A MULTIFUNCTION WEB BASED METADATA TOOL DESIGN AND IMPLEMENTATION FOR NSDI

O. Emem, F. Batuk,

Yildiz Technical University, Department of Geodesy and Photogrammetrical Engineering, Besiktas Istanbul, Turkey
(oemem, batuk) @yildiz.edu.tr

KEY WORDS: Metadata, Standards, Web-based, GIS, Infrastructure

ABSTRACT:

The growing use of spatially referenced digital data has necessitated the development of a statewide metadata standard. Metadata is a term that refers to data about data. Geospatial metadata provides information about data content, quality, spatial reference, etc. A metadata standard will help facilitate data capture, translation, exchange, and documentation. The consistency in metadata content and style is recommended to ensure that comparisons can be made quickly by data users as to the suitability of data from different sources. Without standardization, meaningful comparisons are more difficult to derive without reading and learning many metadata management styles. Also metadata standards are important when browsing and finding required data from NSDI. Therefore, collecting, storing, presenting has become as important as the metadata standards. In this study, design and implementation of a multifunction web based metadata tool which able users to create, edit, view standard metadata records in one tool is presented. Also, design of national draft metadata standard which conform to ISO 19115 is briefly discussed.

1. INTRODUCTION

Geographic information is vital to make decisions at the local, regional, and global levels. However, information is an expensive resource, and for this reason appropriate information and the resources to fully utilize this information may not always be readily available, particularly in the developing world. Many national, regional, and international programs and projects are working to improve access to available spatial data, promote its reuse, and ensure that additional investment in spatial information collection and management results in an ever-growing, readily available and useable pool of spatial information.

Once created, geospatial data can be used by multiple software systems for different purposes. Given the dynamic nature of geospatial data in a networked environment such as NSDI (National Spatial Data Infrastructure), metadata is therefore an essential requirement for locating and evaluating available data.

The term metadata has become widely used over the past 15 years, and has become particularly common with the popularity of the World Wide Web. But the underlying concepts have been in use for as long as collections of information have been organized. Library catalogues represent an established variety of metadata that has served for decades as collection management and resource discovery tools. The concept of metadata is also familiar to most people who deal with spatial issues (GSDI, 2004).

Metadata is first element of NSDI – which enables a user to find spatial data that is available in the different NSDI servers. Metadata serves two major purposes – both for the spatial data producer and for the spatial data user. For the producer, the metadata provides a framework to document the spatial data and declare its content for users. For the user, Metadata serves many important purposes, including finding the spatial data of his need; browsing spatial data; deciding on whether the spatial data will meet the application need and finding how the spatial data can be accessed (ISRO, 2003).

Metadata also helps people who use geospatial data determine how best to use it. As personnel change in an organization, undocumented data may lose their value. Later workers may have little understanding of the contents and uses for a digital database and may find they can't trust results generated from these data. Lack of knowledge about other organizations' data can lead to duplication of effort. Metadata is one of those terms that is conveniently ignored or avoided. They often begin to look at incorporating metadata collection within the data management process (GSDI, 2004).

The growing use of spatially referenced digital data has necessitated the development of a state-wide metadata standard. The consistency in metadata content and style is recommended to ensure that comparisons can be made quickly by data users as to the suitability of data from different sources. Without standardization, meaningful comparisons are more difficult to derive without reading and learning many metadata management styles. Also metadata standards are important when browsing and find required data from NSDI.

2. METADATA AND SPATIAL DATA

The main problem for the management of distributed GI Services is the heterogeneity of geospatial data models and formats required in different GIS applications. A service is described in ISO 19119 Services draft international standard document as "*distinct part of the functionality that is provided by an entity through interfaces*". The definition of service includes a variety of applications with different levels of functionality to access and use geographic information (ISO 2, 2001). One of the well- known GI services is catalogue service which able users to search, access etc. spatial data via metadata records. Currently, one of the popular solutions to the problems of distributed GI services is to create metadata associated with geospatial data items and services, which can be interpreted by users or metadata search engines. Metadata becomes the key to bridge the heterogeneous environments of distributed GIS databases and services and to provide users with the semantics and syntactic of GIS databases (Plewe and Johnson, 1999).

Users of geographic information have faced several difficulties over the last few years including lack of awareness and supply of GI by many producers, pricing levels targeted to large corporate users rather than small users, and difficulties in searching and retrieving GI partly due to the policy of search engines to ignore metadata tags, and index text instead, and partly due to the lack of an agreed standard on metadata for GI. This is hopefully being changed in last years thanks to usage of metadata.

