UDK 528:004.6:004.822:81'37:004.738.5
Pregledni znanstveni članak

Metadata Catalogues in Spatial Information Systems

Miro GOVEDARICA, Dubravka BOŠKOVIĆ, Dušan PETROVAČKI, Toša NINKOV, Aleksandar RISTIĆ – Novi Sad[1]

ABSTRACT. This paper gives the short review of the Open Geospatial Consortium (OGC) metadata catalogue services that have the key role in geospatial resource discovery in Spatial Data Infrastructures (SDI). The notion of Spatial Data Infrastructure comprises a collection of technologies, policies and institutional agreements that provide an easier access to geospatial data. The SDI is suitable for usage in geospatial data discovery, evaluation, and also various applications within government, commercial and non-profit sectors, academic institutions, etc. Metadata catalogue services have been specified in OGC Catalogue Service Implementation Specification. The part of the specification that specifies a web interface that supports the storage, retrieval, and management of data related to web services, is called Catalogue Service for the Web (CSW). Metadata catalogues are service brokers that represent a key component in a service-oriented architecture that manages shared resources and facilitates the discovery of resources within an open, distributed system. OGC services have gained significant popularity in recent years and the number of organizations using them has increased. However, the full potential of metadata catalogues has not yet been reached, not only because of the lack of appropriate documentation of data in the form of standardized metadata, but because the lack of semantics of the data. The analysis of the usage of metadata catalogue services in geodetic information systems has been given and the proposal for a possible solution for improvement has been made.

Keywords: OGC Catalogue, CSW, spatial information systems, metadata, semantics.

[1] Prof. dr. Miro Govedarica, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, RS-21000 Novi Sad, Serbia, e-mail: miro@uns.ac.rs,

Dubravka Bošković, MSc, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, RS-21000 Novi Sad, Serbia, e-mail: dudab@uns.ac.rs,

Prof. emer. dr. Dušan Petrovački, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, RS-21000 Novi Sad, Serbia, e-mail: petrovacki@uns.ac.rs,

Prof. dr. Toša Ninkov, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, RS-21000 Novi Sad, Serbia, e-mail: ninkov.tosa@gmail.com,

Doc. dr. Aleksandar Ristić, Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, RS-21000 Novi Sad, Serbia, e-mail: aristic@uns.ac.rs.

1. Introduction

A service-oriented architecture (SOA) (Erl 2005) is the distributed computing architecture based on loosely coupled interactions of geo-services in which the service interaction model illustrates the interaction between different agents for publishing, discovering, and invoking geo-services, so called "publish-find-bind" model (Fig. 1). This model involves: publishing resource descriptions so that they are accessible to prospective users (*publish*), discovering resources of interest according to some set of search criteria (*find*) and interacting with the resource provider to access the desired resources (*bind*). Within such architecture a registry service plays the essential role of matchmaker by providing publication and search functionality, thereby enabling a requester to dynamically discover and communicate with a suitable resource provider without requiring the requester to have advance knowledge about the provider. Benefits of using SOA is that the monolith software applications are replaced by a set of loosely coupled services which can be reused and combined in various application domains. Those services comply with the standards so their users are not vendor-dependent.

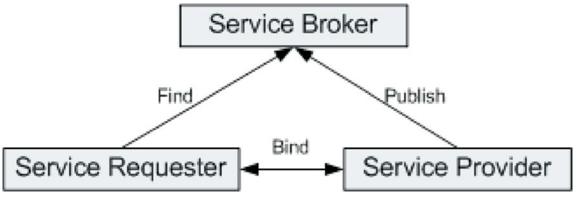

Fig. 1. *"Publish-Find-Bind" service interaction model.*

Service oriented architecture of Geographic Information Systems (GIS) is based on services for geospatial data discovery, access, visualization and processing that implement the OpenGIS Consortium (OGC) (URL 1) specifications and are the building blocks for the development of the Spatial Data Infrastructure (Nebert 2004). The notion of Spatial Data Infrastructure (SDI) comprises a collection of technologies, policies and institutional agreements that facilitate access to geospatial resources. Resource discovery through catalogue services is the key element for the development of any SDI (Nogueras-Iso et al. 2005). Catalogue services are implemented as part of the geoportals which are web portals where geographic content can be discovered (Maguire and Longley 2005, Tait 2005). They provide an access to SDI.

