

Geoinformatics and e-Science

Ralf Bill, Rostock

ABSTRACT

The article introduces science paradigm changes through the ages up to the present data-driven science. Selected terms within the field of today's e-Science are explained. The role of spatial information sciences in the context of this data-driven science is illustrated. Several case studies and examples are presented.

1. SCIENCE PARADIGMS AND E-SCIENCE

1.1. Changing science paradigms over time

Throughout human history, the vision of science and its recognized assumptions and ideas (the so-called scientific paradigm) has been amended several times. Scientific paradigm means a reasonably coherent bundle of theoretical guidelines, concepts, thought patterns, methods, postulates and questions, shared by many scientists at any particular period of time in the development of science. KUHN (1970) defined it as “universally recognized scientific achievements that, for a time, provide model problems and solutions for a community of practitioners.” In the textbook written by HEY/TANSLEY/TOLLE (2009), the four science paradigms over the course of time are described as:

- Thousands of years ago, science was purely empirically oriented. Science tried to observe and describe natural phenomena. [Experiment]
- Over the past hundreds of years, science was based on theories. Science developed models and generalizations of these. [Theory]
- In recent decades, the simulation of complex phenomena was introduced into science. Information technology made it possible to model and simulate. [Simulation]
- Today, the science is using more and more explorative methods, combined with experiment, theory and simulation. Based on the exploration of massively existing or easily to be collected data this paradigm is often called “data-driven science” or “data-intensive science” (GRAY, 2005, 2009)). [Data]

Digital availability and computational analyzes characterize the rise of data-driven science, allowing for the scientific identification of statistical patterns from an endless amount of data. This approach complements the traditional theory and hypothesis-driven science that emanates from creating theories, proofing hypotheses, designing survey instruments, collecting data and finally analyzing these large amounts of data (<http://irights.info/artikel/information-mining-aus-dem-steinbruch-der-wissenschaft/11499>). Prerequisite for such – paradigmatic understood “data-driven science” – is a systematic management of resources, in the sense of infrastructure, regulations, and – what is often forgotten – additional human resources and their skills.

1.2. e-Science

Today's knowledge and information society with its diverse possibilities of effective communication, ease of access to huge amounts of information, and powerful computing technology is producing a new challenge for scientists. Chances to achieve qualitative and quantitative improvement of the scientific results with the new methods have increased significantly. In parallel, the difficulty to control such distributed, dynamic system components has risen.

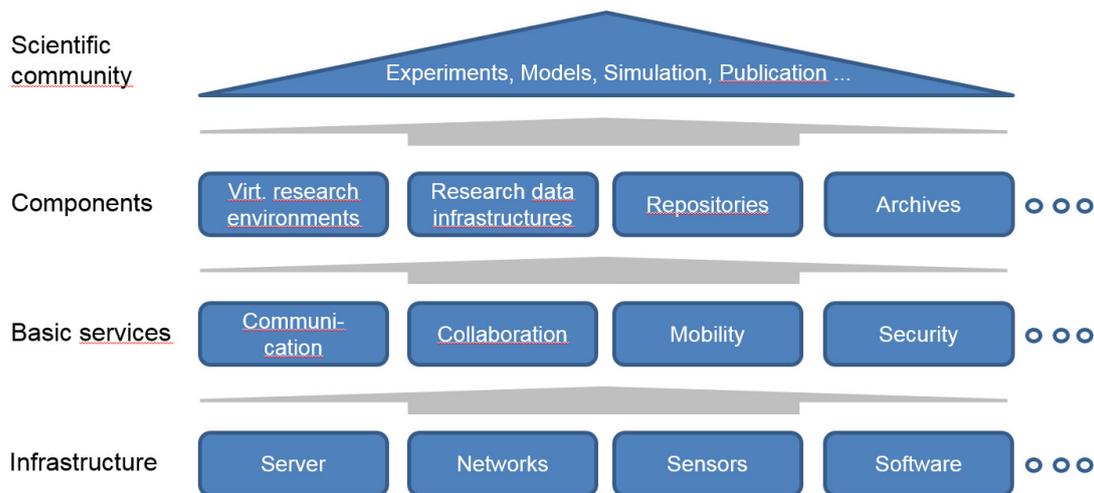


Figure 1: e-Science Framework.

In addition to the data, in today's science the collaboration between scientists plays an essential role. Today the term “e-Science” (Electronic Science) or alternatives such as “(digitally) enhanced science”, “net-based science”, or in the US “Cyberinfrastructure” is established. e-Science is designed to support the entire scientific life cycle. It is aimed at a collaborative research platform based on a given digital infrastructure integrating all relevant resources for a research topic and providing generic services and discipline-specific components for scientists. The e-Science framework (Fig. 1) organizes the distribution of computing resources (such as servers, networks, sensors, software, data, information, etc.) and offers basic services for communication, collaboration, ubiquitous usage and security, which are required in all scientific disciplines. Special components, which will be explained later, combine the basic infrastructure and services to create research environments for scientists supporting their experiments, models and simulation, but also the publication process and the exchange of results.

