

T. KLIMENT

METAINFORMATION INFRASTRUCTURE FOR GEOSPATIAL INFORMATION

ABSTRACT

The Infrastructure for Geospatial Information consists of an information component, where the bearers of geoinformation are geospatial data, and a metainformation component. The bearers of metainformation are metadata, which describe these geospatial data. Metadata serve as a tool for discovery, retrieve and evaluation of geospatial data within the Infrastructure of Geospatial Information. In order to acquire metadata from a metadata producer to a metadata consumer, the implementation of various tools – components of Metainformation Infrastructure for Geospatial Information is required. This paper describes the main principles and components of Metainformation infrastructure and the possibility of using open source tools for their implementation in practice.

1 INTRODUCTION

Geospatial data (GD) represent a great treasure through their form of inputs to various projects in our society. The results of these projects bring new information and knowledge, which can be subsequently used as inputs to other projects. Since the basis of this process is high quality easily accessible GD, it is necessary to provide opportunities for easy discovery, evaluation and retrieval. These are the keywords of one part of Infrastructure of Geospatial Information (IGI), which we can denominate as the Metainformation Infrastructure for Geospatial Information (MIGI). These components are used in the discovery of metadata managed within geocatalogues and thereafter the evaluation of GD and retrieval for the end user of IGI. Therefore, metadata (MD) and Geocatalogue are the basic components that facilitate accessibility and interoperability of the geospatial resources provided by IGI (Nogueras-Iso et al, 2005).

The paper describes the theoretical aspects of MIGI components and the practical testing and modifications of chosen open source tools from the field of MD editors, the Metainformation Systems (MetaIS), the implementation of Catalogue Service and desktop GIS (Geographic Information System) tools, which the author carried out in the written part of his dissertation examination (Kliment, 2009). Main objective is an assessment of the actual status of the implementation of the International standards for digital geographic information (ISO/TC 211 Geographic information), technical specifications (OGC - Open Geospatial Consortium) and a description of the functionality and technology used in the creation of these tools. Open source tools are an excellent solution for the implementation of components of MIGI in practice, particularly in the non-commercial field, because they are usually free and offer open source code for additional development. It is possible to install, test and modify these tools according to one's specific needs.

Tomáš KLIMENT

email: tomas.kliment@stuba.sk
Research field: Spatial Data Infrastructures (SDI) - Metadata, Discovery Service, Infrastructure for Spatial Information in Europe (INSPIRE)

Address: Department of Theoretical Geodesy,
Faculty of Civil Engineering,
Slovak University of Technology in Bratislava,
Radlinského 11, 813 68, Bratislava, Slovakia.

KEY WORDS

- Metadata,
- Metainformation,
- Metadata Catalogue,
- Metainformation System,
- Metainformation Portal,
- Catalogue Service

2 MAIN COMPONENTS OF METAINFORMATION INFRASTRUCTURE FOR GEOSPATIAL INFORMATION

The main components of MIGI are Metadata Catalogue (MetadC), MetaiS, Metainformation Portal (MetaiP), the implementation of the Catalogue Service – these components are commonly implemented as complex solution named as Geocatalogue. Additional components as Geoportals and desktop GIS tools, which have nowadays frequently implemented management and distribution of MD may also be classified as components of MIGI. A conceptual model of IGI with main focus on MIGI components and their relations is shown in Figure 1. The basic “metainformation stores” are MD warehouses, which we called the Metadata Catalogues. Each MetadC may comprise of one or more MD databases. For centralized MD management of particular MetadC within organization is implemented MetaiS. The particular MetaiS for group of organization are centralized to MetaiP within region, state or union. MetaiP provides cross connection of MIGI components to entire IGI and may be either independent component or part of Geoportals – main web access point to IGI either national (NIGI) or multi-national level. The main “engine” to serve metadata among already mentioned elements is catalogue service.