Services that support geospatial data access and visualization include Web Map Service (WMS), Web Feature Service (WFS) and Web Coverage Service (WCS) and have been described in (Bulatović et al. 2010), together with the benefits of using those services in modern service-oriented architecture of GIS. However, the key elements in such architecture are registries of geospatial resource descriptions that allow users to find out about the existence of other geospatial services they can use in their applications. Those services that support geospatial resource discovery implement OGC Catalogue Service (URL 2) and will be presented in the rest of the paper.

Resource descriptions are called metadata, which means data about data. Metadata are registered in the catalogues. Metadata is used to describe geospatial resources (data and/or services). Its purpose is to enable geospatial resource discovery, its evaluation and to provide information how to access and use that resource. Therefore metadata can be divided in three categories: discovery metadata, exploration metadata and exploitation metadata. *Discovery Metadata* is the minimum amount of information that needs to be provided to reveal to the user the content of the resource. This kind of metadata answers the "what, why when who, where and how" questions about geospatial resource. *Exploration metadata* provides sufficient information to determine that a resource that fit for a given purpose exists, to evaluate its properties, and to reference some point of contact for more information. Exploration metadata include information required to allow the user know whether the data will meet general requirements of a given problem. *Exploitation* metadata include those information required to access, transfer, load, interpret, and apply the data in the end application where it is exploited.

This paper is organized as follows: the next Section presents the basic concepts of the OGC Catalogue Service Specification including information models, catalogue interfaces together with examples, application profiles and catalogue implementations. Section 3 describes the usage of metadata catalogues in geodetic-cadastral information systems, lists some of the problems and possible solutions. Finally, conclusions are given in Section 4.

2. OGC Catalogue Service Specification

Since the geospatial community is a wide community and comprises both tightly coupled systems implementing functions in a tightly controlled environment, and various Web based services knowing nothing about the clients, OGC has developed an abstract model of a metadata catalogue that is called *General Catalogue Model* which can be applied in various catalogue operating environments. General catalogue model comprises *Catalogue Abstract Information Model* and *General Catalogue Interface Model* which will be described in the following.

2.1. Catalogue Abstract Information Model

Catalogue Abstract Information Model specifies the minimal query language, a set of core query able attributes (names, definitions, datatypes) and a common record format that defines the minimal set of elements that should be returned in a query, called core returnable properties.

The notion of core catalogue schema is related to Catalogue Abstract Information Model. For the purpose of sharing information within the information community, it is necessary to define the metadata schema that provides a common vocabulary to support search, retrieval and an association between the descriptions and resources. Core catalogue schema comprises core queryable properties and core returnable properties. Its purpose is to enable interoperability among various catalogue implementations which can differ in metadata schemas and protocols used, where the same queries can be executed against any catalogue service without modification and without detailed knowledge of the catalogue's information model. In that way metadata returned from different systems can be used by a single client. This requires a set of general metadata properties that can be used to characterize any resource. The core metadata is expressed using the syntax of Dublin Core Metadata, ISO 15836 (URL 3).

Depending on the needs, developers of metadata catalogues may choose to implement various metadata models to extend Catalogue Abstract Information Model. Most common are ISO 19115 (URL 4) and OASIS ebXML Registry Information Models (URL 5). In the remainder of this section four information models that can be implemented in OGC catalogues will be described.

Dublin Core

The Dublin Core is a metadata element set intended to enable discovery of electronic resources. It is primarily used for author-generated description of Web resources, but it is also used in communities such as museums, libraries, government agencies, and commercial organizations. Dublin Core metadata is specifically intended to support general-purpose resource discovery. The elements represent concepts of core elements that are likely to be useful to support resource discovery. It uses only fifteen base text fields, which are usually inadequate for even basic geospatial resource description and discovery, because there is no mean to declare what type of content is present in the text element (coordinates, date or time, place name, etc). Therefore, a more detailed metadata model is needed to support the discovery of geospatial resources.

ISO 19115 Geographic Information – Metadata

ISO 19100 series of standards is a set of standards that define the basic structure and semantics of geographic data and services. This series of standards also defines the standard for geospatial metadata: ISO 19115 Geographic Information – Metadata, so that users that are searching for the appropriate geospatial data can find them, estimate whether they fit their needs, and to access, transfer and use them. Accompanying standard ISO 19139 – Metadata – Implementation specification (URL 6), defines the Extensible Markup Language (XML) (URL 7) schema for storing metadata. Because XML has become the industry standard for data storage and transmission over the Internet, it is necessary to define an XML schema for describing geospatial data.

ISO 19115 classifies metadata standards into eleven categories and they all belong to the Metadata entity set information. These categories are:

- *Identification information* – provides basic information about the geospatial data set for the purpose of the data identification;
- *Constraint information* – describes how the rights for the access and use of metadata and geospatial data are regulated;
- *Data quality information* – defines metadata about the quality of geospatial data and possibilities of their application with regard to quality;
- *Maintenance information* – describes the maintenance and updating of metadata and geospatial data;

- *Spatial representation information* – describes the mechanism used to represent spatial information;
- *Reference system information* – a description of the used spatial and temporal reference systems;
- *Content information* – describes the content of the dataset and the feature catalogue used;
- *Portrayal catalogue information* – reference to the catalogue that contains methods for the portrayal of geospatial data;
- *Distribution information* – information about access to geospatial data, methods of distribution, and the people responsible for the distribution;
- *Metadata extension information* – describe the structure of user extensions of metadata;
- *Application schema information* – describes the application schema of the data (the conceptual data model, usually expressed in UML).

OASIS ebXML Registry Information Model (ebRIM)

Electronic Business using eXtensible Markup Language, known as e-business XML, or ebXML, is a family of XML based standards whose goal is to provide an open, XML-based infrastructure that enables the global use of electronic business information in an interoperable, secure, and consistent manner by all partners. The ebXML Registry allows businesses to find one another, to define trading-partner agreements, and to exchange XML messages in support of business operations. The goal is to allow all these activities to be performed automatically, without human intervention, over the Internet.

This catalogue information model is based on version 3.0 of the OASIS ebXML Registry Information Model (ebRIM 3.0). This model specifies how catalogue content is structured and interrelated; it constitutes a public schema for discovery and publication purposes. An ebXML Registry is capable of storing any type of electronic content such as XML documents, text documents, images, sounds and videos. The ebRIM uses several standard classification schemes as a mechanism to provide extensible enumeration types which are used to create classifications or ontologies for the catalogue content.

The ebRIM information model is a general and flexible one with several extensibility points. A set of extensions that address the needs of a particular application domain or community of practice may be defined. The ebRIM is more generic and flexible than ISO 19115 and may contain various contents which is not specifically indented for geospatial data, and in that way the relationship between GIS and non-GIS systems is provided.

Web Ontology Language

This information model is based on a semantic markup language for describing ontologies. Ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts (Staab and Studer 2009). It is used to reason about the properties of that domain, and may be used to describe the domain. Their role is to provide a shared vocabulary within a certain domain such as, for example the land administration.

OWL (URL 9) is a standard for ontology on the Semantic Web from the World Wide Web Consortium (W3C) (URL 10). It is built on top of RDF (Resource Description Frame) (URL 11) and RDF Schema (URL 12), a family of specifications for description of web resources. OWL ontologies may be categorized into three species or sub-languages: OWL-Lite, OWL-DL and OWL-Full. A defining feature of each sub-language is its expressiveness. OWL-Lite is the least expressive sub-language. OWL-Full is the most expressive sub-language, but does not guarantee computational completeness or decidability. The expressiveness of OWL-DL falls between that of OWL-Lite and OWL-Full. OWL-DL is much more expressive than OWL-Lite and is based on Description Logics (Baader et al. 2002), hence the suffix DL, and allows expressivity without losing computational completeness and decidability for reasoning.

The information model has two components: one component models RDF and RDF Schema and another the OWL-DL. OWL-DL is used to ensure computational resolvability. The main benefit of this information model over others is that meaning of the concepts are made explicit, so semantic interoperability may also be achieved in addition to syntactic interoperability provided by OGC standards. Ontology may be used to automatically reason about the properties of a domain, and may be used to describe that domain. Its role is to provide a shared vocabulary within a certain domain and therefore avoid semantic disambiguates.

2.2. General Catalogue Interface Model

General Catalogue Interface Model provides a set of abstract service interfaces to support retrieval, access, maintenance and organization of catalogs of geospatial information and associated resources. Specified interfaces allow users or applications to find information that exists in multiple distributed computing environments, including the World Wide Web (WWW).