Metadata is data about data. In other words, it is a structured summary of information that describes the data. It includes, but is not restricted to, characteristics such as the content, quality, currency, access and availability of the data. For spatial information or information with a geographic component, metadata deals with the "what, when, who, where and how" of the data. (ANZLIC, 2001).

The term metadata is used differently in different communities. Some use it to refer to machine understandable information, while others use it only for records that describe electronic resources. In the library environment, metadata is commonly used for any formal scheme of resource description, applying to any type of object, digital or non-digital.

There are three main types of metadata (NISO, 2004):

- *Descriptive metadata* describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords.
- *Structural metadata* indicates how compound objects are put together, for example, how pages are ordered to form chapters.
- *Administrative metadata* provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. There are several subsets of administrative data; two that sometimes are listed as separate metadata types are:
  - *Rights management metadata,* which deals with intellectual property rights, and
  - *Preservation metadata,* which contains information needed to archive and preserve a resource.

Although there are many types of metadata in many different arrangements, it is important to choose the right type after detailed examinations according to the requirements of the users. The usages of the metadata seem to be very different at first. However, the main difference lies in the level of abstraction or generalization applied to the dataset. From that point of view the border between data and metadata begins to disappear: metadata can be seen as a generalization of a dataset to the desired level of detail (Vckovski, 1998).

3. METADATA STANDARDS

A metadata standard is needed to facilitate the exchange of data and to help ensure that users are aware of the limitations imposed by the methods and accuracy of its collection and the decisions made during its development. A geospatial metadata standard helps GIS developers describe the data they create which increases the data's value. Without a metadata standard, it is difficult to determine what spatial data exist, the quality of the data, how appropriate the data are for a given use, and who to contact about the data.

In order to share data effectively, it is essential that the data providers and data users choose common metadata elements to describe a dataset. Studies are currently being undertaken to establish international standards relating to the composition of metadata (Salgé, 1999).

Some examples of international standards are as follows:

- ISO 19115 Geographic Information – Metadata,
- CEN / TC 287,
- United States Federal Geographic Data Committee (FGDC) standard, the Content Standards for Digital Geospatial Metadata (CSDGM),
- Dublin Core – Online Computer Library Centre,
- OpenGeospatial Consortium (OGC).

Since the metadata is the key element of a NSDI, the countries which are struggling to implement national SDI's or using it for years have constituted their national metadata standards. However, positive effects of the globalization are leveraging those countries to conform to the international standards such as ISO 19115. For example, FGDC has been carrying out activities to harmonize ISO 19115 with FGDC's CSDGM (FGDC-STD-001-1998) (FGDC, 2005). Besides, other countries such as Australia, Denmark, and Norway etc. are creating their own metadata profiles based on ISO 19115.

3.1 ISO 19115 Metadata Standards

Many organizations and groups are active in the field of standardization. The "official" standards are published by the ISO. Through the work of its Technical Committee 211 (ISO/TC 211) the ISO has taken the leading position in the standardization of geographic information. The ISO/TC 211 has created a complete suite of standards for vector-based GIS which integrate all major developments in this field.

Most of the ISO 19100 standards contain abstract solutions. Standards on the implementation level have been defined by other organizations such as OGC. In many cases the implemented solutions are well established existing formats or environments updated. (Kresse, 2004).

The ISO 19115 "Metadata" is probably the best known standard of the ISO 19100 family. It provides a large basket of metadata elements that are needed in applications of geographic information. The ISO 19115 consolidates well known sources such as the metadata listing of the FGDC in one standard. The large software providers for GIS have begun to incorporate the ISO 19115 in their system enabling the automated and backgrounded generation of metadata elements while the GIS data are updated. (Kresse, 2004).

ISO 19115 was published as international standard in 2003. The metadata-model of the ISO 19115 distinguishes between about

20 core metadata elements and the comprehensive listing with about 400 elements (ISO/TC 211, 2005).

4. STUDY OVERVIEW AND METHODOLOGY

The consistency in the metadata content and style is recommended to ensure that comparisons can be made quickly by data users as to the suitability of data from different sources. Without standardization, meaningful comparisons are more difficult to derive without reading and learning many metadata management styles.

