Metadata Management with ISO 19115 -
The key to success of spatial data infrastructures (SDI)

Wassilios KAZAKOS

Abstract

Spatial data infrastructures are among the much discussed subjects in the field of geodata management. At the forefront is the issue of providing and marketing geodata. The reasons are firstly the intention to lower the cost pressure on public geodata providers and secondly the anticipation of new markets by private-sector providers. We believe that in spite of a certain recent boom the role of metadata is still not sufficient discussed, although metadata are the essential basis and precondition of a functioning geodata market. Metadata are needed to systematically document geodata and are crucial for their integration in marketing processes and workflows. Actually without well documented metadata no SDI will ever work. In addition, with the availability of large metadata sets new methods can be applied and developed to measure quality of available geodata and new research can be conducted for advanced handling of geodata within everyday workflows and processes. With this paper we contribute to the current discussions about SDI and put the focus on Metadata Information Systems (MIS) within a SDI which we are convicted no SDI can do without.

1. Introduction

Increasing cost pressure and the necessity to provide more geodata for routine tasks in public authorities require the shared usage of existing geodata. To allow this, the development of spatial data infrastructures (SDI) is presently at the top of the agenda in many authorities. A reduction of costs and a better marketing in a larger scale can be achieved only by a systematic and structured description of the metadata which makes them researchable for others. For this purpose standardized metadata are essential. They improve the self-organization within an authority and at the same time provide information about the available geodata for other authorities. Serving to describe metadata is the metadata standard ISO 19115 implemented in several national, regional and thematical profiles.

In addition to public authorities a private-sector market for geodata is emerging. Even when this market is not as rigorously coupled with the international standardization efforts, it meets the same challenges. The initial point is the realization that spatial data infrastructures used in their own do not primarily provide a benefit but are the precondition for the developing geo-information market. SDI aims at offering a direct access to and the usage of geo-information and geo-services of miscellaneous providers.

The discussion often neglects the fact of metadata being the hub of an efficient SDI. Metadata are frequently seen as a spin-off product of geodata which in our view is a mistake. For only a systematic documentation of geodata by metadata allows informing interested parties about offerings. Carrying the much-quoted example to extremes: Quite as no library could function without metadata, no geodata infrastructure could reach their aim without high-quality and meaningful metadata.

2. The Role of a Metadata Information System for Geodata

2.1 Metadata for Documentation and Information Purposes

The precondition for the utilization of geodata is an IT infrastructure which allows their collection, investigation, and management. This is where Meta Information Systems (MIS) on the basis of ISO 19115 come in, where – like in a library – available geodata are structurally described and managed. The ultimate goal of SDIs in public authorities must be to enable the shared usage of geodata. Seeing the enormous costs of generating and updating geodata, even selective measures can positively influence the budgets, because the costs of geodata collection amount to only a fraction of the costs of geodata generation.

Vice versa, the absence of an appropriate metadata infrastructure can lead to additional costs: For follow-up projects data must often be collected over again which increases the costs. Poorly documented data, at the latest loose their value at staff changes, because their significance can hardly be assessed any more. In the worst case, consequential costs emerge due to decisions based on incorrect data.

Thus, the goal of a MIS is to purposefully provide information about character, availability, and quality of the geodata. In order to enable this, the metadata of geodata are stored in a database and are usually provided via a web interface. To support the re-use of geodata, they ought to be described in a way for potential users to see at a glance whether they are applicable for them.

But metadata descriptions can contribute even more: In addition to the systematic description of geodata, all information required for their marketing ought to be stored. Apart from the attributes defined in ISO 19115 more standards to describe marketing information should be included. This possibility is for example offered by XML Complex Pricing Format (XCPF) which supplements metadata with machine understandable pricing information. By using XCPF a future integration into a broker infrastructure for example via a Web Pricing & Ordering Service (WPOS) is also easily possible.