The reference architecture for the development of the OGC Catalogue Interface is shown in Fig. 2. It is a multi-tier architecture consisting of clients and servers. Client application communicates via an interface to the catalogue service using the OGC catalogue interface. Service catalogs can communicate to one of the three sources in order to respond to the request it received from a client: the metadata repository that is local to the service catalog, resource service, or other catalogue service. The interface to the local repository metadata is internal to the

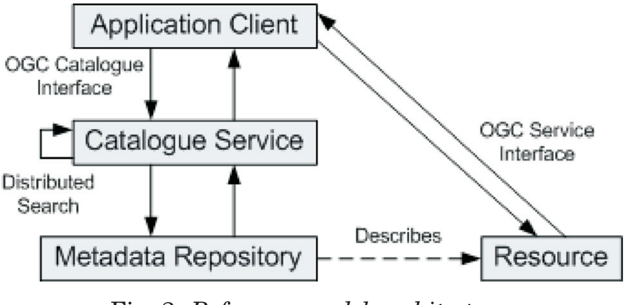

Fig. 2. *Reference model architecture.*

catalogue service. The interface to the resource service can be private or some OGC interface, such as WMS or WFS. The interface between the catalogue services is the OGC catalogue interface. In this case, the catalogue service acts as a client and a server. Data returned from the OGC catalogue service is handled by the catalogue service that sent the request and it returns them to the original request. In this way distributed search is accomplished.

General interfaces can be bound to several application protocols. OGC Catalogue Service specification provides the possibility of implementing a catalogue service using one of the following application protocols:

- HTTP protocol (URL 13) binding – involves mapping the operations of a General model to the message requests and responses that are common to all web-based catalogue services. When the catalogue service implements HTTP protocol binding it is called Catalogue Services for the Web (CSW) and it will be given in the following section in more detail.
- Z39.50 protocol binding – it uses client-server architecture based on the messages implemented using ANSI / NISO Z39.50 Application Service Definition and Protocol Specification (URL 14) for searching and retrieving information from remote computer databases.
- CORBA/ IIOP protocol binding – CORBA (Common Object Request Broker Architecture) (URL 15) is a standard defined by the OMG (Object Management Group) (URL 16) that enables software components written in different programming languages and that run on different computers to work together.

2.2.1. HTTP protocol binding (Catalogue Services for the Web, CSW)

In HTTP protocol binding, the interaction between the client and the server is accomplished using the standard request-response model of HTTP protocol. This means that the client, such as web browser, submits an HTTP request message to the server using the HTTP protocol, and expects to receive a response message from the server. The basic message exchange pattern is illustrated in Fig. 3. The server stores content or provides resources, which it delivers to the client.

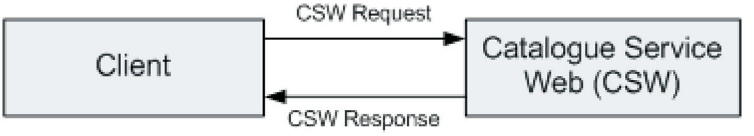

Fig. 3. *Catalogue Service for the Web.*

There are two ways for encoding request and response messages of the CSW. First, they can be encoded as pairs of keywords (request parameters) and values within the URL address of the server. This method is called Keyword-Value Pairs (KVP). Secondly, they can be encoded using XML, an industry standard for the exchange of data on Internet. CSW client's requests can also be included in a framework based on messages such as Simple Object Access Protocol (SOAP), which is a protocol specification for exchanging structured information in the implementation of web services in computer networks and relies on XML for its message format and other application protocols, such as HTTP, for message negotiation and transmission.

HTTP protocol supports several methods for submitting requests to the server. HTTP protocol binding for the CSW only uses GET and POST methods. Former means that parameters and data that are sent are encoded into a URL, and is basically for just getting (retrieving) data, while the latter means that the parameters and data are sent within a message body and may involve anything, such as storing or updating data. Table 1 lists CSW requests and their allowed methods in HTTP protocol binding (GET or POST), as well as method of data encoding (KVP or XML). These requests will be described in detail in the following.

Table 1. *CSW requests and HTTP method bindings.*

Request	**HTTP method bindings**	**Data encodings**
GetCapabilities	GET (POST)	KVP (XML)
DescribeRecord	POST (GET)	XML (KVP)
GetDomain	POST (GET)	XML (KVP)
GetRecords	POST (GET)	XML (KVP)
GetRecordById	GET (POST)	KVP (XML)
Harvest	POST (GET)	XML (KVP)
Transaction	POST	XML

The *GetCapabilities* is a mandatory operation that allows CSW clients to receive service metadata from a server. The response to a *GetCapabilities* request is an XML document that contains service metadata about the CSW server. This XML document describes the capabilities of the CSW service.

The *DescribeRecord* is a mandatory operation that allows CSW clients to find the elements of the information model supported by the target catalogue service. This operation allows the individual parts or the entire information model (metadata schema) to be described. The response to this request is XML metadata schema that describes the appropriate record that is requested.