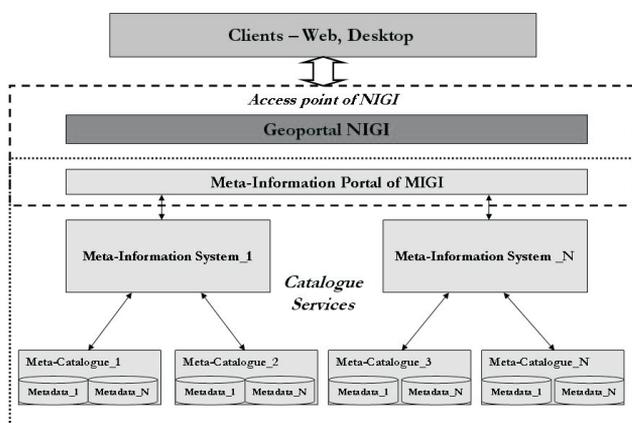


Fig. 1 A conceptual model of IGI – metainformation part – MIGI

2.1 METADATA AND METADATA-CATALOGUE

The term “metadata” can be most easily defined as data about data, respectively information about data (metainformation). MD can be regarded as summary information that can be attributed to an abstract or real object, process or phenomenon. This means that the MD must have meaning and application in various fields of human

society. Even in the domain of IGI the use of MD is most important, because it improves the quality and quantity of GD gained within the infrastructure. Detailed aspects of the MD domain are described within authors written part of dissertation examination (Kliment, 2009), which among others defines the basic characteristics of MD, namely, that metadata should:

- reflect the most actual state of a described source,
- constitute an auxiliary element for understanding the functionality and content of the source,
- be portable and consistent,
- facilitate discovery, comparison and assessment of the sources of a variety of points of view,
- be entered in an appropriate way for communication, distribution, interpretation and processing,
- facilitate the sharing and exchange of information about the source,
- provide information about the point of access to the source (URL, etc.).

The result of this approach is a set of the MD records, which describes the GD. These MD records are commonly centralized within a storage area. The storage area can be defined as a Metadata Catalogue, in which we can imagine a large set of MD records in a database. The basis for MetadC is that they store a large number of heterogeneous MD (descriptive information about different resources). MD management in MetadC can be performed using a tool for the administration, which can be a part of the installation of database management system (DBMS), or by using a Meta-Information System. Second approach is standard within IGI.

2.2 METAINFORMATION SYSTEM AND METAINFORMATION PORTAL

A MetaiS is a general name for any information system, of which the database is a MetadC (Ožana, 2007). The key features and functionalities of any MetaiS are shown in Figure 2. A MetaiS should provide an attractive, user-friendly interface for the effective discovery of MD and the display of their content. In addition to these basic functions, MetaiS should provide the functions for MD management (creating, editing and deleting of MD records). For the distribution of existing records it should provide functions to import records from other systems or export existing records for further use in other systems. From the point of view of the system and the MD administrator, a MetaiS should provide the functions for its administration and also the end-user functions for user support in the form of documentation and tutorials. The key function from the point of view of the common user (MD consumer) of MetaiS is the discovery and view of MD records content. With

regard to a query for GD, it is necessary to ensure that a MetaiS should provide classical full-text discovery and also so-called extended discovery, where one can define the additional discovery parameters, for example, spatial, temporal and thematic parameters in order to enhance the relevancy of the results of the discovery process. A MetaiS may be divided into public and private MetaiS and corporate and department MetaiS (Ožana, 2007). These MetaiS may be connected to MetaiP within MIGI. A MetaiP is a tool which makes MD records from various MetaiS available (for example, the government department MetaiS in Slovakia) in a uniform method within a central interface (See Figure 3). A MetadC of a MetaiP maintains only a minimum set of MD about geospatial sources and MD about federated MetaiS. The aim is to simplify the discovery of MD records from a variety of MetaiS from a country or a region (for example, the European Union) for the end-users of IGI. It unifies the display method of MD content and management of the MD records of various types and natures (heterogeneous MD records). A MetaiP is the central point of access to MD, e.g., on a national or multi-national level of an IGI.