Turkey is one of the countries which are struggling to establish an NSDI. For this purpose several organizations are discussing about the legislative part of the infrastructure, while technical arguments are being carried out in GIS communities.

When the spatial data is considered in Turkey, the situation is mostly unclear. One of the big problems in this area is unorganized institutions. These organizations are managing their own data for local purposes. However, the solutions for the spatial problems are looking effective, data sharing and spatial standardization problems are appeared from the other side. Especially, since there is no complete metadata for the spatial data, browsing and find the required data is a dilemma. It is not easy to find the data from the institute's databases, therefore many users and producers are forced to collect the duplicate data.

When the duplication and data searching problems are considered, the solution seems to store the metadata in NSDI concept.

In this study, it is aimed to standardize the metadata elements, create the schemas and create a tool for collection and creation of standard metadata tool. Details of the proposal metadata framework are not given in this paper, since the web based multifunction metadata tool is mainly focused on.

During the study conceptual, logical and physical models are implemented using UML.

The carried out steps for the study are as follows:

- Creating draft metadata profiles,
- Designing and coding of schemas for XML,
- Modelling and implementing the multifunction metadata tool.

4.1 Creating Draft Metadata Profile for Turkey

During the study, existing metadata standards and their implementation frameworks were examined and requirements of national institutes were revealed after detailed investigations. Those metadata standards are used for the generation of operational, hierarchical metadata framework. In this framework, the metadata elements are categorized in 14 packages according to ISO 19115 (Figure 1). For the beginning, the elements of ISO 19115 Metadata standard were taken as base, while some elements were elected which are not required for Turkey. However, those metadata elements are being extended according to the requirements of national institutions with the conformance to metadata extension methodology of ISO. During the process other national metadata standards, FGDC standards, Dublin core metadata standards and ISO 19115 metadata standards were examined individually.

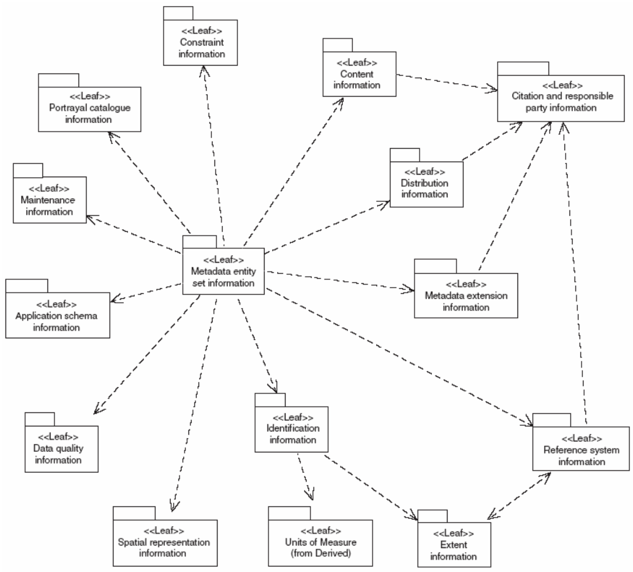

Figure 1. 14 metadata packages of ISO 19115 (ISO 1, 2003)

As the elements of the metadata packages, all the sub elements of packages are created as classes and attributes of it in logical model (Figure 2).

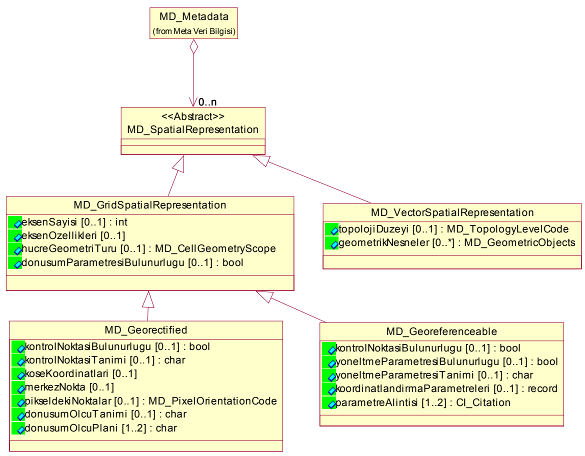

Figure 2. A part of a metadata logical model

4.2 Designing and Coding of Schemas for XML

For the metadata record to be truly useful, it must be capable of being readily exchanged and of being read by software that indexes, searches and retrieves the metadata records. To achieve this, the metadata record must be available in a well-structured and reliable format.

