free for the federal administration and European Union as far as no further business usage of the data exists.

In the near future Germany also will have to pick up the requirements of international developments (INSPIRE, GMES, GEOSS) and fill those with actions. On a national scale the INSPIRE guideline has to be transferred into German law by 15.5.2009. The data sets proposed in the annex of the INSPIRE guideline have to be recorded until 2019 on all administration levels.

4. Future Developments: European and International Challenges

In an international context, similar however also diverging tendencies are prevalent compared to Germany (Fornefeld et al. 2003). The strategy to optimise access to administrative geodata via inter ministerial organisation by developing geodata infrastructures is an ongoing task within Europe and for example the United States. While in the US this has reached a well-established level, most other countries are still in the process of development. Since internationalization takes place in the field of private data acquisition as well as in spatial techniques provision the global market is growing together. In addition an increasing number of international guidelines on European as well as across European borders requires the comparable usage of thematic geodata, such as the European Union Water framework directive. International initiatives, such as the INSPIRE guideline, GMES and EOSS will thoroughly change the handling of geodata in Germany and Europe. The resulting potentials for geodata usage in research and business have to be exploited as far as possible. These developments across national boundaries lead to an increasing importance of international
exchange of geodata beyond ministerial infrastructures.

5. Conclusions and Recommendations

Geodata and spatial techniques are a rapidly developing field due to technology developments of data and methods as well as due to the new political attention. This makes it difficult to keep track of the present developments, nonetheless it is even more important to regularly analyse the situation and develop recommendations. After all, it needs to be stressed that geodata is data like any other dataset and the artificial separation between geodata and meteorological, juridical, demographic data for example is no more adequate considering that 80% of all information has a spatial reference. It is only the combination of information that offers multiple benefits. Hence, most of the issues addressed by the other expertises are also of relevance for geodata. In the following, selected recommendations concerning geodata, influencing factors of geodata accessibility, and spatial techniques are phrased.

5.1 Geodata and influencing factors of geodata

The amount and variety of geodata is constantly growing, hence, the main challenge is to offer access to geodata in such a way that they can be combined with others to provide information for research and policy (Bundesregierung 2008). A geodata infrastructure based on geoportals is of great significance but also new sources of internet-based and private data provision need to be considered (MICUS 2003; Bundesregierung 2008).

Regarding the data, a reliable update of administrative spatial base data is missing. Even though a 5 years rhythm may be envisaged, in reality it is most often less. Furthermore, it is also the provision of historical data which is of relevance for longitudinal studies, at the best, comparable data. This may be an administrative task since it is not covered by private data providers. Data gaps regarding an areawide coverage of spatial base data in Germany (not to think of Europe or even beyond) should be closed, for example diverging data quality between urban and rural topographical data. In addition, research requires comparable data, hence, object catalogues should be developed in further detail for spatial base data and spatial thematic data. Since linking geodata has been identified as a major task conversion codes between different datasets should be available. Spatial reference is one key to possible data integration, hence, data should as far as possible be equipped with a spatial reference. While indirect reference via postal codes or election districts might be more feasible, the spatial outline and position however can change; hence, a direct spatial reference
seems to be a better solution as far as data security issues are untouched or can be dealt with (see Schaar in this publication). Only then, spatial benefits of thematic data such as official statistics, microcensus or particularly the Census 2011 can be fully exploited. **Concepts for the Census 2011** are necessary to enable linkage and integration of Census data with further data sets based on explicit georeferencing via the personal address on the one side and at the same time to prevent the extraction of individual level information via techniques such as data aggregation on grid or thematic object based level.

**User rights**, particularly for reusage and further usage as well as regulations for fees and **price models** within Germany but also international European wide and beyond need to be pursued.

Consistent and up-to-date technical standards continue to be an important subject.

With the growing amount of available data and the enhanced combination of data of different sources, **quality measures for geodata** sets have to be striven for. Users miss reliable measures of available datasets for administrative data as well as for privately offered data. Imperfect data is better than no data, however, it is of particular importance to estimate possible limits of explanatory power.

### 5.2 Spatial Techniques

On the one side sophisticated spatial analysis and integration of geodata with additional data within interdisciplinary projects opens up new research opportunities and needs to be exploited. On the other side, mapping techniques for non-professionals offering user-oriented techniques for their specific tasks are a challenge and require an overview of existing software and tools for non-professional spatial analysis supported by best-case-studies.

New techniques such as freely available GoogleMaps, Picasa or GoogleEarth open up a wide field for **internet-based data access and tools** which need to be exploited. However, issues of **quality and reliability of publicly available and free tools** need to be investigated.

### 5.3 Politics

The main challenge seems to be the exchange between the rather segregated field of **geoinformation and the information infrastructure** such as the RatSWD. Parallel developments in terms of GEOdatainfrastructures, GEOdataportals, and GEOinformationmanagement should be integrated within a German approach of an overall information infrastructure. A **round table on geodata/regional together with the RatSWD** should be initiated to bring together the different approaches and initiatives of data providers.
and research.

Issues to be addressed should comprise user and price concepts, data security, as well as technical standards developments to further enhance data exchange. Also, exchange is needed within the Geoinformation sector between the public, private and the research sector to get new impulses for and out of research.

**Awareness of opportunities of geodata and spatial techniques** is the prerequisite for a successful usage in transdisciplinary if not interdisciplinary socioeconomic and policy related research. **Joint research projects** going along with presence in journals and media should be initiated to exploit the potential of integrating geodata in integrated analyses. Integrative modules across department (and thematic data) boundaries within Universities may be one possibility. **International exchange** should include successful initiatives of geodata usage within scientific information infrastructures such as the Center for Spatially Integrated Social Science (USA, http://www.csiss.org/), the SEDAC, Socioeconomic data and application center (USA, http://sedac.ciesin.columbia.edu/) or the Center for Geoinformation (Ireland, http://ncg.nuim.ie/).
References: