Geonetwork opensource as an application for SDI and education

Stanislav Grill, Michal Schneider

Department of applied geoinformatics and cartography, Faculty of Science, Charles University, Albertov 6, 128 43, Praha 2, Czech Republic
Kontakt.grillst@natur.cuni.cz

Abstrakt. Tento článek se zabývá teoretickými a praktickými možnostmi použití software Geonetwork opensource pro správu infrastruktury prostorových dat a zároveň jeho využitím ve výuce. Geonetwork opensource je primárně systém pro správu metadat pro prostorová data. Se stále rostoucím objemem existujících prostorových dat rostou i potřeby uživatelů data vyhledávat a organizovat. Toto platí i v rámci akademického pracoviště. Nasazení systému Geonetwork na PřF UK má za cíl nejenom vyhovět potřebě uživatelů po centrálním katalogu pro prostorová data, ale zároveň umožnit studentům samotný systém poznávat a rozvíjet. Software Geonetwork podporuje standardy ISO pro práci s metadaty a tím může být vhodným nástrojem pro studenty při poznávání geoinformačních aplikací, služeb a standardů. Znalost standardů je velmi důležitá v průběhu geoinformačního vzdělávání, ale i v pozdější praxi absolventů. V první části je popsán způsob použití systému na Geografické sekci Přírodovědecké fakulty Univerzity Karlovy po technické stránce. Následuje představení základních případů užití pro organizaci, vyhledávání a získávání příslušných informací o geodatech. Pro správné využití musí být systém nejenom funkční, ale zároveň musí obsahovat i vhodnou dokumentaci nebo tutoriály. Systém Geonetwork opensource poskytuje možnosti sdílet data, ale pro jeho efektivní provoz v rámci akademické instituce je nutná spolupráce a použití ze strany vyučujících i studentů.

Klíčová slova: SDI, výuka, Geonetwork opensource.

Abstract. The paper describes theoretical and practical possibilities of using Geonetwork opensource as a basic tool for spatial data infrastructure (SDI) in education within university environment. There are several reasons why this system was chosen. Geonetwork opensource is primarily a system for administration of metada records. The amount of geodata is still increasing nowadays and is followed by similarly growing needs of users. The paper should discuss whether there exist possibilities to establish fully working SDI system not only for satisfaction all academic users (staff, students, researchers, etc.) but also on the side of the student body with a particular emphasis on student's participation in a development. The Geonetwork opensource has been developed with an intention to show connection between geodata, metadata and their standards. The Geonetwork opensource software come out from ISO standards and it can therefore improve students' understanding of current GIS applications, services and standards. It can help in education because the knowledge of the standards is a very important part of each education process. The first part of the article describes how the Geonetwork has been introduced at our department (Department of Geography, Faculty of Sciences, Charles University in Prague) with regard to its technical implementation. In the second part we present some concrete case studies of how to organize the metadata catalog and request, search, and receive correct set of spatial data for given purposes. In the final part of the paper we shall discuss preparations of tutorials and documentation of this system since Geonetwork enables sharing data but this can be done only in co-operation with all staff members and students.

Keywords: SDI, education, Geonetwork opensource.

1 Introduction

Spatial data infrastructure (SDI) has become a very relevant topic nowadays. Organizations working with spatial data have used this concept in many varying meanings and on different hierarchical levels. International standards for spatial data and related services like catalogue's services provide us with appropriate interface to chain individual organization's infrastructure. The situation in the field of searching or integrating spatial data is very heterogonous and there are international efforts developed for meaningful organizations of the SDI.