Through e-Science, there is growing understanding that the value of the research is mainly in the data. Therefore the scope of work should be expanded on the primary object, the research data (BÜTTNER & RÜMPEL, 2011).

In this sense e-Science has, in addition to the technological aspects (virtualization of hardware and software, interoperability of digital tools, etc.), above all, a social and political science aspect (cooperative research, reusability of data, open access, etc.). For the scientist, new challenges arise with regard to a general, interdisciplinary, multiple nearby use of current digital information infrastructure. It requires more than just the professional scientific expertise (Fig. 2). Scientists must become familiar with search strategies within the network, develop advanced information processing skills, educate themselves on web technologies and service orientation in distributed systems and be aware of media skills required to share information with each other.

2. COMPONENTS AND EXAMPLES OF E-SCIENCE

2.1. Terms and components

In the context of the e-Science various concepts and components play an important role, a few of which will be presented here shortly.

(Primary) Research Data (PSD) is data that is generated during the research process, collected, used or put together by scientists. It is the basis to create models or theories and to proof scientific

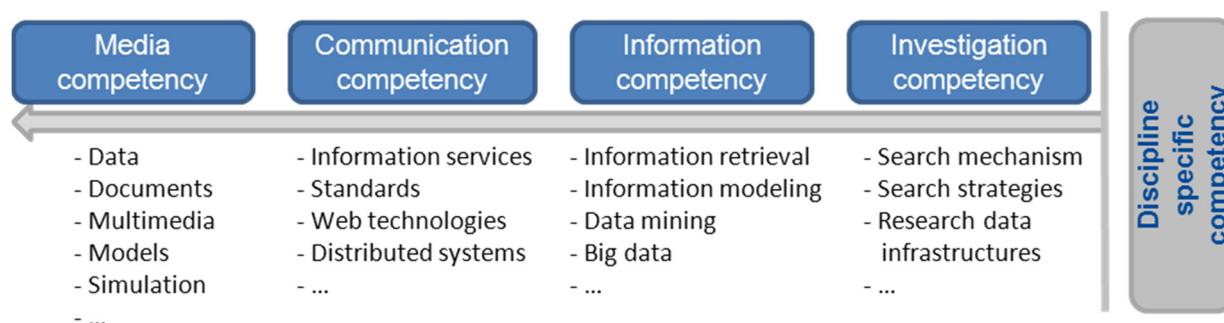


Figure 2: Enhanced competences needed in e-Science.

hypotheses. These include, in the broadest sense, primary data, secondary analyzes, visualizations, models, analysis tools, collections of objects or other products. (Primary) Research data provides a valuable repository of information that is collected with high financial cost. Depending on the subject area and method data is replicable or based on non-repeatable observations or measurements. In any case, the data collected should be publicly accessible and freely available after completion of the research. This is an essential prerequisite that data can be used again in the context of new research questions and that, in the event of doubt as to the publication, the data can be used for the review of the published results (DFG, 2006, 2008). Research data has a certain life cycle, that – from the view point of a researcher – covers the data production, the data usage, the publication and goes far beyond that in accordance with the data-driven science, namely to provide for and to reuse it in other scenarios (see for instance the data life cycle under data-archive.ac.uk/create-manage/life-cycle). Requirements are, among others, to guarantee the reliability of the data, to document the information processing chain and to describe the data sets with sufficient metadata.

For the provision and reuse of research data **Research Data Infrastructures (RDI)** have to be built as components of a research infrastructure (BMBF, 2013). Large research initiatives such as the “Initiative on Scientific Cyber Infrastructures” of the US National Science Foundation (NSF 2007) or the “European Roadmap for Research Infrastructures” (ESFRI 2008) accelerate the development of research data infrastructures for the exchange of scientific data. Research data infrastructures are combinations of infrastructure facilities that provide a service to the scientific community and store and disseminate research data for secondary research analysis, at little or no cost to the researchers. The data made available are mostly aligned to scientific questions (DFG, 2012 or www.allianzinitiative.de). The decisive factor is the interoperability of tools and research data as well as their long-term availability. They may be located at one single site (centralized) or distributed on several sites (distributed) and, by this, look like a virtual system (virtual). Mostly research data infrastructures arise at major research institutions or in the context of large research projects. The construction and operation of research data infrastructures at temporary organizational structures (such as university institutes, special research groups) is much more difficult. On the other hand, the establishment of structures for the permanent provision of quality-assured research data may become a key strategic objective for a larger research facility. Required properties of RDIs are the interoperability of its tools and its research data, their long-term availability and the possibility to reuse the data at different levels. The currently available research data infrastructures specialize in the provision in certain domains, such as the social network MyExperiment for the exchange of processes in the field of Bioinformatics or the iPlant Collaborative Cyber Infrastructure for Plant Science. In the field of Earth and Environmental Sciences, there are publishing platforms such as PANGAEA (Data Publisher for Earth and Environmental Science), Geoshare and the Global Collaboration Engine (GLOBE) supporting search and exchange of research data. The central features of these platforms are searching through catalogues, providing data access, performing visualization,

utilizing computing capabilities, as well as supporting the interaction and collaboration between scientists. Examples including spatial aspects will be called in the next section.

