InGeo-IC: The portal to Geodata

Dr. Uwe JASNOCH, Stefan GÖBEL, Dirk BALFANZ
Fraunhofer-IGD
Rundeturmstr. 6
64283 Darmstadt

Abstract. In Germany there exist innumerable geodata which, on the one hand, represent a great economic value but, on the other hand, lie idle to a great extent with (public and private) data suppliers so that they cannot be disposed of by potential geodata appliers. This is the reason for a platform or a turntable to be found connecting the groups belonging to the geodata market: geodata providers and appliers as well as GIS service providers. The information and cooperation forum for geodata (InGeoForum) has taken over to create such a platform as a project and make it available to all groups. Within the development of the InGeo-IC project methods and concepts have been prepared allowing to find, compare, and assess geodata by means of the metadata technique.

1. Introduction

The more information and data[1] are produced by the actual information society, the more important become mechanisms and systems which organize the data and include information where to find and access the wanted data. Most popular peculiarities of such information systems are metadata information systems (MIS) and catalogue systems (CS).

The main difference between these systems is the extent of the information spectrum covered. Whilst e.g. catalogues for department stores or phone books are typical representatives of CS covering a relatively clearly outlined information spectrum (address, phone number, article no., price etc.), other systems cover a quite larger spectrum. Examples for this are environmental MIS such as the UDK (German: Umweltdatenkatalog, engl.: environmental data catalogue, see [19]) on a national basis or GELOS (Global Environment Information Locator Service, [13,9]) and EIONET (European Environment Information and Observation Network, [8]) on an European basis.

This paper describes general aspects around the topic, outlines typical strengths and problems of existing systems and gives some impressions about trends and current activities in field of MIS for geospatial data.

[1] Between 80 and 85% of all data are geospatial data, e.g. maps, administrative areas, postal codes or addresses.

In the first part, some definitions and background information about the topic metadata and MIS are given. Then, different metadata tasks and views are listed: whilst from the user point of view, MIS are used as search engine in order to locate suitable data and to integrate it in current GIS projects, on the other hand geodata provider use it as marketing instrument – here MIS build the basis for E-Commerce applications. Afterwards, several examples of existing MIS and current activities and organization involved in the field of MIS for geospatial data are presented. Finally, aside from a comprehensive summary, some trends referring to further development and technical research in that field are mentioned.

2. The Information System

The present situation on the geospatial data market could be characterized as follows: On the one side, there are data suppliers who want to provide as much geospatial data as possible, on the other side, there are users who are insufficiently informed to benefit from the data. Typical lacks of information on the users' side are: which data are really needed, which data are available, how and where to get this data, which GIS systems are best suited for an application and how to integrate the data into a GIS application. The interests of the data suppliers are to advertise their data and to improve their presence on the geospatial market.

A solution for these demands are metadata information systems or catalogue systems like the InGeo-MIS which is part of the InGeo-IC project.