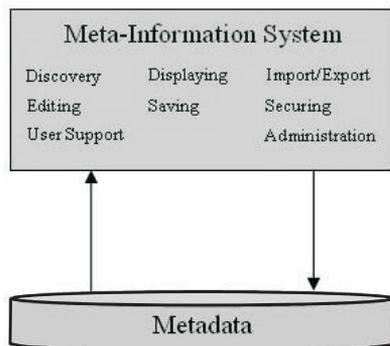


Fig. 2 Metainformation System

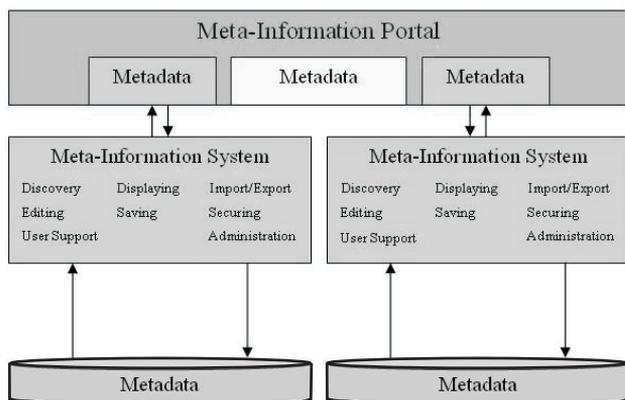


Fig. 3 Metainformation Portal

2.3 CATALOGUE SERVICE

A Catalogue Service implementation on the web is at the top of MIGI, particularly an example of a CS called the Catalogue Service for the Web (CSW) (OGC, 2007a). With CSW a MetaiS or MetaiP sends user query against MD records content, which are stored in MetadC. The user query can be also send to federated CSW to perform distributed discovery. The service responds to the request of a client by displaying relevant MD records based on the query's parameters. A reference model of the catalogue services adopted by the OGC (OGC, 2007a) for the development of interfaces of these services is shown in Figure 4. The model defines both the possibilities of the discovery of geospatial resources from three possible ways (a local service MetadC, services (mapping, downloading, etc.) within geospatial sources and other CS in the distributed network) and two interfaces (OGC and e.g. proprietary solution).

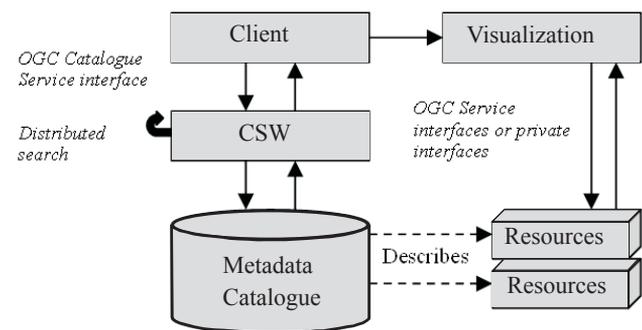


Fig. 4 A reference model of the Catalogue Service defined by OGC (OGC, 2007a)

2.4 GEOPORTAL

Geoportals represent a gateway to the IGI on the web. It is in fact a portal, that may provide the possibility for users to discover, view, download, transform and process geospatial sources. The MetaiP can be a part of the Geportal. Geportal's architecture is based on web services that are invoked by users within its interface. Hence, it concerns the Service-Oriented Architecture (SOA), the principles of which are described within written part of dissertation examination (Kliment, 2009). The general implementation of the Geportal complex is defined by the 4 main services (Portal Services, View Service, Data Services, and Catalogue Services) (Ožana, 2007). The architecture of the Geportal divides the whole system into smaller logical elements that are autonomous

and provide for the partial functionality of the system. These elements are easier to modify and adjust. The architecture ensures the openness and accessibility of the system, particularly on the basis that these services are built on ISO standards and OGC specifications, which have gradually been introduced into most systems within the geospatial domain.

2.5 DESKTOP GIS

Nowadays many desktop GIS tools have implemented functions for the creation, management, distribution and publishing of MD about geospatial sources, with which we may be working in these environments. It is possible to actually work with geospatial sources and their descriptions and then publish MD for instance in the MetaIS. These tools represent a desktop client of IGI.

3 MAIN DESCRIPTION AND FUNCTIONALITY OF CHOSEN OPEN SOURCE TOOLS FOR THE IMPLEMENTATION OF THE MIGI COMPONENTS

Following part of paper describes open source tools that were chosen according to their wide usage within IGI domain (<http://geonetwork-opensource.org/geonetwork-nodes>) and author practical experiences achieved within bachelor, master and parallel doctoral study.