A **Repository** is a managed directory for storing secondary documents that should be made accessible for science and the public. Colloquially, this term is used for the storage of data and documents by means of modern information and communication technologies and making them sustainably available on the Internet. General examples are, for instance, the open source project Flexible Extensible Digital Object Repository Architecture Commons (FEDORA), developed at Cornell University, that is able to manage any kind of digital object (data, text files, metadata, images, videos, websites, etc.). At the Karlsruhe Institute of Technology (KIT), as part of a DFG-funded project, a Registry of Research Data Repositories named re3data was established (www.re3data.org/). Further examples of repositories are located at the World Data Centres (e.g. WDC-RSAT for remote sensing data or WDC-Climate for climate data).

In a **Research Data Archive**, however, data, documents, media etc. are managed for long-term storage. Data archives are indexed and provided with a search function for the retrieval and transfer of sections or the entire files. In many cases, formats used for archives differ from those used for repositories. The selection criteria must be comprehensible and the maintenance of the data over a very long period must be ensured. Some of the information stored in such data archives usually contains older data as a reference for future research. Archiving is a severe problem currently not solved in the digital world.

A **Virtual Research Environment (VRE)** “is an online system helping researchers collaborate. Features usually include collaboration support (web forums and wikis), document hosting, and some discipline-specific tools, such as data analysis, visualisation, or simulation management. In some instances, publication management, and teaching tools such as presentations and slides may be included. VREs have become important in fields where research is primarily carried out in teams which span institutions and even countries: the ability to easily share information and research results is valuable.” (en.wikipedia.org/wiki/Virtual_research_environment). It should enable collaborative research activities by several scientists working from different locations at the same time. VRE should have the potential to support the entire research process: from collecting, analyzing, discussing and refining the data to the publication of the results. Technologically it is based primarily on software services and communication networks and gives access to real research resources, such as data analysis and processing tools, scientific equipment etc. The scientists of the future attacks from any location on a virtual environment and finds programs, research data and secondary sources (such as publications, databases and services), which he needs for his current work. He stands in connection with his colleagues, directly accesses new findings, performs evaluations, adding his new findings and making them available "on the fly" for discussion and further use (see also NEUROTH ET AL. 2007, Definition AG „Virtuelle Forschungsumgebungen“ der Allianz-Initiative “Digitale Information” under www.allianzinitiative.de/de/handlungsfelder/virtuelle-forschungsumgebung.html).

2.2. RIsources-Examples with spatial reference

The German Research Foundation (DFG-Deutsche Forschungsgemeinschaft) operates the Internet portal RIsources (risources.dfg.de/). RI stands for Research Infrastructures and refers to dedicated devices offering scientific services and information infrastructures such as libraries, data repositories or scientific collections. RIsources offers concise information on national research infrastructures that provide the scientific resources and services for research projects. RIsources makes scientific research infrastructures accessible that offer recognized, established scientific and technological information, which allow free access or a transparent selection process based on scientific quality and feasibility of the project that have sustainable management and a long-term perspective. RIsources should help to utilize existing resources more effectively to facilitate scientists' access to required technologies and services and to develop technical information. Overlooking the providers of these resources, the

DFG hopes to improve visibility and – in the medium term – therefrom increased utilization of infrastructures often operated at very great expense. The portal also allows a general overview of research infrastructures at German academic institutions. Querying this portal for German research repositories and data archives in the earth sciences, there are a large number of hits, of which those should be mentioned here which cover space-related issues:

- Aktuelle Wetterwerte deutscher Stationen (WETTER)
- Animal Tracking Data (Movebank)
- Data Publisher for Earth & Environmental Science (PANGAEA)
- Deutscher Wetterdienst – Klimadaten (KlimaD)
- GLUES Geodateninfrastruktur (GLUES GDI)
- Historische hydrographische Daten des BSH (ICDC)
- Virtuelles Kulturlandschaftslaboratorium (VKLandLab)
- World Data Center for Climate (WDCC)
- World Data Center for Remote Sensing of the Atmosphere (WDC-RSAT)

3. GEOINFORMATICS AND E-SCIENCE

3.1. Spatial reference in e-Sciences – the role of geoinformatics

The spatial reference plays a central role in many fields of science, of course for geospatial disciplines such as earth science, planning and environmental sciences, but also a move to the space or the geographic area as a cultural dimension is reflected in the cultural and social sciences. Since the late 1980s, the term “spatial turn” became popular (DÖRING/THIELMANN, 2008). In the context of the increasing digitization of science (called Digital Humanities or eHumanities) a few years ago, the term “Spatial Humanities” or “GeoHumanities” occurred (BODENHAMER ET AL. 2010), explicitly integrating geospatial references and spatial information systems in human research work. The potential of geospatial methods is, for instance, highlighted in the text-based geographic analysis and visualization using cartographic methods. B. PIATTI (2009) speaks of the “literary geography”.