Geographic section of the Faculty of sciences is one of the biggest departments at Charles University organizations which use spatial data on daily basis. The geographic section consists of four departments – social and regional geography, physical geography, applied geoinformatics and cartography, demography. All of these units need spatial data for their research and teaching activities. As the world is changing, geographical disciplines have tried to respond to these changes by integrating these changes in their applications. These developments are reflected in employing methods for better description of our world and processes around us. These are often similar or we can clearly find overlap in using same sources, especially in spatial data sources. We can point the collaboration all departments with spatial located information. The localized information should be well organized. The organization of spatial data is a matter of SDI. This paper describes SDI in the institution from the catalog point of view. Technically the Geonetwork opensource software is used for this purpose. The overall conceptions are still being developed (as described in Kolar, et.al. [2]). The metadata catalog is crucial part of SDI and it is sustained by users preferences. Such a catalog is the first interface where present and future users touch the infrastructure. The concept of this catalogue is discussed in greater detail in this paper, especially with a particular emphasis on educational purposes.

2 Spatial data infrastructure for academic institution

2.1 Motivation

The development of the spatial data infrastructure is important in due to the general needs to share spatial data. From the point of the motivation there is demand on knowledge or more precise awareness. It is important to have the knowledge about the existing datasets within faculty, at least for the reasons of minimizing the redundancy. Many spatial data can be already found and stored at the faculty. There exists a big heap of the data between individual person at the faculty during the time (a large number of research projects, grants, etc.). Therefore, the notion of awareness is very important for building SDI. This concept can be divided into several steps or phases (as shown in the table 1). These steps are adapted from Thellufsen [3]. Sometimes is more difficult to induce staff to share their data in co-operative procedure. These steps are sometimes group by overall main processes – motivation, coordination and outcome [3].

Table 1. Phase of awareness with process of SDI building (adapted from [3])

Phase	State	Description
Collaboration	Problem-setting stage	Definition of the needs of potential users. All actors must agree on what exactly the common issue is.
Cooperation	Direction-setting stage	Each owner of the spatial data begins to discuss
Coordination	Structuring stage	The structure of the planned infrastructure is proposed. All individual data owners and users have to agree on a common solution.
Implementation	Problem – solving stage	The project of the common spatial data infrastructure is implemented and tested in the real organization's environment.
Evolution	Relation – maintaining stage	The users assess the resulting outcomes of the SDI. They can come up with new suggestions and influence the future development of the system.

The second reason in decision to create SDI for an academic institution concerns the role of the interactions between the following terms [8] – institution, technology and implementations. In order to understand our proposed perspective we need to distinguish between different points of view. SDI is created for users who are working or studying somewhere (where), SDI is created for certain purposes with a planned solution (what) and it is implemented in a chosen way (how). These three question – where, what and how – are crucial for building an appropriate information system (SDI is not an exception).

It is necessary to start with assessing needs of the institution ('where' question). The SDI concerns the sharing of spatial data. However, requirements for SDI vary for different institutions. For example, the "necessities" of cadastre or municipality are different from universities. Therefore, a particular attention must be paid to specific requirements. These requirements could be formed as two different types - general and specific requirements on SDI. The first type ensures general schema and functions for cataloging systems or services. These functions support storing, managing and querying data in database. The second type (specific needs) then largely depends on institutions. These specific purposes of each SDI can be defined with regards to the type of academic institution – extended catalog for academic outputs (students' maps, semester thesis, models, applications, new script or applications, diploma thesis, etc.), introducing SDI from implementation point of view in the form which is easy to understand or, at least, in a more direct form, easy to extend or adapt application to student's project, etc. Building up SDI represents a very complex application process (for above mentioned reasons). On the other hand, its application can be seen as relatively simple since we can exclude some business-oriented aspects and questions (e.g. business goals, selling or e-shop functionality, time constrains).

As it is shown above, the main motivations for building SDI lie in the principle pivoting on the data sharing. The decision to share all spatial and non-spatial data leads to define problems and questions and to break down the sharing activities from different perspectives. This involves different types of partitioning problems. This can be divided into three parts – institution, technology and implementation (Yeung [8]). These perspectives are shown at the figure 1.