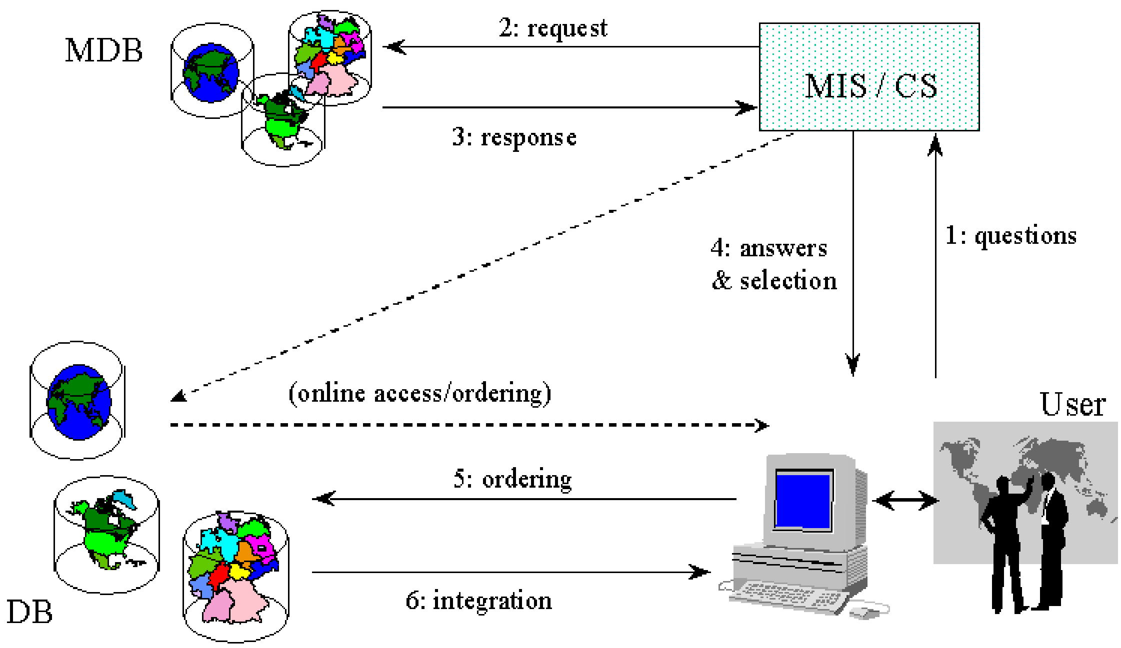

Figure 1: Scenario of a metadata information system

Figure 1 presents the scenario of a metadata information system and the different steps on the way from a metadata query to the access of the respective geospatial data. The general starting point for the usage of a metadata information system is a user looking for geodata in the context of a certain project. He needs information about existing geospatial data and where to get it. The user sends a query to a MIS, which translates his lingual query into a valid request. The MIS sends the request to one or more meta-databases, collects the results and prepares an adequate presentation/visualization.

The user compares these results and decides which one meets best his requirements. Finally, he may contact the data supplier to order the topical geospatial data (not metadata) and to integrate them into a GIS.

Another possible ordering process could be the online access to data servers containing both metadata and original datasets. In this case, the complete interaction (from metadata query to geodata retrieval) could be handled between user and MIS.

The main difference between the existing systems is the extent of the information spectrum covered by these systems. Whilst e.g. catalogues for department stores or phone directories/books are typical representatives of CS covering a relatively clearly outlined information spectrum (address, phone number, article no., price, etc.), other systems cover a quite larger spectrum.

3. History, Origin and Background

The concept of metadata has already existed a long time before the arising of the Internet/WWW and has been used in the 'library world' in form of catalogue cards and digital library information systems, which contain information about the contents, formal bibliographical aspects and the usage of books. Hence, the origin of metadata and protocols is settled in library information systems: In order to give users the possibility to search for data (books, reports, etc.) both in a local library and 'foreign' libraries resp. databases networked library information system have been established and the protocol Z39.50 [21] has been defined for the data exchange between the different participating databases integrated in the network. An example of such a distributed library information system represents the DBV-OSI II (German library of Congress, [6]).

Z39.50 is a comprehensive standard with high functionality (several services for searching and presenting result sets of requests), but otherwise it is very general. Therefore, applications often do not need to exploit complete functionality of Z39.50, but require some further specific additions relating to the thematic context. So-called profiles and attribute sets have been defined, providing harmonized terminology and encoding of terms referring to special applications like GILS (Governmental Information Locator System, [14,9]) and WAIS (Wide Area Information System, [20]), or special 'communities' such as musicians community, biology community or EO community (Earth Observation). There are numerous relationships between the data protocol Z39.50, several profiles (see [1,5,14,16,20,21]) and attribute sets (see [2,21]) playing an important role in the sector of geospatial data.

One of the first and most well- known metadata formats is the FGDC format, which is originated by the Executive Order 12906 [10], *Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure*' signed by Clinton on April 11, 1994 and has been approved by the Federal Geographic Data Committee [11] on June 8, 1994 to the metadata standard ‚*Content Standards for Digital Geospatial Metadata*‘ (abbr. FGDC format resp. standard).

Aside from the FGDC format, which has its origin on the national basis (USA) and is widespread across North America, according to metadata standards for geospatial data on international basis the activities/efforts of the two standardization organizations ISO/TC 211 Geographic information/Geomatics (International Organization for Standardization, worldwide, [15]) and CEN/TC 287 (European Committee for Standardization, European wide, [3]), as well as the OGC (Open GIS Consortium, [17]), which primarily pass/drop the results of their endeavor into ISO/TC 211, have to be mentioned and analyzed. In the context of environmental systems in Germany and Europe, the most wide-spread metadata formats are the UDK and the Catalogue of Data Sources (CDS, [9]).