The optional *GetDomain* operation is used to obtain runtime information about the range of values of a metadata record element or request parameter. The runtime range of values for a property or request parameter is usually much smaller than the value space for that property or parameter based on its type definition. This type of runtime information about the range of values of a property or request parameter may be useful for generating user interfaces with meaningful lists from which the user can select one or more values. This operation does not always give accurate results.

The mandatory *GetRecords* operation is used for searching and presenting information from the catalogue. The searching part of the *GetRecords* operation is encoded using the *Query* element. The *Query* element includes the parameters that specify which entities from the information model of the catalogue are queried, and may also specify which query constraints shall be applied to identify the request set. It is specified using OGC Filter specification (URL 17). The presenting part of the *GetRecords* indicates which schema i.e. information model (ISO 19115, ebRIM...) is used to generate the response to the *GetRecords* operation and which properties of that schema should be included in each record in the *GetRecords* response.

The mandatory *GetRecordById* request retrieves the catalogue records using their identifier. In order for this operation to be performed a previous query has to be performed in order to obtain the identifiers that may be used with this operation. For example, records returned by a *GetRecords* operation may contain references to other records (its identifier) in the catalogue that may be retrieved using the *GetRecordById* operation. This operation is a subset of the *GetRecords* operation and is suitable for retrieving and linking to records in a catalogue.

There are two optional operations that may be used to insert, delete or update records in the catalogue: *Transaction* and *Harvest*. The *Transaction* operation is used to “push” data into the catalogue whereas the *Harvest* operation “pulls” data into the catalogue. That is, this operation only references the data to be inserted or updated in the catalogue, and the catalogue service should resolve the reference, fetch that data, and process it into the catalogue, immediately or later depending of the mode of operation (synchronous and asynchronous).

Fig. 4 shows a conceptual architecture that illustrates the relationship of CSW interfaces to metadata consumers and producers. The arrows show the CSW requests that producers and consumers of metadata can generate. For example, in order to create metadata, metadata producer may invoke *Transaction* or *Harvest* request. Similarly, the user of metadata may invoke *GetRecords* request to perform queries on the catalog.

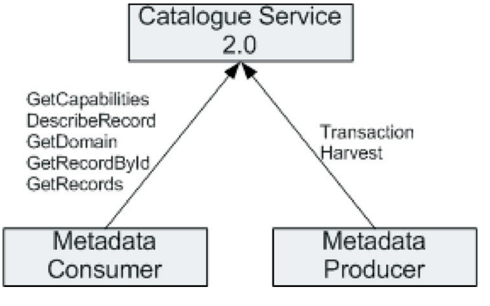

Fig. 4. *CSW conceptual architecture.*

The following examples illustrate *GetCapabilities*, *DescribeRecord* and *GetRecords* requests. In each request the part before the symbol “?” represents the address of the service, whereas the other part represents parameters of the service expressed as keyword-value pairs. The order of parameters is arbitrary.

The *GetCapabilities* request example is as follows:

```
http://localhost:8080/deegree-csw/services?REQUEST=GetCapabilities&version=2.0.2&
service=CSW
```

Listing 1. *The GetCapabilities request.*

Parameters of this request are:
- *REQUEST* – the type of request to the CSW, in this case *GetCapabilities*, it is a mandatory paremeter;
- *service* – the type of service, in this case CSW, it is a mandatory paremeter;
- *version* – the version of the service, in this case it is the CSW service version 2.0.2, it is a mandatory paremeter.

The *DescribeRecord* request example is as follows:

```
http://localhost:8080/deegree-csw/services?REQUEST=DescribeRecord&version=2.0.2&
service=CSW&TypeName=gmd:MD_Metadata
```

Listing 2. *The DescribeRecord request.*

Parameters of this request are the same as the parameters of the *GetCapabilities* request with the additional parameter:
- *TypeName* – a list of type names that are to be described by the catalogue. It is a mandatory paremeter. In this example, the request demends a whole metadata schema according to the ISO 19115 / ISO 19139 standard.

The KVP *GetRecords* request example is as follows:

```
http://localhost:8080/deegree-csw/services?REQUEST=GetRecords&version=2.0.2&
service=CSW&TypeNames=gmd:MD_Metadata
```

Listing 3. *The KVP GetRecords request.*

Parameters of this request are the same as the parameters of the *GetCapabilities* request with the additional parameter:
- *typeNames* – a list of one or more names of entities, from the information model of the catalogue, that will be queried. It is a mandatory paremeter. In this example, the request demends all metadata records from the catalogue.