The first tool is the metadata editor CatMDEdit (www.catmdegit.sourceforge.net), which was developed in Spain at the University of Zaragoza. The tool can be used for the creation of metadata records on a local basis, with a special focus on the description of GD. This tool provides broad functionality for the creation, management and distribution (export/import) of metadata with regard to the standards ISO, namely ISO 19115:2003 Geographic information – Metadata (ISO, 2003) for MD content and ISO/TS 19139:2007 Geographic information – Metadata – XML schema implementation (ISO, 2007a) for MD content coding. Tool supports current activities in Europe in IGI domain as well in metadata creation possibilities according to INSPIRE (Infrastructure for Spatial Information in Europe) (Directive, 2007) CatMDEdit was developed in the Java programming language and distributed as open source software through the web. It is multiplatformed, and can be installed in Microsoft Windows and UNIX operation systems. It is a multilingual tool and currently supports six European languages and can be transferred into the Slovak national environment. It provides a very user-friendly graphic user interface (GUI) (See Figure 5). Its additional functionality is described in (Kliment, 2009).

Another tool is the MetaIS GeoNetwork opensource (www.geonetwork-opensource.org), which began as a catalogue system

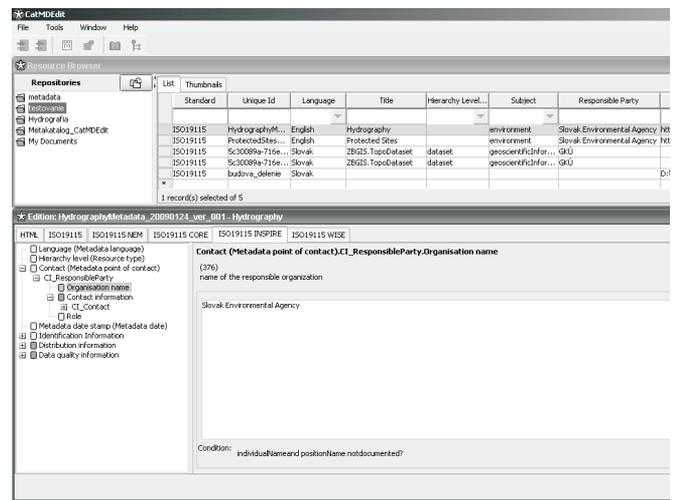


Fig. 5 GUI of the MD editor CatMDEdit

for geospatial sources - Geocatalogue. Currently, this software is widely used as an IGI base worldwide. The GeoNetwork opensource is an application, which allows users to implement the comprehensive management of metadata in a distributed form via the Internet. It is based on standards and provides geospatial data through MD, which can be freely shared among different communities. The system offers interfaces for easy and quick access to georesources. GeoNetwork was developed as a web application and among all technologies that are used within its development Java platform is their connection link. The system has a Jetty web server within the installation package and also supports the Apache Tomcat server. GeoNetwork stores all data in the relational DBMS. It supports MySQL, PostgreSQL and Oracle DBMS. To access the DBMS system uses the Java Database Connectivity (JDBC) application programming interface (API), which defines how the client may access the database. The tested version of GeoNetwork is v. 2.0.2 from April of 2008. It provides a number of upgrades from the previous 2.0.1 version from September 2007. The current version offers a more intuitive and user-friendly interface and tools for user interaction with the system. The discovery interface has been completely redesigned for easy and interactive discovery. The system has implemented GeoServer as the map server. Users no longer need the data only overlap with the available Web Map Service (WMS) (ISO, 2005 and OGC, 2006), but can directly create their own WMS for other users of the system without installing additional plug-ins. The GeoNetwork is in accordance with standard ISO 19115 (ISO, 2003) and MD are encoded regarding to ISO 19139 (ISO, 2007a). The system has a built-in MD editor, which contains MD schemas such as the Dublin Core (ISO, 2009), the

Federal Geographic Data Committee (FGDC, 1998) and ISO (ISO, 2003) for the creation of MD records. MD can be created in the default, advanced or XML views. The current version provides a number of different interfaces for harvesting, which allow users to connect their server to various catalogues around the world in the distributed network. This is the result of open-source references of the OGC consortium for CSW. Harvesting is fully compatible with version 2.0 of GeoNetwork nodes CSW.