In the modern human sciences the spatial location of objects of research often is an important further dimension, this is by specifying the place name as an indirect spatial reference, by spatial descriptions or by direct positioning on maps as a direct spatial reference. For the humanities, the possibility of spatial location of their research subjects, spatial analysis and visualization (distribution pattern propagation maps) can be well coupled with a temporal dimension.

According to the IMAP functional chain of Geographic Information Systems (BILL, 2010) geoinformatics takes a primary role at different places in e-Science. In the most varied fields of science, data is recorded spatially (I=Input). By direct measurement in the field or through georeferencing remote sensing images a direct spatial reference is given, whereas positioning the data on the basis of place names creates indirect spatial references. In the data administration (M=Management) the spatial reference can be used as structuring and ordering systems, so as to promote fast and efficient search mechanisms on large amounts of data. With spatial data mining methods, the data of the most diverse areas can be spatially analyzed (A=Analysis), for example to identify spatio-temporal clusters. Presentation of results and disseminating the data via web technologies is then processed cartographically or by computer graphic visualization (P=Presentation).

With some selected examples, the role of geoinformatics in e-Science should be explained:

- Spatial references as ordering criteria beside semantics and time: Geographical names
- Spatial data repositories and virtual research environments: Virtuelles Kartenforum 2.0
- Spatial data research infrastructure including spatial functionalities for research: GLUES

3.2. Spatial references as ordering criteria: Geographical names

„Whatever occurs, occurs in space and time.“ (HILL, 2006). The spatial dimension is – in addition to the thematic and temporal dimension – a central concept of context description of information resources. To query, link and evaluate information resources, it is necessary to tap geographical references in a structured form. In the media of cultural tradition and science, spatial references are usually made visible textually using place names. Place names, e.g. toponyms or geographical names, designate names of topographic objects and can be mapped to the real world using geographical coordinates. Similar the search for relevant documents via a text-based search in titles, full-text, key words or classification systems is based on geographical names. The spatial reference here is relevant and is often established about gazetteers in catalogue stocks. Building such information resources provides an important research base for scientific disciplines that work in space and time (eg, historians, geographers, planners, ecologists, statisticians, demographers). However, they gain only sustainable value when they are georeferenced, i.e. clearly identified with geographical coordinates. Gazetteers in the classical sense are listings of toponyms or geographical names in a specific region in order to identify them based on additional spatial information and assign these to space. Current web-based services on place names exist with different functionality (eg geocoding / reverse geocoding via toponym, address, zip code, etc.) and in some cases specific applications (eg genealogical information, crowdsourcing, official certificates, etc.). For use in the context of virtual research environments and map portals, a number of geographical name-based service directories (eg GeoNames, OpenStreetMap Nominatim) have been established for a space-based research in the offered data sets. The place names contained in map sheet series (both toponyms in general and especially settlement names) and its locations can be used to link map information with other existing directories such as libraries.

In the last years Chair of Geodesy and Geoinformatics (GG) at Rostock University has been creating a geographical names directory especially for the area of the former Deutsches Reich (historical German Empire between 1871 and 1933). Therefore we checked and combined the following online gazetteers:

Table 1: Gazetteers and relevant geographical names

Name of the gazetteer (URL)	# of items
GeoNames (www.geonames.org/)	345.331
Genealogisches Ortsverzeichnis (gov.genealogy.net/search/index)	283.496
OpenstreetMap (Nominatim) (www.openstreetmap.org/)	218.913
Bundesamt für Kartographie und Geodäsie (www.geodatenzentrum.de/geodaten/)	45.061
GeoBeacon (HONV-MV, www.vklandlab.uni-rostock.de)	ca. 91.000
Aggregated and filtered number of geographical names	377.946

The directory of geographic names results in a collection of spatial objects, which are identified by their location. By keeping the foreign identifiers of all external data sources, it is also possible to recycle all combined information in an object based on their respective source. The result is a set of URLs and links with reference to the corresponding record in the website of the gazetteer provider. The method is described in WALTER & BILL (2015).