4. Metadata tasks and views

Figure 2 characterizes the current situation on the geospatial data market: On the one side, there are data suppliers who want to provide as much geospatial data as possible, on the other side, there are users who are insufficiently informed to benefit from the data. Typical lacks of information on the users' side are: which data are really needed, which data are available, how and where to get this data, which GIS systems are best suited for an application and how to integrate the data into a GIS application. The interests of the data suppliers are to advertise their data and to improve their presence on the geospatial market. Corresponding to these different user groups and needs there are different views and tasks of metadata.

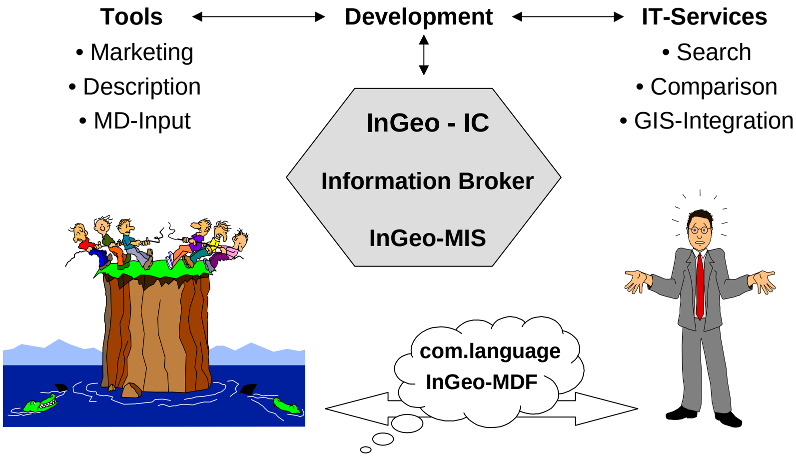

Figure 2: Current situation on the geospatial data market

4.1 Metadata: User point of view

From the end user's point of view metadata and MIS are instruments and mechanisms to locate required data, to compare various different available/suitable data and finally to integrate the data into special GIS applications or projects. Apart from comprehensive, reliable and expressive information, easy and intuitive graphical user interfaces (GUIs) are most important in the information retrieval process.

Especially in the context of geospatial data, spatial metaphors such as maps help users both to formulate appropriate geographical searches and to understand data characteristics (extent information). With regard to a comparison between several appropriate data sets, preview images and symbols, keywords and further metadata information need to be presented, in order to get an impression which data best fits user requirements. Often MIS or any other kind of information system with search instruments –for instance web search engines such as AltaVista or Yahoo- offer a sorted list of appropriate information with best matches at the top, but it's quite hard to understand the list, because the list contains thousands of matches (which results in the 'lost in hyperspace' phenomena) and the ranking mechanisms are not transparent to the user. The latter aspect is very important in the context of MIS for geospatial data: Users want to know which parameters of the formulated query impinges on the result in which way and subsequently how to modify the query.

Several workshops on metadata for geospatial data and comprehensive studies on user behavior in MIS have shown that the most important search parameters are the spatial and temporal extent, keywords and categorical information on geospatial data and application areas (e.g. GCMD valid [12], CEO discipline keywords [4] or ISO theme codes [15] build categories such as cadastral, topography, geodetic control, remote sensing, earth science, atmosphere, environment or forest & natural vegetation). All existing MIS offer some kind of these search mechanisms, but there are differences in usability and user support. For example some MIS offer thesauri (defined set of hierarchical and semantically related terms) and gazetteers (thesauri for geographic names and places, e.g. administrative areas) in order to improve query formulation and information retrieval in general. Another point are enormous differences relating to result presentation: Whilst most MIS are limited to result lists with the titles of matching datasets, other systems offer graphical support referring to navigation in metadata structures or getting impression of datasets by visualizing the spatial extent on maps or using preview images.

4.2 Metadata: Provider point of view

From the point of view of geodata provider metadata represent an instrument for marketing and web-based MIS build the basis for E-Commerce and geodata online. In the last decade, parallel to the growth of the internet (studies estimated the growth of the internet connections in Germany at 30% up to 2002 and expect an amount of 66,6 million Internet and online user in western Europe) there is an enormous trend referring to geodata marketing. Here, geo information builds the basis for new business areas and technologies, such as telematics, environmental monitoring, environmental forecasts, telecommunication or knowledge discovery.