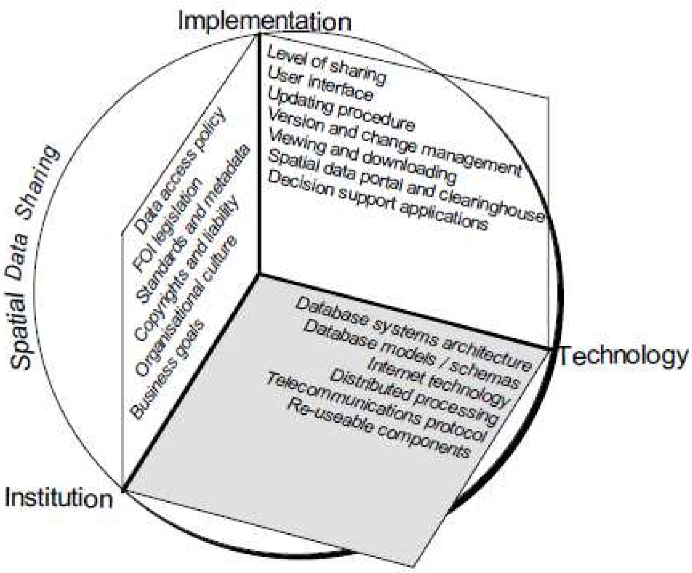

Figure 1: Spatial data sharing from different perspectives [borrowed from 8]

Yeung [8] describes various benefits derived from these different perspectives. Spatial data sharing concerns not only reusing the data among number of users but is also even more so about system. This approach can be very useful in our case. It leads to the increasing influence of more subjects on our curricula and it is, in fact, the second reason to development of the SDI. SDI has to support research (spatial data sharing) and learning activities (knowledge and skills from different subjects are used).

2.2 The key features of the SDI

SDI consists of three main parts - database system, catalog system and visualization system (user interface). These components are not created from one unit of software. The principal software used in our case is shown in the table 2. Geonetwork (catalog and metadata application) as a key tool used in this paper will be discussed later in the text. Let us first consider the others parts of the system.

The basic feature of any SDI is the database. The knowledge of the theory and technology of current databases represents crucial condition for actual learning in the field of the geoinformatics. Only database platforms that can be used together with SDI (also without additional cost for students in lessons) are chosen. The support for spatial data is essential. The PostgreSQL with spatial extension PostGIS was chosen for our SDI concept and its implementation. This combination is very well known and involves very stable and big community. It ensures future development and also easy cooperation with other academic institutions. PostgreSQL is often preferable database at non/commercial organizations.

From our point of view, the visualization of the outputs entails to have one point (starting point or gateway) in which users are able to see the data itself. The appropriate mapping server application fulfills these needs. The ArcGIS server was chosen for this part of the infrastructure. There are two reasons for this choice – software license politic and a level of difficulty involved in the implementation. The software license politic in this case refers to the conformity between our department and ESRI Company. The family of the ESRI software is used in a number of subjects in our educational curricula on GIS. Additionally, students are already familiar with this type of software. The growing number of the people participating in implementation and dissemination of data within the system correspond to the increasing level of potential users' familiarity with the environment. The level of difficulty can be different for this software – students can start with applications development very fast thanks to the "out of the box" principles. Anyone else can use API for these software's functions and write his own programs. Despite the complexity of the ArcGIS server application, webpage for mapping server capabilities was implemented only for its basic functionality (to observe data and to use WMS service). The work on using web processing capabilities of this software is still in progress.