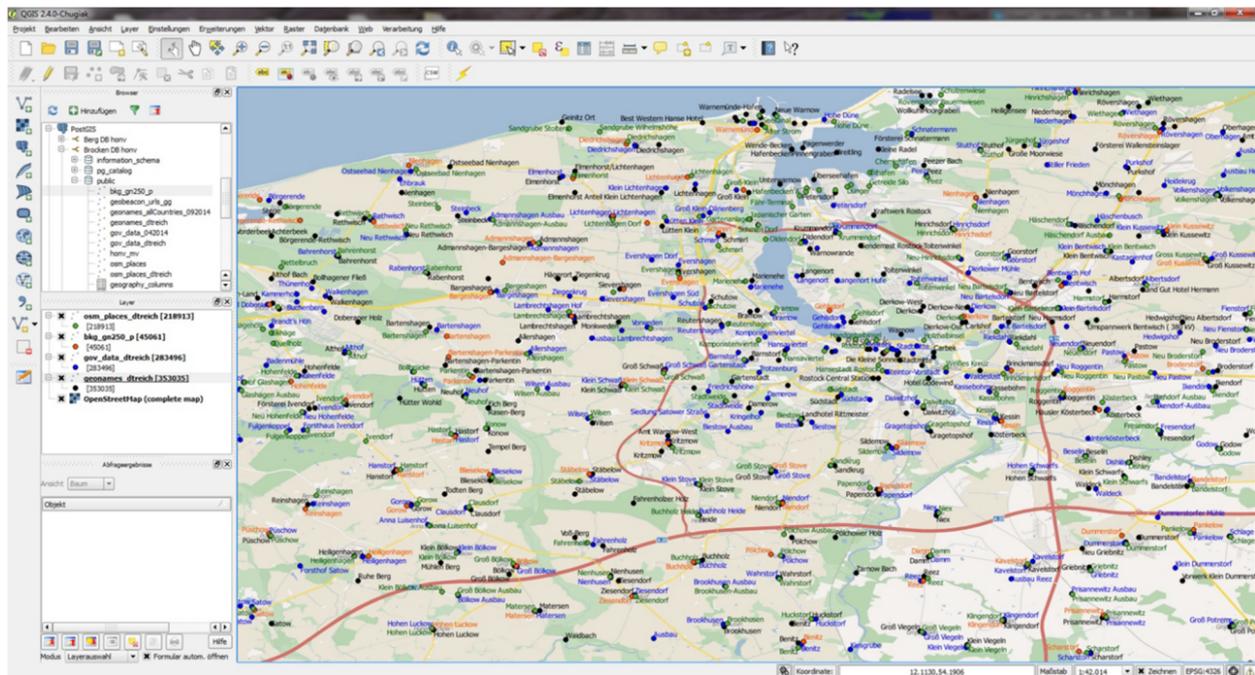


Figure 3: Directory of geographical names, aggregated from different gazetteers.

3.3. Spatial data repositories and virtual research environments: Virtuelles Kartenforum 2.0

The Chair of Geodesy and Geoinformatics (GG) at Rostock University has been involved for many years, inter alia, in DFG-funded projects, aiming at the development of web-based virtual research environments and research data infrastructures for cultural landscape research (BILL, 2012). Scientists (eg planners, historians, geographers) should have access to current spatial data as well as old georeferenced maps through OGC web services and portals.

In recent years, a comprehensive historical map series was created, first covering the area of the historical Mecklenburg (DFG project VKLandLab), now on the complete area of the former Deutsches Reich (DFG project Virtuelles Kartenforum 2.0). VKLandLab's time slots of old maps and geoinformation started with the maps of Wiebeking (of approximately 1786, 48 sheets at a scale of 1: 24,000) and Schmettau (by 1788, 16 sheets at a scale of 1: 50,000), the topographic map sheets of the Prussian land survey (approximately 1890, 168 sheet scale 1: 25,000) to the today's Topographic Maps (DTK 10) of the Authoritative Topographic-Cartographic Information System (ATKIS). This supports landscape researchers to analyze landscape changes for more than 230 years in a large area (15,000 km²) with modern information and communication technologies in combination with other data sources, for example publications and historical statistical data from libraries. The above mentioned historical place names directory (HONV) offers scientists both spatio-temporal and bibliographic work on their research questions. A demo access for everyone is available under <https://kvwmap.geoinformatik.uni-rostock.de/VKLandLab/index.php>.

Based on the results of VKLandLab together with Saxon State and University Library Dresden (SLUB) the project Virtuelles Kartenforum 2.0 creates an access to a collection of old maps covering the area of the former German Empire.

The SLUB inventory includes a comprehensive digital map series, among others a set of scanned topographic maps 1: 100,000 (674 sheets) and the topographic maps 1: 25,000 (about 6,000 sheets, formerly called Messtischblätter). For space-time researchers, the developed virtual platform provides an access to these map series for spatial analysis and visualization, and it enriches the librarianship to relevant ISO meta-information. Virtuelles Kartenforum 2.0 offers an access to the old georeferenced maps as one virtual resource using OGC-compliant WebMapService (WMS).