The format should be software and hardware independent, in a particular standard, flexible and receptive to the extensions (W3, 2005).

XML covers those requirements. XML is a plain-text format that is independent of computing platforms, vendors, and software. XML is a so-called "meta-markup language" that has been developed to convey data. It is used to encapsulate data into files that can be (ANZLIC, 2001)

- displayed within web browsers,
- exchanged across the Internet between different computer applications and businesses, and
- Stored in and retrieved from databases.

XML has international support, and there is now consensus within the geospatial community that metadata should be exchanged in this format. This approach has been incorporated in the ISO 19115.

Therefore, in order to store the metadata XML is selected as standard in this study. In this context, the Document Type Definition (DTD) which defines the set of structural rules and relationships and allows for the validation of metadata structure is coded. Additionally XSD schemas are created from DTD files for forthcoming studies, since the XSD is more advanced schema language than DTD.

XML DTD schema is prepared for standard metadata forms using UML (Figure 3). For flexible structure all the metadata components are defined as elements in DTD, however domains are attributes.

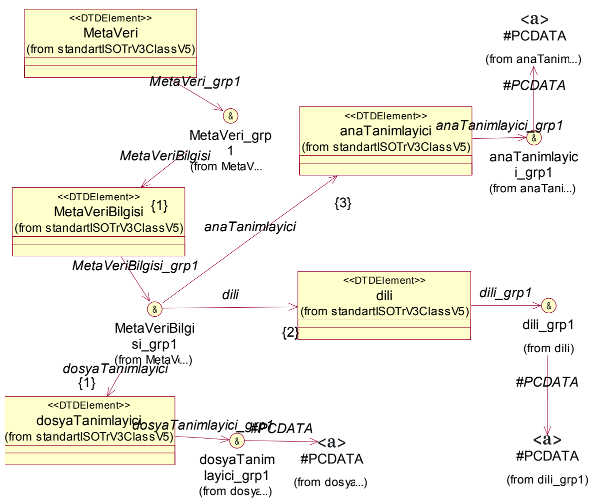

Figure 3. A part of DTD design using UML

Since the metadata tool was considered web based, the structure should have been conformed to the web environment. Therefore XSL files are prepared for presenting the metadata on web browsers (Figure 4). XSL schemas are coded flexible as well, since users may want to avoid seeing the null data or unrelated elements of the metadata.

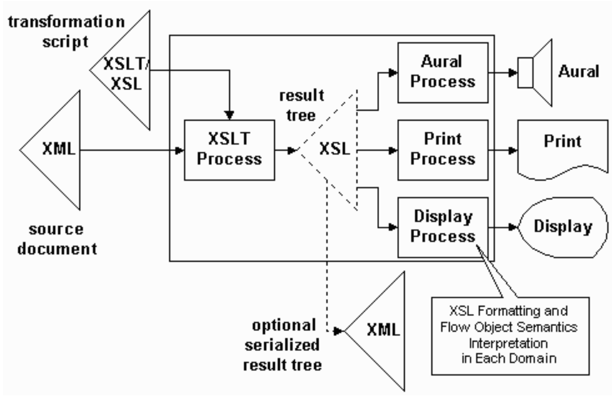

Figure 4. Process of XML (Holman, 2000)

4.3 Modelling and Implementing the Multifunction Metadata Tool

A NSDI includes several tools to enable users search, find, and access and retrieve data from the related spatial servers. One of those tools is the metadata tool. Metadata tool is a necessity when users require processing the metadata. These processes can be creation of the metadata or editing of an existing metadata from the database. Those processes are distributed on a large dovetail.

The metadata tools include functions for entering and editing metadata and utilities for preprocessing, extracting, post processing, validating, and viewing metadata. Most of the tools were designed to help complete CSDGM metadata in the world, but several have been tuned to produce specific local metadata profiles. There are also many varying examples of metadata tools with different functions.

In this study it is aimed to design and implement a complete web based tool for processing metadata in one tool.

4.3.1 Design of Tool: The metadata tools are important when the metadata of the spatial data is considered. Although, metadata elements were standardized during the projects, it is required to collect the metadata files in a standard structure. Those metadata are used for especially search and comparison of the metadata records. Therefore collecting and storing the metadata is as important as the metadata standards, because, it is not possible to make meaningful comparison between different formatted metadata files.