Table 2. Software used in SDI and its license

Application	Usage	License
PostgreSQL	Database platforms	Open source, free
Postgis	Extension of the database	Open source, free
Geonetwork opensource	Application server for metada catalog (Catalog services)	Open source, free
ArcGIS Server 9.3	Web GIS solution or framework (mapping server, WMS, WFS, etc.)	Commercial license (ESRI)
Data interoperability extension	Conversion (transformations) tool for data formats (co-operate with ArcGIS)	Commercial license (ESRI, Safe software)

Geonetwork opensource is a crucial application for infrastructure of spatial data at the Faculty of Sciences. Cataloge services fulfill the requirements for the data. This application is supported by FAO UN organization and its development is coordinated with OGC. Many aspects of Geonetwork and its development are described by Ožana [5]. It ensures that the catalog services are done with using international standards for geographic information. This allows apply Geonetwork in our local environment. The technology aspects of this software will be now introduced.

2.3 Technology aspects

Geonetwork opensource is a web application based on servlet technology. A specific web server is needed in order to launch successfully this application. This web server must support servlets and database system, in which metadata of spatial data and other data essential for running the

Geonetwork opensource application can be stored. Geonetwork opensource is by default installed with the Jetty web server and McKoiDB database system.

Jetty web server and McKoiDB database system are insufficient for the real application of Geonetwork opensource within the SDI. Therefore, the default settings have been reconfigured for Apache Tomcat web server, which supports servlet technology. Database system has been reconfigured to PostgreSQL. Due to the security reasons, the Apache Tomcat web server and the PostgreSQL database system are installed on different servers - Geonetwork opensource with Apache Tomcat web server are installed on application server along with other server applications (such as ArcGIS Server), PostgreSQL is installed on data server. Disclosure of the application server of the extranet is currently being negotiated. Data server will be located in the intranet. It will be the only application to which server connects from the extranet.

The basic configuration of Geonetwork opensource can be done by using GAST, which is a part of the installation process. This simple application can be used for basic settings of applied web server and database system. Advanced settings should be made in the XML configuration files.

Geonetwork opensource supports language localization (by default, the Czech localization is missing in the installation). Geonetwork open source is very modular and to add any further language to the existing languages constitutes a very simple task. All texts are stored in XML files, so you may just copy files with an existing localization, translate and place them in the directory named "cs". If you want to load Czech as the default language of Geonetwork opensource instead of English, it is necessary to set the value of "en" instead of "cs" in the index.html file and config.xml. The Czech localization is currently edited on the basis of existing localization from an earlier version of Geonetwork opensource from colleagues from the Institute of Geoinformatics VSB-TU Ostrava.

Access to Geonetwork opensource is addressed through groups, roles and users. Editor and controller roles can be so very well presumed for the management of metadata. Editor has the rights to create, edit and delete of metadata. Controller has only one, yet more important, task: to inspect and to approve the metadata for publication. Students will have typically editor rights and their metadata records will be checked by teachers.

3 Geonetwork opensource in the education

3.1 The goals for education with Geonetwork opensource

To discuss the curricula in geoinformatics represents a very common activity at most departments around the world. The changes come up with rapid development in this field and lead to regular evaluation of subjects. The idea of using Geonetwork not only as a tool for providing catalog for spatial data but also as a possible tool for students stems from its key position within SDI system. We can look at the system from different point of views. This means that we can classify the ways of usage. The following groups are defined based on:

1. type of users
2. relations to subjects in curricula
3. methods of teaching

In fact, we can find three types of users – normal users (often called only users), contributors (user with higher privileges) and developers. Normal users constitute the majority because all they usually need is only to search, select and obtain or show the data. These users do not have the access to write or modify data or metadata records and can not analyze the functionality of the systems. The set of the available data differs depending on license providing the data. There is no difference made between users – whether these are members of the stuff or students. This question is more related to the agreement with data providers. It is clear, for example, that the access to all census data is more restricted here then to exercise or tutorial data.

Relations to subjects determinate the extent of how much is the SDI connected with subjects. This is a relation similar to the one between heterogeneous data and interoperability capabilities. The subjects can be loosely or tightly coupled with SDI. Loosely coupled subjects concern only matters of theory. This can be mathematics, statistics, methods of acquisition spatial data, remote sensing theory, basics of geography etc. Tightly coupled subjects are the ones which directly describe principles and methods of the SDI within their contents or subjects dealing with knowledge and skills necessary for understanding methods and underlying technologies. Examples of these subjects and relations between them are described later in this chapter.