Operator of the platform (Fig. 4) is the SLUB. Currently nearly 6,000 georeferenced topographic maps of scale 1: 25,000 covering the time range from 1868 to about 1965 are available. The georeferencing of these maps has been done in two ways. The first uses a crowd sourcing approach in a web client, and the other utilizes an automatic image processing tool. The results of both approaches have been compared and showed for both cases an excellent geometric quality in the range of 10 to 20 m in reality.

3.4. Spatial data research infrastructure including spatial functionalities: GLUES

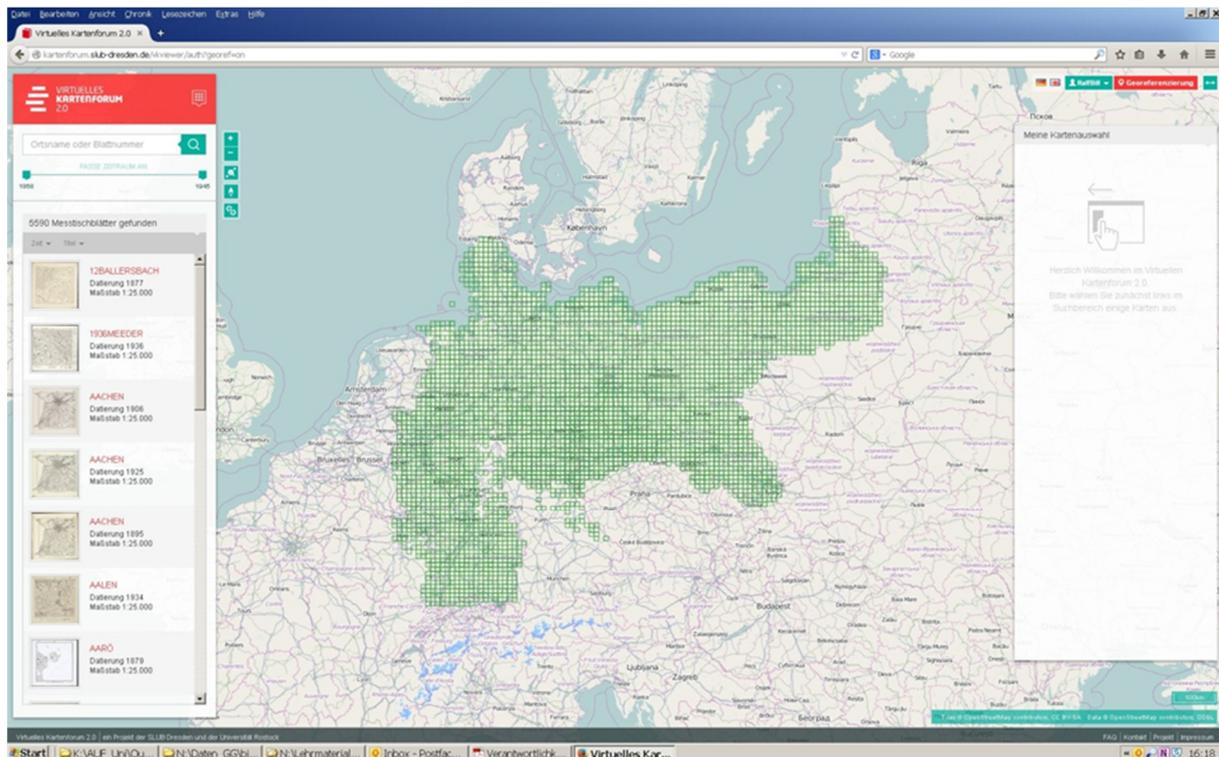


Figure 4: Virtual map forum (Virtuelles Kartenforum 2.0).

MÄS ET AL. (2014) are developing a scientific spatial data infrastructure supporting the data exchange within a land management and climate change research program of BMBF. The project „GLUES“ (Global Assessment of Land Use Dynamics on Greenhouse Gas Emissions and Ecosystem Services) supports international multidisciplinary collaboration in research projects through the communication, coordination and integration of the achieved results through a common data platform and through the development of standardized scenarios in land use, climate change and socio-economic changes. The GLUES Spatial Data Infrastructure (GLUES GDI) organizes the exchange of scientific data from models and simulations. As an exchange platform for the scientific data GLUES GDI has the following objectives:

- Publication and sharing of model data, analysis results and basic scenarios of the involved research groups.
- Seamless integration of existing data sources through the GDI, for example, for the calculation of scientific models or comparative analysis.
- Stakeholders of different areas are supported by spatial search and analysis tools to find research results and to understand and to use them for their own planning and management activities.