More and more people –especially on the private sector- notice that fact and see the value of geospatial data as well as its corresponding market potentials.

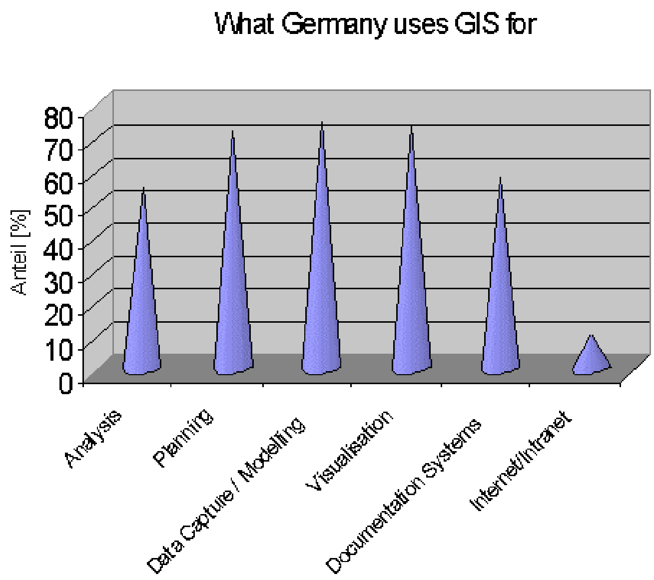

Figure 3: GIS in the internet and intranet in 1998, source: Smallworld Systems

The growth of the geo market in western Europe based on an analysis by the DDGI [7]. Here the increase rate of the market for digital geo information is estimated at 140 % in the time period between 1996 and 2001. Another estimation says that the volume of the German market for geo information has reached 220 million DM in 1999 and the annually increase lies between 10 and 30 % which results in a value of about 300-560 million DM until 2001. Corresponding to that expansion the competition between data provider (including both public and private sector) increases, too. Therefore, all marketing concepts and instruments are used in order to improve presence on the market. This tendency is catalyzed by the propagation of web-based GIS systems, which are settled in different application areas (see figure 3) and make enormous use both of the internet and digital geo information.

Of course, there are always problems with these kinds of forecasts, but one can conclude that there is a strong relationship between the figures of GIS and the geospatial data they consume.

The open question for geospatial data is always their value in terms of price for datasets. Per definition, the price of a product is directly linked to the value of the information for the potential customer. Consequently, if there is no customer for geospatial data, these data sets do not have any cash value.

Obviously, this last scenario does not meet reality and there are customers for geospatial data. But the initial statement should be considered in order to conclude, that a fixed sales structure for the product "geospatial data" does not meet the needs of the market. In order to visualize these elements, a following table shows some examples demonstrating the relation of information, the added value of this information and a fictive price for a possible product with geospatial data.

Application	Hotel	Urban-planning	Real Estate
Value of Decision (€)	500	High	100000
Added Value by geospatial data (€)	15		
Possible price of the product (€)	7,5	100	1
# of decisions per year (Mio.)	4	0.3	160

Table 1: Value-adding and prices of geo information

As one can conclude, there is a high diversity in price and value. In addition, the resulting price politics depends also from the usage of the product, thus the number of possible decisions based on the product. The stated figures are not based upon a market analysis and may vary from reality. Nevertheless, they visualize the different factors of influence and provide somehow the basis for the hypotheses stated before, that a fixed price politics may not meet the demands of the market.

There is one more week point within this hypothesis, it somehow implies that the geospatial data could be found and transferred easily to the customer. As a matter of fact, finding the appropriate data is one of the highest obstacles in the geospatial data market. The technology introduced here in the chapter, the catalogue systems and the meta information systems are possible ways to overcome this barrier.

On the other hand there is still the disadvantage of costs and endeavor involved in the process of describing geospatial data respectively generating meta information. Therefore, sometimes provider resign to generate metadata and there exist some geospatial data – anywhere in archives, books, files, etc.- but that data is neither documented nor locatable. Especially this situation arises in countries such as Germany where (apart from the environmental sector, see [18]) there are no laws or executive orders, which instruct Federal agencies to document its geospatial data with metadata and to provide these metadata to the public.