Another tool is the implementation of CSW deegree Web Catalog Service v. 2.2 (www.deegree.org), which is being developed in Germany at the Department of Geography at the University of Bonn. In the deegree project Java solutions are developed, which provide the main building blocks for the IGI. All the Java modules are controlled by one central servlet. Most web servers support this technology, so this fact makes the deegree solutions a universal product. The actual version of the deegree CSW 2.2 service implementation works in a Java Runtime Environment (JRE) in versions 1.5.x and is controlled by Apache Tomcat servlet in versions 5.5.x and higher. DBMS is PostgreSQL with a PostGIS extension, which enables the saving of spatial objects to a database. The entire architecture is developed according to the OGC and ISO standards mentioned above. The implementation of deegree CSW is able to provide metadata in different standards in parallel with the same database. This is possible because the CSW uses the Extensible Stylesheet Language Transformations (XSLT) process to transform the requests and responses to the required format. The service does not have its own module to access the MD, but uses the OGC Web Feature Service operations (OGC, 2005) (See Figure 6). CSW deegree provides users with a high degree of flexibility in view of its configuration options and output formats. The service works with a wide range of data sources and server environments. The configuration of CSW deegree requires editing XML files, which manage and control the functionality of the server.

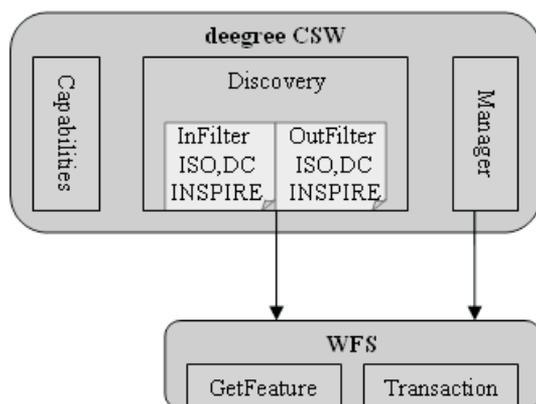


Fig. 6 Architecture of deegree CSW

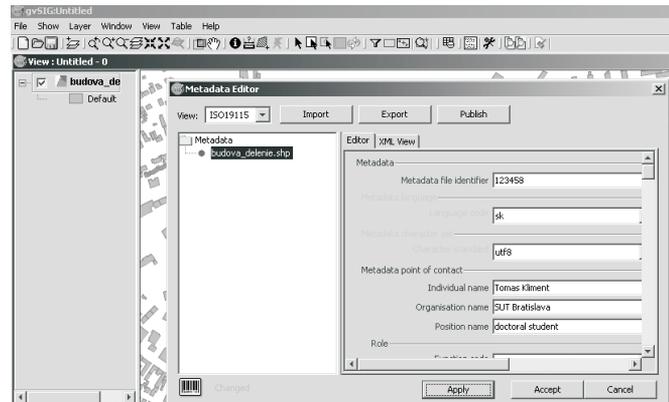


Fig. 7 GUI of metainformation component of GIS tool gvSIG

The installation of the CSW contains a simple Web client. Using XML messages the user can discover, insert, delete and edit MD from a MetadC. The responses are also answers in the form of XML views

The latest tool is gvSIG a desktop GIS (www.gvsig.gva.es), which is a part of the projects under the Open Source Geospatial Foundation (www.osgeo.org). It allows the user to work in the user-friendly GUI, in which it is possible to create interaction with the OGC Web services (WMS, CSW). The architecture is based on the Java technologies. In the version gvSIG 1.1.1 a metainformation component is also implemented (See Figure 7). The program allows the user to discover and view MD in the basic installation from remote MetadC by defining the URL (Uniform Resource Locator) address of CSW. It is possible to install the prototype of a MD extension, which allows creating, importing, editing and publishing the MD in the ISO Core schema (ISO, 2003).

4 TESTING – OVERALL DESCRIPTION AND TEST OBJECTIVES

A main objective for above described tools is usability testing for workflows that could be performed either by MD producer or MD consumer or establisher of IGI. To achieve this main objective 4 testing scenarios were proposed and performed:

1. MD producer creates a MD record from the scratch by CatMDEdit following ISO (ISO, 2003) and INSPIRE (Directive, 2007) specific constraints and then distributes (export/import) created record in XML format to the MetaIS GeoNetwork opensource.
2. MD producer extracts MD from geospatial data (provided by Slovak Environmental Agency) in Geography Markup Language

- GML (OGC, 2007b and ISO, 2007b) and distributes to the MetaIS GeoNetwork.
- 3. MD consumer searches, discovers and views MD records created and imported in previous 2 sections via both the web client of the MetaIS GeoNetwork and the desktop client tool gvSIG.
- 4. IGI establisher inspects and verifies the functionality of the implementation of degree CSW and modifies GUI of MetaIS GeoNetwork opensource.