2.2 Support during Data Generation

MIS not only enhance the overview over an organization's available geo products but they can also be profitably utilized in the process of data generation. When metadata are collected during every stage of generating and updating geodata, selective queries concerning current campaigns and their status can be answered promptly. Thus, data collections performed in parallel can be coordinated. Many organizations experience the lack of an overview over currently available geodata, and they do not know which campaigns are running or planned in which departments.

Frequently, simply the absence of knowledge or coordination causes the multiple collection of the same geodata. Having a comprehensive overview over running activities, their status and their subject helps, both in identifying the potential for re-use in time and in synchronizing campaigns. But this is only possible when the metadata are maintained in different stages of their generation and update.

2.3 Simplified Information Exchange

Besides designing local MIS the information exchange between different organizations is also growing more important. Via so-called broker architectures a distributed access to different sources is realized. So, instead of regularly mirroring data it is sufficient to tie the MIS to a broker architecture via standardized interfaces. In the meantime, such architectures are an integral component of a geodata infrastructure.

Even though different realizations are possible, the Catalog Service Web interface (CS-W) proposed by the Open Geospatial Consortium (OGC) is establishing, via which the access to the connected MIS is realized. Apart from the central GeoMIS.Bund used in Germany, there are a lot of other infrastructures thematically or regionally defining the access to MIS. For environmental issues there is for example the "Umweltdatenkatalog" (UDK; catalog of environmental data) documenting both geodata and the most important datasets with environmental reference. Another example is the "Nordsee-Ostsee-Küsteninformationssystem" (NOKIS; North Sea and Baltic Sea Coastal Information System), which can be used to research both geodata and time series or projects referring to German coastal information.

In order to keep open the option to offer one's metadata and thus one's geodata via various portals, it is therefore essential to bank on the standard interfaces.

3. Tool Support

3.1 Standard Functionality

In the meantime, technology is so far advanced that setting up a MIS by using standardized tools drastically lowers the costs of a new development. Whilst only ten years ago each MIS was practically newly developed, today more and more standard tools and methods establish which simplify and support the set-up of a SDI with a MIS. Not only the implementation periods are getting shorter and shorter, but the productization also causes better and better MIS with regard to performance, user guidance, and functionality. An important contribution is also given by the growing standardization in this area, and in the meantime, with the ISO 19115 standard for the description of geometadata there is a widely accepted international standard.