5. Summary and Conclusion

The more information and data are produced in the actual information society, the more important become mechanisms and systems which organize the data and include information where to find which data. Especially in the field of geospatial and environmental data, data provider notice the enormous value of its data and the potentials offering web-based information systems and geo data online. Subsequently, more and more digital libraries, catalogue and metadata information systems arise. Further on, several activities, organizations (e.g. InGeoForum or IMAGI in Germany) and technical boards such as CEN, ISO or OGC have been established on national, European or worldwide basis in order to enhance the geospatial market and to coordinate the establishment of MIS using harmonized metadata formats and data exchange mechanisms.

On technical point of view, recent initiatives to geospatial libraries and MIS/CS provide access to a wealth of distributed data, but offer only basic levels of interactivity and user assistance. This includes all steps of the information retrieval process: Query formulation, query modification, comparison of (metadata) result sets and detailed presentation resp. visualization of result sets. Therefore further research on MIS for geospatial data probably will focus on usability and ergonomics aspects as well as knowledge discovery which helps users not only to locate and compare but also to understand geospatial data.

References

[1] C. Best, L. Hayes, D. Nebert, L. Reich, G. Percival, J. Hylton, S. Smith. Alignment of CIP and GEO Profiles. *Proc. EOGEO´98*, Salzburg, Austria, 1998.

[2] Bib1: Attribute Set bib-1 (Z39.50-1995): Semantics, ftp://ftp.loc.gov/pub/z3950/defs/bib1.txt

[3] CEN/TC 287: Geographic Information, http://forum.afnor.fr/

[4] CEO: Centre for Earth Observation, CEO Programme, http://www.ceo.org

[5] CIP: Catalogue Interoperability Protocol, Protocol Task Team (CEOS-PTT), Committee on Earth Observation Satellites, http://harp.gsfc.nasa.gov/ptt

[6] DBV-OSI II: Deutscher Bibliothekenverbund (German Library of Congress), - Open System Interconnections, http://www.ddb.de/partner/dbv-osi_II.htm

[7] DDGI: Deutscher Dachverband für Geoinformation, http://www.ddgi.de/ -> english version.

[8] EIONET: European Environment Information and Observation Network, http://www.eionet.eu.int/

[9] ETC/CDS: European Topic Centre on Catalogue of Data Sources, incl. document *'Comparing the ETC/CDS Core Data Model to GILS, GELOS and DUBLIN CORE'*, http://www.mu.niedersachsen.de/cds/start_documents.html

[10] Executive Order 12906: http://www.fgdc.gov/publications/documents/geninfo/execord.html

[11] FGDC: Federal Geographic Data Committee, ftp://fgdc.er.usgs.gov/

[12] GCMD: Global Change Master Directory, http://gcmd.nasa.gov/, contains EOSDIS.

[13] GELOS: Global Environment Locator System, http://ceo.gelos.org

[14] GILS: Application Profile for the Governmental Information Locator Service: Technical Topics and Other Information, http://www.usgs.gov/gils/topics.html

[15] ISO. Draft International Standard on Metadata for Geographic Information, ISO 19115-Metadata (status: 2nd CD, first notation was ISO 15046-15), see http://www.statkart.no/isotc211/.

[16] D. Nebert, GEO Profile, Z39.50 Application Profile for Geospatial Metadata, Version 2.1 b Initial Review Draft, http://130.11.52.178/GEO21a.html

[17] OGC: Open GIS Consortium, http://www.opengis.org.

[18] W. Schober, H. Lopatta. Umweltinformationsgesetz. Verlag Österreich, Wien 1994.
EU-Directive (90/31/EC): Freier Zugang zu den von Behörden erhobenen oder verwalteten Umweltinformationen.

[19] W. Swoboda, F. Kruse, R. Nikolai, W. Kazakos, D. Nyhuis, H.Rousselle. The UDK Approach: the 4th Generation of an Environmental Data Catalogue Introduced in Austria and Germany. *Proc. IEEE Metadata '99*. Washington, D.C., 1999. Koordinierungsstelle UDK: Umweltministerium Niedersachsen, Hannover, http://www.mu.niedersachsen.de/udk

[20] WAIS: Wide Area Information Servers, ftp://ftp.loc.gov/pub/z3950/profiles/wais-profile.txt

[21] Z39.50: ANSI/NISO Z39.50-1995 Information Retrieval (Z39.50): Application Service Definition and Protocol Specification, Z39.50 Maintenance Agency Official Text for Z39.50-1995, http://lcweb.loc.gov/z3950/agency/