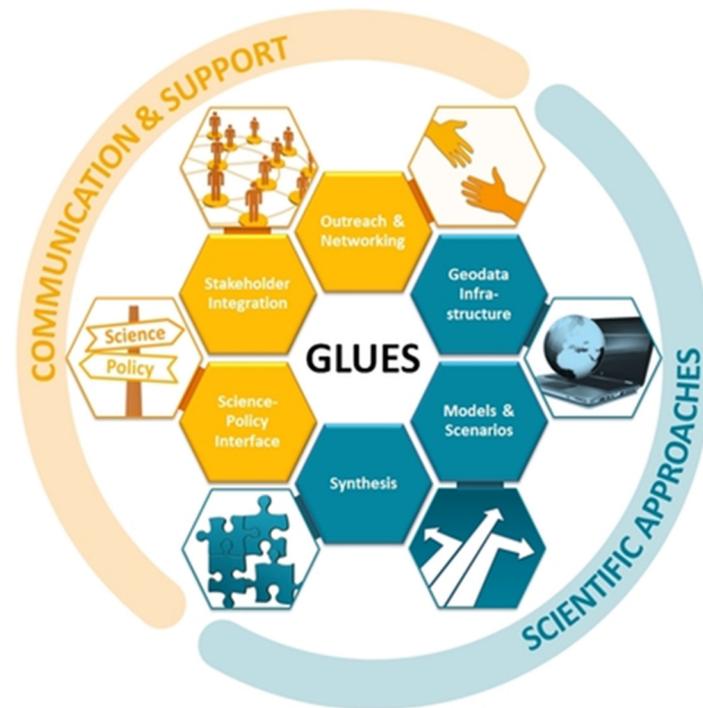


Figure 5: GLUES (MÁS ET AL. (2014)).

The GLUES catalogue as the central component for searching in the GDI includes nearly 3,000 items of metadata. This high number of records mainly comes from results of the simulation models. The models are usually implemented so that they provide a large number of result records with different thematic content in each pass. These data sets are each individually listed in the catalogue and linked by a hierarchical structure as a complete data set or data series. In order to understand the individual result records and their interrelationships, the links between the records are particularly important. Data are organized as decentralized data storage systems in the scientific institutions which produce the data, and the provision of data using decentralized services. Of the twelve regional projects metadata and data from six regional projects can be currently found in the GLUES GDI. Most of these projects operate their own geoportals or catalogues, and are integrated via the CSW interface in the search at the GLUES catalogue. Most of the data are disseminated via the GLUES GDI. Special requirements for the content and especially on the functionalities are provided by the scientific GLUES-GDI:

- Much of the data described and provided in the GLUES-GDI are different climate, land use and biodiversity simulations. These time series data have different temporal resolutions and are available for certain time slots. Tools for an accordingly flexible visualization of time-series data exist only sporadically in commercial or complex expert systems. For the visualization and analysis of time-series data in the GLUES-GDI a web application has been implemented, which both meets the requirements of users, as well as findings of the latest standards and technologies. The various groups of users of the GDI, such as stakeholders or partners of different areas of the program, can use the client as a software package in their own GDI and may adapt the client or use it parameterized with their data and services via the GLUES-GDI. Within GLUES a loose coupling of Web clients to different applications of the GLUES-GDI was required, for example, as a visualization component for the metadata catalog. To ensure these different scenarios, standards-based interfaces are used.
- The simulation models are usually run by scientist several times with different parameters and input data sets. For this reason, information about the history of the simulation data is a

particularly important part of the descriptive metadata. The ISO 19115 standard includes the element Lineage Description to describe the development processes and input data. In most applications of the ISO standard, however, this detailed modeling is dispensed with and the history is best displayed on a free text field. The metadata of the GLUES catalogue includes a detailed modeling of the genesis of the data, so that for each data set the respective simulation models, the input records and references can be written to the relevant scientific publications. Each input record also has its own set of metadata which is then linked to the history described in the catalogue. One advantage of this is that not only the history but also the reuse of a data set is recorded as an input into a simulation in the metadata.

- The integration of services for data processing in research infrastructures allows for interactive analysis of the existing data. The GLUES-GDI provides access to a wide range of climate simulations. In the GLUES-GDI this is provided in three different ways:
 1. as an ad hoc process executable on a WPS
 2. as an interactive client, on a predefined set of records
 3. as a well-defined, exchangeable code which can be downloaded from experts and embedded in their own systems and applications.

4. CONCLUSIONS

In the context of e-Science, interesting new fields of activity arise for geoinformatics. For many research questions the spatial reference is representing an ordering criterion. Also in many e-Science components such as research data repositories or virtual research environments geoinformation play an important role. First examples can be found at selected research institutions. In this contribution some examples and aspects have been presented illustrating the high potential as well as the interesting research and development issues for GI scientists.

5. ACKNOWLEDGEMENTS

The author would like to thank the Deutsche Forschungsgemeinschaft (DFG) for funding projects on Virtual Research Environments (funding reference Bi 467/21-2, Bi 467/21-2 and Bi 467/22-1).