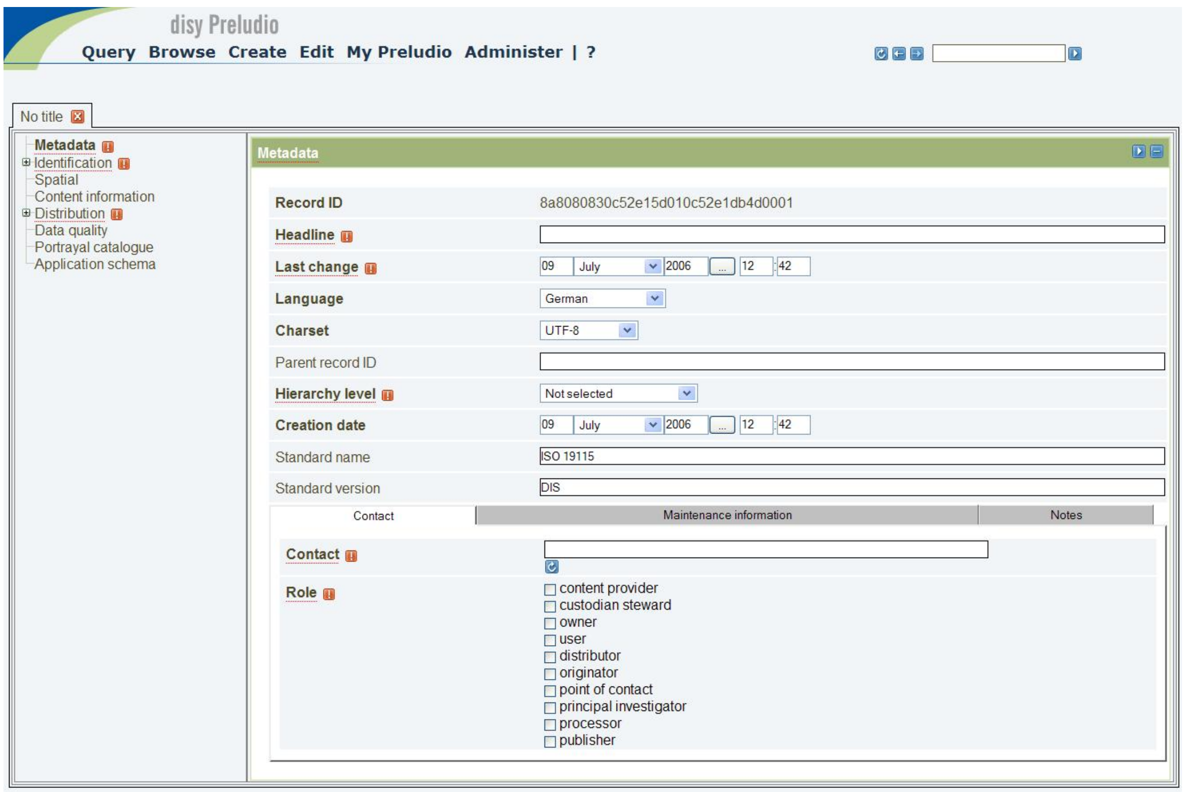

Figure 1: Example of an input mask for data collection in a MIS with disy Preludio

While previously a good deal of the discussions in projects revolved around how many and what type of attributes to describe, modern SDI tools can focus on this standard and offer improvements in user guidance or data management. The acceptance of MIS by the users can be achieved only like this. In the meantime, topics like the intuitive collection of the comprehensive ISO 19115 metadata, a simple and extendable search over space, time and theme, a clear management in a central database, and various import and export functionalities are common place, and the current trends regard the areas of the integration of MIS in the daily workflows, the connection of services (web services) and the stronger integration in Non-Geographic-Applications and Geographical Information Systems (GIS).
From a certain amount of descriptive information it is essential for the user to be directly supported in collecting the metadata. Apart from functionalities like context-sensitive help which accurately tells the user what to enter where, it

is getting more and more important to give him directions to guide him through the metadata collection. Modern tools could for example validate the metadata while they are collected and show the user exactly where to complete or correct data or which rule his entry is breaking. When employing measures like this not only the user is supported in entering data, but also the quality of the metadata is enhanced which in the end is deciding for success or failure of geodata marketing.

3.2 Configuration instead of New Development

In the meantime it is clear that the support of standards is not sufficient anymore. The detailed description of geodata by metadata, even with the purpose of internal documentation, again and again demands adjustments and complements of the profile responding to the specific requirements of an organization. Apart from differences in the used data base system or in the layout it is growing more and more important to react fast to new requirements for the profile. Generally the standard permits conformal extensions. The question remains how such profile extensions can be added to the application and consequently to the collection and search masks in a simple way but without leading to essential changes in the core of the application. In NOKIS for example these are coast-specific elements like tidal information which are not needed in other areas. The standard delivers the framework, but generally has to be expanded in so-called profiles. This does not affect the metadata exchange, for local expansions need not be passed on.

In order to meet this goal we pursue an approach where parts of the application are generated from the given XML schema. This technology comes from research at the Forschungszentrum Informatik (FZI; Research Center for Information Technologies) and was successfully transferred to and enhanced by disy. Among others HyperJAXB is used, an open source software developed by us and distributed via the sun website. HyperJAXB supports the mapping of XML to relational database schemes. Through this it is possible to adjust a MIS to new requirements relatively quickly without the need for a new development.

4. Example: NOKIS

In this context the project NOKIS is leading the way. In the project an information system for the coasts of North Sea and Baltic Sea is implemented. With NOKIS a standardized documentation of data and information from coastal regions was set up. Currently the information basis is extended to add standardized, web-based software tools to the functionality of the NOKIS portal www.nokis.org. Doing this, requirements of the user community are taken on which where phrased during the operating time of NOKIS. The conceptual and technological extension of NOKIS comprises the dynamic linking of further metadata profiles and the implementation of web-based methods in the existing NOKIS web portal.