During the design of the metadata tool, existing metadata tools were examined in order to define the requirement functions, addition to the local researches. Finally it was decided to include the following main functions in the metadata tool:

- Tracking the users for implemented processes,
- Creating a new metadata document,
- Editing the existing metadata document,
- Checking the available metadata documents,
- Viewing the existing metadata documents,
- Retrieving the schema files.

The main functions arranged above also include sub functions for additional processes. Some sub functions are given in use case diagram Figure 5.

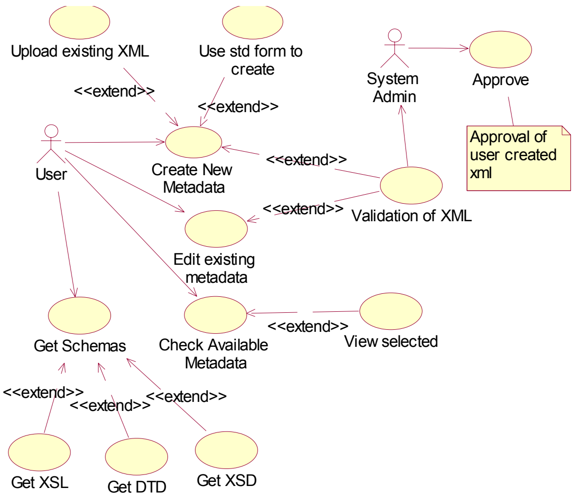

Figure 5. Use case diagram of the metadata tool

The tool was designed as flexible as possible. For example, user can upload a metadata documents file in two ways. First, user can fill a standard form to create a metadata document in XML file. Secondly, user can upload an existing or a self-created metadata document via tool. Filling form option provides minimum requirements of standards. Therefore, user can add or delete records according to the schema files. To avoid any error in user created XML files when uploading, user is obliged to validate XML file according to the schemas.

During the study, all life cycle of the metadata tool is designed via UML.

4.3.2 Implementing the Multifunction Metadata Tool: Since the metadata tool is one of the components of a NSDI, it must be strictly related to the other components like catalogue service. Those components should interact; therefore they must be in a similar structure. When the complete system considered, the metadata tool creates the standard records, catalogue searches the entered criterions, and redirects to the related WMS, WFS etc. services.

For this study, implementation of the tool is settled by .NET Framework and SQL Server 2000. The tool was designed web based in order to provide platform and software independency. The web based design also provides users to use one tool from one source without downloading and installing any component to the personal computers. In this way, it was aimed to ensure metadata processes to be in a standard structure.

In order to provide software independency, it is avoided to use any object library of a commercial company. So, standard object library of .NET framework was used, and in some cases additional components were developed when required. SQL server is used for storing the metadata XML files. So, a simple database table structure was designed for storing procedure (Figure 6). This table structure was connected to spatial data which was stored in SQL by commercial software. This relation gives design to be tested with other software packages.

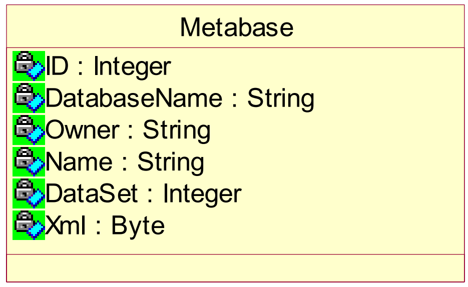

Figure 6. Class diagram of simple table

In this study, XML files are stored in database as binary, since the field of table is limited. During the store procedure XML files are converted to BLOB, and then uploaded to the database field. BLOB is an acronym for binary large object. It refers to any random large block of bits that needs to be stored in a database, such as a picture or sound file. The essential point is that a BLOB is an object that can’t be interpreted within the database itself (Calsavara, 2002).

Also, it was considered, when selecting the BLOB, to allow users to search column data for text strings. And full-text indexing enables searching of binary data types by associating the column with its target application (Otey, 2002). This assumption will be required when the metadata is connected to the catalogue service for further studies.

The tool was designed to mainly work with Draft Turkish Metadata Standard. However, there are some institutions which have created metadata, such as metropolitan municipalities, according to the FGDC Metadata Standards. However, there was no legislation to create metadata for spatial data; those institutions were created them by themselves. Therefore, tool was designed to upload metadata files created according to FGDC-STD-001-1998 standard as well. Users are able to upload existing XML files which are created according to Draft Turkish Metadata Standards or FGDC-STD-001-1998 schemas. Only necessary process to be done by user is validating the XML file according to a schema by the tool. Then system allow user to upload it to the database.