The third type of usage Geonetwork (SDI) in education depends on teaching methods. One can identify three learning options here: instructor-led lecture, self-study and complex projects. The first two methods have general meaning mainly depends on inputs knowledge of the students. The principles of technology are better to teach with lecturer because of the time constraints. Good understanding is necessary for all students and other stuff. Instructor-led lectures or special introduction courses can overcome the possible difficulties by starting to use Geonetwork or other misunderstanding. This approach goes to more detailed topics where students are encouraged to pick up some problematic issue within the system. The solution has to be found while using self-study but still with a supervision of the lecturer. The most complex and most advanced option is a project type of education. The term "project" means that students are using knowledge from more subjects to find a solution. This study supports group work and cooperation. The typical projects can address questions of implementation of Geonetwork for specific purposes, to use selected functions of the spatial database or to evaluate some web processing services or functions which can be integrated and to create metadata description in Geonetwork. Specific type can be multimedia tutorial (animation, voice, etc.) for new users who are just normal users. These tutorials are aimed especially at the staff members. The crucial thing is to persuade new users (staff or students) that common interface to the spatial data through Geonetwork can fully satisfy their needs.

Geonetwork at Faculty of Sciences is now working only with local databases. There are security reasons for this restriction. Despite this temporary limitation, it has been proposed to use this as one of the geonetwork node. Horáková [1] discussed functionality in context of general metada catalog on the national level. It would be interesting to create net of the Geonetwork's nodes also between different academic institutions. This network would include not only spatial framework data (similar datasets used for teaching like ArcCR database) but also the results of master and bachelor thesis. Although it is possible to find catalogs of the thesis in libraries, there are no possibilities to find any further practical details of the thesis. Assuming that the student captures some specific data or develops new mapping services, it would be highly beneficial to be able to share these outputs. The security and authorization questions have to be carefully considered. The issue of the security plays very important role at all academic departments. Stored data are usually protected and using these is only allowed for education or research non-commercial purposes. The matters of security are moved to spatial database because searching or querying metadata through Geonetwork is unprotected. To show all available datasets might appear as better option, which also ensures more easily evidence and statistic about data within an institution. The metadata records contain the information, which be accessed by the users. The system has to decide between calling from mapserver (or from catalog) to show data (or to direct download). Based on the resulting decision, data are provided in an appropriate form.

The above mentioned points show that Geonetwork can serve for practical as well as for training purposes. In spite of both goals in usage of Geonetwork one has to admit that for operational and a security purpose is better to have more instances installed. The production version and the training (development) version are good tools for ensuring both goals. Since education is a long-term developing process the goals related to Geonetwork's uses in education are evolving at similar pace. The goals can be summarize in following points:

- establish solid state catalog system for SDI and to bring them to the students and other users
- give to the students tool which can be used to disseminate information about their work to the community
- use applications as a tool to introduce international standards and services for spatial data (ISO 19115, ISO 19139, WMS, WFS, WCS, etc.)
- improve students' understanding of open source software for GIS
- improve programming skills (XML based languages, Java)
- show examples of webserver and servlet applications, introduce relations between data in database and applications developed above this databases.

We believe that the Geonetwork can significantly help students in understanding principles behind web and geoinformation services. The benefits and shortcoming of this approach will be reviewed on an annual basis. The other possibility indicating the results of this approach can be found in final students' thesis and their defenses. Better understanding of how the theory works and how it can be implemented will be the great award for these efforts.