Thus both the collaboration of Coastal Protection, Water Management, Environmental Protection, and Waterway Maintenance and the information exchange between research establishments can be intensified, and the required public relations can be supported. The necessary information infrastructure is largely based on standardization of data exchange and data handling. Existing structures, systems, and resources in the respective departments are to be systematically employed, reasonably complemented and enhanced. With web services controlled by metadata the flow of information in the context of multidisciplinary collaboration is to be optimized.

The benefit of this shared NOKIS portal lies in overcoming the sectoral view of specialized information systems and in the possibility to research data in their context by using an interdisciplinary virtual metadata base representing the most important data providers and information sources with regard to the coastal areas. The portal's range of functions comprises practical methods aiming at effectively supporting the departments in performing their tasks.

Based on the technologies of disy Preludio, NOKIS has become an important building block of the metadata infrastructure of most organizations in the German coastal regions. Via the OGC CS-W interface both a distributed access is realized and NOKIS is tied to the GDI-DE (Geoportal.Bund). In the same way data can be exchanged with the Umweltdatenkatalog (UDK). A significant point is the fact that due to consequent communication with the users the acceptance of the software, compared with other projects, could be increased so with about 2000 metadatasets NOKIS became one of the larger metadata sources in GeoMIS.Bund in best time.

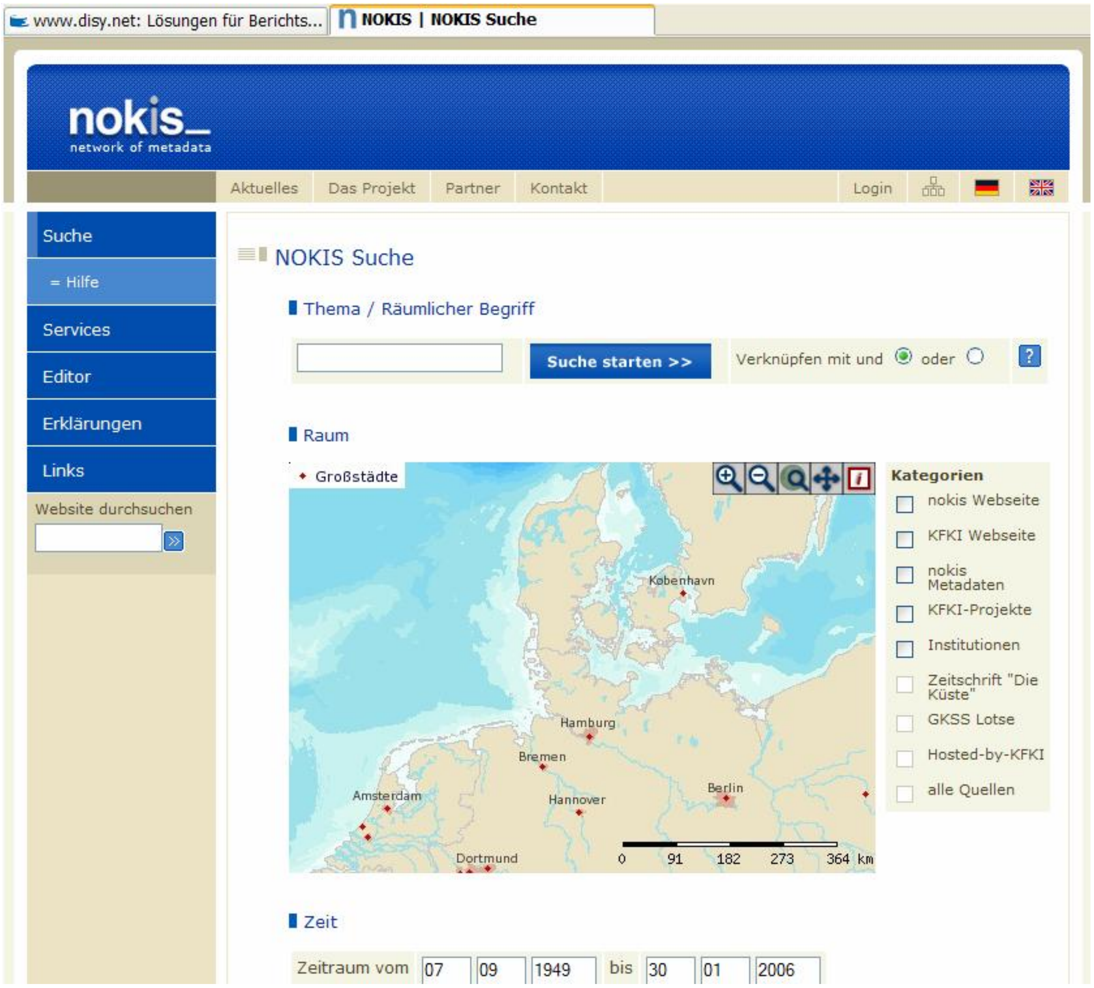

Figure 2: The NOKIS portal

5. Conclusion

Geodata infrastructures promise a simplified access to geodata. But still a crucial aspect is neglected in our point of view: For a working SDI and thus for a working geodata market metadata are the basis. They are not to be simply extracted from geodata as a spin-off product but must be processed systematically in order for them to satisfy their real purpose of documenting the data. Furthermore, metadata are the precondition for advanced concepts of a SDI like marketing or integration of workflows for daily tasks including geodata.