But, when users want to create metadata using the standard form prepared in the metadata tool, there is no option to choose the schema standard. In that case metadata is created according to the national draft standard.

For existing metadata records in database, user can view according to a schema and edit it using form. However, editing metadata via form works only for metadata which conforms to the national draft standard.

After all creation or before uploading processes, metadata files needs to be validated according to the related schema. Then, process is informed to the system administrator for approval of process.

During the all processes, users are tracked for the implemented works for ownership. But system will be worked according to the authorizations to avoid keeping system administrator busy, addition to the login procedure.

5. CONCLUSION

Metadata is the key element of a SDI. It is used for mainly searching and comparing datasets in order to find the required data when available.

A metadata standard is required to facilitate the exchange of data. A geospatial metadata standard helps GI communities to describe the data they create which increases the data's value. Without a metadata standard, it is hard to define what spatial data exist, the quality of the data, how appropriate the data are for a given use, and who to contact about the data. Therefore it is essential to create metadata standards. Especially OGC and ISO/TC 211 are working on spatial data standards, and ISO published 19115 Metadata Standards as the international standard. It presents very detailed metadata elements, sometimes more than necessary, for spatial data. But it is also give chance to create own profiles to the national standard makers. Therefore it is talent of standard makers to create a perfect profile. However, it is effective source to take as base.

In this study, to stimulate the creation of basic metadata records for the cataloging of spatial data sets, multifunction web-based metadata tool has been written. This tool has been developed for authoring, creating, viewing, editing and publishing metadata of spatial data. This tool is being developed and it is planned to be a part of cataloging service for querying, redirecting etc.

Since the collection of metadata is as important as metadata standards, metadata tools are inevitable to be used for geospatial data.

REFERENCES

ANZLIC, ANZLIC Metadata Working Group, 2001, *ANZLIC Metadata Guidelines: Core metadata elements for geographic data in Australia and New Zealand*, Version 2, Australia.

Calsavara A. P., 2002, *A portable alternative for BLOBs*, http://www.builderau.com.au/architect/database/soa/A_portable_alternative_for_BLOBs/0,39024547,20267938,00.htm.

FGDC, 2005, *FGDC/ISO Metadata Standard Harmonization*, http://fgdc.er.usgs.gov/metadata/whatsnew/fgdciso.html.

Government of India Department of Space (ISRO), 2003. *NSDI Metadata Standard*, Indian Space Research Organization NNRMS Secreteriat, Version 3.0, India.

GSDI, Technical Working Group Chair, 2004, *Developing Spatial Data Infrastructures:The SDI Cookbook*, Version 2.0, www.gsdi.com.

Holman G. K., 2000, *what is XSLT?* http://www.xml.com/pub/a/2000/08/holman/index.html?page=1.

ISO 1, 2003, *ISO 19115 "Geographic information - Metadata"*, ISO Central Secretariat, Geneva

ISO 2, 2001, ISO 19119 *"Geographic information- Services"*, Draft International Standard, Version 2.0, ISO Central Secretariat, Geneva

ISO/TC 211, 2005, ISO Technical Committee Web Site, www.isotc211.org.

Kresse W., 2004, Standardization of Geographic Information, In: *Proceedings of the XXth ISPRScongress*, Istanbul, 2004.

National Information Standards Organization (NISO), 2004, *Understanding Metadata*, NISO Press, USA, ISBN: 1-880124-62-9, 2004.

Otey M., 2002, *Return of the BLOB* http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnsqlmag01/html/blob.asp.

Plewe, B.S., Johnson, S.R., 2000, *Automated Metadata Interpretation to assist in the use of unfamiliar GIS data source,* In: Goodchild M.F., Egenhofer, M.J., Fegeas, R., Koffman, C.A. (Ed.), Interoperating Geographic Information Systems. Kluwer Academic Publishers, Boston, pp 203-214.

Salgé F. , 1999, *National and International Data Standards,* In: P. Longley, M. Goodchild, D. Maguire, and D. Rhind (Ed.), Geographical Information Systems. 1 Principles and Technical Issues, John Wiley & Sons, New York, pp. 693-706.

Vckovski, A., 1998, Interoperable and Distributed Processing in GIS, Taylor & Francis, London.

W3, 2005, World Wide Web Consortium, *Extensible Markup Language (XML)*, www.w3.org.