In the following part, paper provides the main assessment characteristics of tested tools regarding to proposed testing scenarios and description of modification of GUI performed on MetaIS GeoNetwork opensource. All test scenarios are described in detail in (Kliment, 2009)

4.1 MODIFICATIONS OF THE GEONETWORK OPENSOURCE

It is possible to modify many things in open source software, because we can freely adjust all the configuration files (mostly in XML form) and source code as well. We can change the GUI (logos, colors, fonts, “GEO look”) or we can localize these tools in the national environment and, at last but not least, we can extend the functionality of this tool by source code modifications. The GUI were changed and localized to national environment of Slovakia by author. The visual presentation of GeoNetwork can be adjusted through the editing:

- Documents for visualization – eXtensible Stylesheet Language (XSL),
- Cascading Style Sheets (CSS).

By editing XSL documents which are stored in the local directories of a software installation, it is possible to change the generated HyperText Markup Language (HTML) code and thus change the arrangement of the web pages generated. By editing CSS can be accomplished modifications of the visual presentation of the generated HTML web page. The CSS are stored in the local directories of the software installation folder as well.

The actual installation of the GeoNetwork is completely translated into four languages (English, French, Spain and Chinese). The author localized GUI to Slovak national environment. The process of localization consists of three basic steps. In the first step it was necessary to modify an the existing menu of GeoNetwork, used to change the application language. This menu is defined in the file banner.xml. The author added the required code to the appropriate site in this file. In the second step the author modified the contents of the string.xml files in each language folder and added code: `<sk>Slovenčina</sk>` to these files. In the third “main” step a folder was created for the Slovak language, and all the files from English language folder were copied. In the final localization process the strings in the elements of all XML files were translated. The final

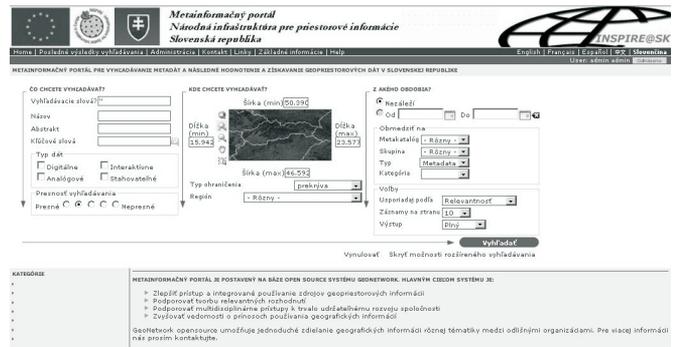


Fig. 8 A new GUI of GeoNetwork localized to the Slovak environment

GUI, which was localized to the Slovak national environment is shown in Figure 8.

4.2 EVALUATION OF THE TESTED TOOLS

CatMDEdit is a very user-friendly and simple tool, which provides broad functionality for the creation and management of MD records for geospatial resources. It has a simple installation process and does not need to use any Internet connection during its use, which can be on a local basis when creating a large number of MD records, respectively their generation, complementation and, validation. Their export to XML files, which is in accordance with ISO (ISO, 2003 and ISO, 2007a) standards, is an advantage in phase of distribution of generated MD records between editor and MetaIS. First testing scenario was passed using CatMDEdit without any system and functional constraint. Created MD record was exported to ISO 19115/19139 XML file and uploaded to GeoNetwork database – prepared for discovery via the web.