3.2 Subjects in curricula and Geonetwork opensource

Geoinformation disciplines have recently developed into a very progressive but very interdisciplinary curriculum. This curriculum needs to incorporate topics from different fields. The concept of spatial data infrastructure and sharing spatial data generally tends to use many aspects of applied informatics. This trend is enlarged by cartographic evolution and modern needs for disseminating spatial related datasets. The disseminations process is moving definitely in the direction of internet environment. All these trends are also reflected in changes in curriculum. We will discuss changes from a common infrastructure point of view. This perspective will not cover all subjects in curriculum in geoinformatics at the Faculty of Sciences (for a full overview see [7]). The SDI (narrow down to Geonetwork) has become important part of many subjects at the department. Selected subjects with tight relations with Geonetwork are shown on the figure 2.

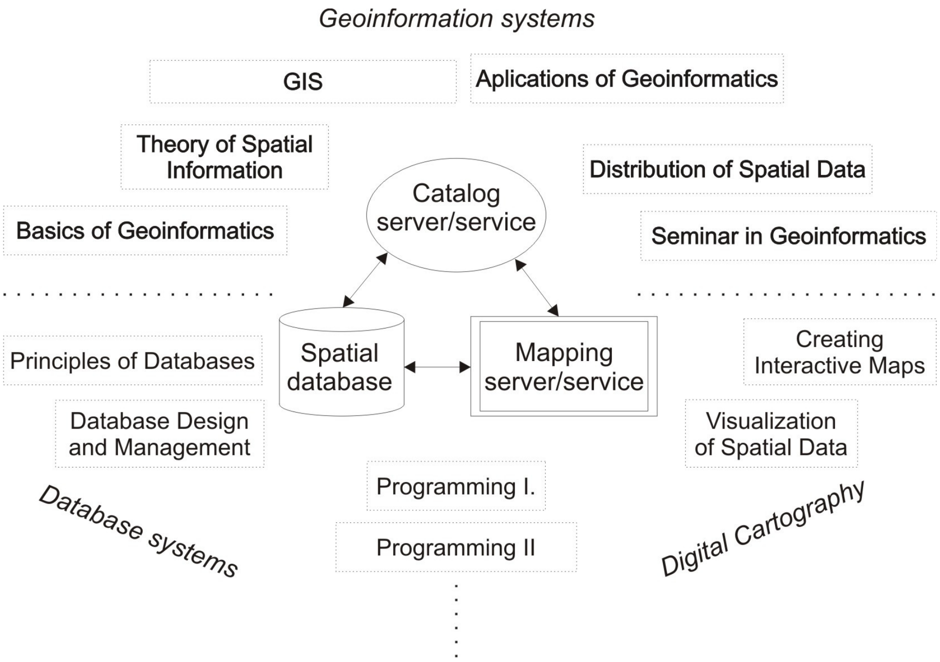

Figure 2. Selected subjects tightly related to Geonetwork (SDI)

These selected subjects can be divided into three branches that can overlap and are separated by porous boundaries.

- Geoinformation systems
- Database systems
- Digital cartography

Groups are defined in accordance to the traditional concept of learning geoinformation science. To put it simply, the professionals dealing with spatial data have to know how to store, how to analyze and how to visualize inputs data. Although this is by far not an exhaustive definition, it is still adequate for incorporating Geonetwork in the education process. On the institutional level, Geonetwork needs some database to store data, some application allowing mapping services and to come up itself with catalog and portrayal service. These requirements are fully satisfactory in Geonetwork. While the above-mentioned subjects are already in the curricula, the Geonetwork integration is still at the very beginning or at the planning stage.