6. REFERENCES

- BERNHARD, L., MÄS, S., MÜLLER, M., HENZEN, C., BRAUER, J. (2014): Scientific geodata infrastructures: challenges, approaches and directions. In: *International Journal of Digital Earth*. Volume 7, Issue 7, pp. 613-633 (<http://dx.doi.org/10.1080/17538947.2013.781244>).
- BILL, R. (2010): *Grundlagen der Geo-Informationssysteme*. Wichmann Verlag: Berlin. 806 p.
- BILL, R. (2012) [Hrsg.]: *Kulturlandschaftsforschung in einer virtuellen Forschungsumgebung auf Basis von Internet-GIS-Technologien*. Wichmann Verlag: Berlin. 186 p.
- BMBF (2013): *Roadmap für Forschungsinfrastrukturen*. Pilotprojekt des BMBF.
- BÜTTNER, S., RÜMPEL, S. (2011): Bibliotheken und Bibliothekare im Datenmanagement. In: S. Schomburg, C. Leggewie, H. Lobin, & C. Puschmann [Hrsg.]: *Konferenz: Digitale Wissenschaft. Stand der Entwicklung digital vernetzter Forschung in Deutschland*. Köln, Deutschland 20.-21. Sept. 2010. 2. Aufl. Köln, S.107-114.

- BÜTTNER, S., HOBOHM, H. C., MÜLLER, L. [Hrsg.] (2011): Handbuch Forschungsdatenmanagement. BOCK + HERCHEN Verlag: Bad Honnef.
- BODENHAMER, D. J., CORRIGAN, J., HARRIS, T. M. [Ed.] (2010): The Spatial Humanities – GIS and the Future of Humanities Scholarship. Indiana University Press: Bloomington.
- DEAR, M., J., KETCHUM, S. L., RICHARDSON, D. (2011): GeoHumanities: Art, History, Text at the Edge of Place. New York: Routledge.
- DÖRING, J., THIELMANN, T. [Hrsg.] (2008): Spatial Turn. Das Raumparadigma in den Kultur – und Sozialwissenschaften. Transcript Verlag: Bielefeld.
- DFG (2006): Wissenschaftliche Literaturversorgung und Informationssysteme. Schwerpunkte der Förderung bis 2015. DFG-Positionspapier.
- DFG (2008): Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten.
- DFG (2012): Positionspapier der Deutschen Forschungsgemeinschaft. Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme. Die digitale Transformation weiter gestalten – Der Beitrag der Deutschen Forschungsgemeinschaft zu einer innovativen Informationsinfrastruktur für die Forschung.
- ESFRI (2008): European roadmap for research infrastructures. European Strategy Forum on Research Infrastructures. ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri_roadmap_update_2008.pdf (Zugegriffen am 30.5.2014).
- GRAY, J. (2005): Scientific Data Management in the Coming Decade. Microsoft Research. Technical Report MSR-TR-2005-10.
- GRAY, J. (2009): 'eScience: A Transformed Scientific Method.' In: HEY, T., TANSLEY, S., TOLLE, K. [Ed.]: The fourth Paradigm: Data-Intensive Scientific Discovery, Redmond, Washington. : Microsoft Research. Page xvii-xxxii.
- HEY, T., TANSLEY, S., TOLLE, K. (2009): The Fourth Paradigm – Data-Intensive Scientific Discovery. Microsoft Research. Redmond, Washington.
- HILL, L. L. (2006): Georeferencing. The Geographic Associations of Information. Cambridge, London: The MIT Press.
- KUHN, T. S. (1970): The Structure of Scientific Revolutions (2nd Edition). University of Chicago Press.
- MÄÄ, S., HENZEN, C., MÜLLER, M., BERNARD, L. (2014): GLUES GDI – Eine Geodateninfrastruktur für wissenschaftliche Umweltdaten. In: *gis.Science* 27 (4), pp. 129-137.
- NEUROTH, H., ASCHENBRENNER, A., LOHMEIER, F. (2007); „e-Humanities – eine virtuelle Forschungsumgebung für die Geistes-, Kultur- und Sozialwissenschaften“; *BIBLIOTHEK Forschung und Praxis*, 31 (3), pp. 272-279, DOI: 10.1515/BFUP.2007.272, /December/2007.
- NSF (2007): Cyberinfrastructure vision for 21st century discovery. US National Science Foundation Cyberinfrastructure Council. <http://www.nsf.gov/pubs/2007/nsf0728/> (accessed 30.5.2014).

PIATTI, B. (2009): Die Geographie der Literatur. Schauplätze, Handlungsräume, Raumphantasien. Wallstein Verlag: Göttingen.

WALTER, K., BILL, R. (2015): Aufbau eines crowdsourcing-basierten Verzeichnisses für historische Ortsnamen. In: AGIT 2015. Wichmann Verlag: Berlin/Offenbach (to appear).

www.allianzinitiative.de (2010): Grundsätze zum Umgang mit Forschungsdaten.

www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/Tagung_Digitale_Wissenschaft.pdf .