Currently another aspect is getting to the top of the agenda: Additionally to the description of geodata it is getting more and more important to define geodata for further datatypes such as time series and projects. Naturally they cannot be entirely separated from geodata. Thus it must for example be possible to assign specific geodata to a project and to document their processing state.

6. Bibliography

Kazakos, W., Lehfeldt, R., Michl, C., Heidmann C., Valikov, A. (2003), Metadata Repositories in Governmental Information Systems for Coastal Regions Proceeding on The Information Society and Enlargement of the European Union, 17th International Conference Informatics for Environmental Protection, Cottbus 2003

Technical Comitte ISO/TC 211: Geographic Information – Metadata – ISO 19115:2003 Final. – http://www.iso.org

Lehfeldt, R., Heidmann, C., Piasecki, M. (2002a), Metadata in Coastal Information Systems, in Holz, KP, Kawahara, M, Wang, SY (eds) Advances in Hydro-Science and -Engineering Volume 5. Proceedings of the 5th International Conference on Hydro-Science and -Engineering, Warsaw. Abstract Volume p181.

Lehfeldt, R., Sellerhoff, F., Piasecki, M. (2002b), Components of Web Portals in Coastal Engineering, in RA Falconer, B Lin, EL Harris, CAME Wilson, ID Cluckie, D Han, JP Davis, S Heslop (eds) Hydroinformatics 2002. Proc. 5th Intl. Conf., Cardiff. IWA Publishing, London, pp 1501-1506.

Kazakos, W., Valikov, A., Schmidt, A. (2002), Automation of Metadata Repository Development with XML Schema. In Werner Pilmann and Klaus Tochtermann (editors), Proceeding of the 16th International Conference Informatics for Environmental Protection - Environmental Communication in the Information Society., volume I, pages 400 - 407, Vienna, Austria

Kazakos, W., Kramer, R., Schmidt, A. (2000): Coastbase - The Virtual European Coastal and Marine Data Warehouse. In Armin Cremers and Klaus Greve, editors, Computer Science for Environmental Protection '00. Environmental Information for Planning, Politics and the Public, volume II, pages 646-654.

Swoboda, W., Kruse, F., Nikolai, R., Kazakos, W., Nyhuis, D., Rousselle, H. (1999): The UDK Approach: the 4th Generation of the Environmental Data Catalogue for Austrian and German Public Authorities, Proc. IEEE Meta-Data’99, http://computer.org/proceedings/meta/1999/papers/45/wswoboda.html