Firstly, the backbone of any GIS systems represents the database. The two subjects – Principles of Databases and Database Design and Management describe databases structure, principles and SQL language. No matter what the choice of the database platform for learning basic of SQL is, the Geonetwork uses PostgreSQL. PostgreSQL is for free and it is stable, which makes it a very appropriate candidate for teaching. It incorporates many examples of various exercises for databases modeling and “SQL queries”. In the final stage of these subjects, the database schema for Geonetwork is shown and described. It is also important to show different views on the metadata record. It is shown as a table or XML tree view within web interface but the distributed storing in database’s tables is shown as well. Students have to find how every piece of information is organized in one ISO metadata records in Geonetwork. It is not only a very good example but also a way of explaining SQL query against the database. Diagrams and database models are the output of the second subjects and should force students to think about spatial data storing and sharing. This exercise shall start with classical non-spatial examples and models (e.g. books libraries) but the examples based on PostgreSQl and PostGIS are next in row.

The groups of the subjects under the umbrella of the Geoinformation systems share a common basis in the approach of extracting new information with a new quality. This extraction is done by analyzing existing data from database. The only one integrated catalog and easy to search possibilities are usually the most appreciated elements by users. Theory of spatial information, Basics of geoinformatics, GIS and Application of Geoinformatics are subjects dealing with usage of data. The skills how to find the best data for projects or tasks is the first. Students can do the best analysis and find meaningful solutions only with the right inputs. This is the part where Geonetwork is very helpful. The second benefit is chance for students to learn about the contents of ISO standards or any metadata profiles based on ISO. The metadata are stored in Geonetwork in that way. The other type of using allows teacher explain and teach about the system itself and also about metadata records creation and editing. These tasks (especially metadata creation) lead to adopt necessary habits or skills by students. They have to know the relations between spatial data and their metadata.

Digital cartography is changing with internet. Geonetwork is running in web environment too. Geonetwork integrates capabilities to show data or results of searching through the map server. Digital cartographic projects use very often similar web technology as the mapping server applications. The introduction about web programming techniques is added to these subjects. The students need to learn web server, server side and client side programming tools which can be used for building high quality and dynamic cartographic products on internet. The knowledge of Geonetwork opensource together with map servers' technology can be very helpful.

4 Conclusion

Application Geonetwork opensource has proved to be a very useful tool for direct support research and education activities at the Faculty of Science. The usage and implementation of this application has been naturally accompanied with some difficulties. It requires spending more time with tutorial preparations and also the supervisions for its users. It was find out that the members of stuff without enormous geoinformation knowledge can use it. These teachers and researchers from many geographic disciplines may use spatial data and they do not need to build their individual data stores. Geonetwork makes it easy principally in sharing data. Students benefits from Geonetwork in the same manner but in addition to that they can learn a lot about technology, theory and standards behind this system. Despite encountering some difficulties, it still comes up with more benefits to offer.

Reference

1. Horáková, B., Růžička, J., Ožana, R. Development of MetaPortal Prototype and Communication Interface for Czech national environment. GIS Ostrava 2007. ISSN: 1213239X.
2. Kolář J., kol. Infrastruktura prostorových dat geografické sekce Přírodovědecké fakulty UK v Praze. GIS Ostrava 2007. Ostrava 2007. ISSN: 1213239X.
3. Thellufsen, C., et al. Awareness as a foundation for developing effective spatial data infrastructures. Land Use Policy (2008). [Online] 2008. http://dx.doi.org/10.1016/j.landusepol.2008.03.002
4. Open Geospatial Consortium. [Online] 2008. http://www.opengeospatial.org/.
5. Ožana R., Horáková B. Actual State in developing GeoNetwork opensource and metadata network standardization. GIS Ostrava 2008. Ostrava 2008.
6. PostgreSQL. [Online] 2008. http://www.postgresql.org/.
7. Potůčková M., kol. Curricula in Geoinformatics at the Faculty of Science, Charles University in Prague. FCE Geoinformatics 2007, 2007 Prague.
8. Yeung K.W. A., Hall G. B. Spatial Database Systems: Design, Implementation and Project. Springer, 2007 Dordrecht, The Netherlands. ISBN 10 1-4020-5392-4.
9. GeoNetwork opensource community website. [Online] 2008. http://geonetwork-opensource.org/.