The GeoNetwork provides a powerful tool for the implementation of all components of MIGI, and the application environment is more than user-friendly. It offers well-arranged MD management, which provides for their creation directly through the web interface, their distribution with imports and exports, their display and finally their discovery. It uses CSW to communicate with remote Geocatalogues. The system is in compliance with the fundamentals of distributed discovery. Using this system is particularly beneficial in building components of MIGI. MD records from testing scenario 1 and 2 were uploaded to the metadata catalogue of GeoNetwork without any constraints. First part of scenario 3 was passed and system appeared imported MD records after searching performance.

gvSIG is a desktop GIS tool, which in addition to a great functionality in the field of GD allows MD discovery from the remote Geocatalogues (tested only from GeoNetwork). The user



can retrieve the location of geospatial sources investigating the MD records content through CSW. After that the user can view, download, process etc., these data. This desktop GIS tool can be used as the discovery client within the infrastructure. Second part of scenario 3 was passed and tool appeared MD records from GeoNetwork after searching performance within gvSIG GUI.

degree CSW introduces a solution for CSW implementation. It does not provide user-friendly GUI for communication with the service (discovery and management of MD records) currently. The installation is more complicated than in GeoNetwork. It is necessary to install the servlet and then DBMS separately. Finally, one can configure the service via a set of XML type files. By eliminating these „shortcomings“ and developing an optimal user-friendly GUI for the end-user this architecture can be classified as a powerful tool for implementing the components of MIGI into practice.

5 CONCLUSION AND FUTURE WORK

The paper provides summary of authors knowledge gained during the first stage of doctoral study finished by his dissertation exam. The main objective of the topic, which has been described within this paper, is metainformation about geospatial resources retrieval.

Metainformation is the result of the MD content investigation. In order to deliver the MD to the end user, their workflow consists from a large number of procedures, from their creation, storage, distribution and implementation through the environment of infrastructure. This infrastructure must allow the user to discover and retrieve MD. This environment can be defined as Metainformation Infrastructure for Geospatial Information, which is part of Infrastructure for Geospatial information, which deals with metainformation. For its implementation into practice open source tools can be used, because they are usually free and offer an open source code. It is possible to install, test and modify these tools, according to the specific needs. The results achieved in (Kliment, 2009) and summarized in this paper describe the advantages of the implementation of these open source tools into practice for building a National Infrastructure for Geospatial Information in the Slovak Republic for its near future connection to INSPIRE.

For the future work author would like propose and implement reference Geocatalogue solution for Geodesy and Cartography domain with all requirements arising from European legislation (Directive, 2007) and following standardization in technical and semantic level.

This contribution has been made thanks to VEGA No. 1/4025/07.

REFERENCES

- **Kliment, T. (2009):** Metainformation infrastructure for spatial information. Written part of dissertation examination. Department of Theoretical Geodesy, Faculty of Civil Engineering, Slovak University of Technology, Bratislava
- **Nogueras-Iso, J. et al. (2005):** Geographic Information Metadata for Spatial Data Infrastructures - Resources, Interoperability and Information Retrieval. Springer Verlag
- **Ožana, R. (2007):** Evaluation of GeoNetwork Opensource Features and Suitability for National Metaportal Design. Geoinformatics Institute, VŠB-TU Ostrava
- **Directive (2007):** Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)
- **FGDC (1998):** Content standard for digital geospatial metadata, Document FGDC-STD-001-1998. Federal Geographic Data Committee (FGDC), Metadata ad hoc Working Group.
- **ISO (2003):** ISO 19115:2003 – Geographic information – Metadata. ISO, Switzerland, 2003
- **ISO (2005):** ISO 19128:2005 - Geographic information - Web map server interface, ISO, Switzerland, 2005
- **ISO (2007a):** ISO/TS 19139:2007 - Geographic information – Metadata – XML schema implementation, ISO, Switzerland, 2007
- **ISO (2007b):** ISO 19136:2007 - Geographic information - Geography Markup Language (GML), ISO, Switzerland, 2007
- **ISO (2009):** ISO 15836:2009 - Information and documentation - The Dublin Core metadata element set. ISO, Switzerland, 2009
- **OGC (2005):** Web Feature Service Implementation Specification, v. 3.2.1, OGC 04-094, Open Geospatial Consortium, 2005
- **OGC (2006):** OpenGIS® Web Map Server Implementation Specification, v. 1.3.0, OGC OGC 06-042, Open Geospatial Consortium, 2006
- **OGC (2007a):** OpenGIS® Catalogue Services Specification, v. 2.0.2, OGC 07-006r1, Open Geospatial Consortium, 2007
- **OGC (2007b):** OpenGIS® Geography Markup Language (GML) Encoding Standard, v. 3.2.1, OGC 07-036, Open Geospatial Consortium, 2007