The XML *GetRecords* request is as follows:

```
<?xml version="1.0" encoding="UTF-8" ?>
<csw:GetRecords service="CSW" version="2.0.2" outputFormat="application/xml"
 outputSchema="http://www.isotc211.org/2005/gmd" resultType="RESULTS"
 xmlns:csw="http://www.opengis.net/cat/csw/2.0.2"
 xmlns:ogc="http://www.opengis.net/ogc"
 xmlns:apiso="http://www.opengis.net/cat/csw/apiso/1.0">
 <csw:Query typeNames="gmd:MD_Metadata">
     <csw:ElementSetName>full</csw:ElementSetName>
     <csw:Constraint version="1.1.0">
       <ogc:Filter>
         <ogc:PropertyIsEqualTo>
           <ogc:PropertyName>apiso:title</ogc:PropertyName>
           <ogc:Literal>Cadastral municipalities</ogc:Literal>
         </ogc:PropertyIsEqualTo>
```

Listing 4. *The XML GetRecords request.*

This example is similar to the previous one except it has a constraint (<csw:Constraint>) defined using OGC filter encoding (<ogc:Filter>) which specifies that the title of the data set must be equal to “Cadastral municipalities” (<ogc:PropertyIsEqualTo>). Therefore, this request does not return all the metadata records as the previous one, but only those metadata records that contain the title equal to “Cadastral municipalities”.

2.3. Application Profiles and Catalogue Implementations

The notion of application profile has been defined in ISO TR 10000-1:1998 (URL 18) which specifies a general framework for functional standardization. An application profile may be developed from one or more base specifications in order to address particular needs or requirements. It should also specify conformance tests to check compliance to the base specifications, because an application profile must not contradict the base specifications. An application profile also specifies the use of an application-layer protocol such as HTTP in order to provide the transfer of information between systems.

The general OGC catalogue model defines common behaviors and interfaces that have general utility, but in practice there is no single solution that will satisfy everyone’s needs. Therefore, OGC has developed several application profiles of the general OGC catalogue model in order to satisfy specific implementation communities. These communities use application profiles as standards for conformance. The general OGC catalogue model is a platform-neutral specification, whereas application profiles are platform-specific. They are bound to a particular distributed computing protocol, in this case they use one of the protocol bindings defined in the catalogue specification. The relationships between base specifications, application profiles, and catalogue service implementations are illustrated in Fig. 5. The platform-neutral specification is one of the base specifications with which the application profile complies, and a given catalogue implementation conforms to one or more application profiles.

Fig. 5. *Relationships between base specifications, profiles, and implementations.*

Catalogue implementations specialize the general OGC catalogue model through protocol bindings and application profiles. They are constrained by the protocol bindings of the catalogue specification. Each protocol binding includes a mapping from the general interfaces, operations, and parameters specified in the general OGC catalogue model to the constructs available in a chosen protocol. Application profiles further specialize the implementation of these interfaces and their operations, including adding classes and parameters. However, they are specializations of parental protocol bindings and the names of the operations and parameters cannot be changed. An application profile is based on one of the protocol bindings in the base specification and in the case of the Catalogue Services Specification it may be CORBA/IIOP, Z39.50, or the HTTP protocol binding. The relationship of general model, protocol binding and application profile is shown in Fig. 6.

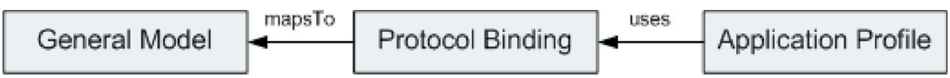

Fig. 6. *Relationship of general model, protocol binding, and application profile.*

OGC has developed several application profiles including:

- OpenGIS Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile (1.0.0) (URL 19) – this application profile implements ISO 19115/ISO 19139 information model
- CSW-ebRIM Registry Service – Part 1: ebRIM profile of CSW (1.0.1) (URL 20) – this application profile implements ebXML Registry Information Model
- Catalogue Services Standard 2.0 Extension Package for ebRIM Application Profile: Earth Observation Products (URL 21) – it describes the mapping of Earth Observation products, defined in the OGC GML 3.1.1 Application schema for Earth Observation products (URL 22), to an ebRIM structure within an OGC Catalogue 2.0. implementing the CSW-ebRIM Registry Service – part 1: ebRIM profile of CSW. This standard defines the way Earth Observation products metadata resources are organized and implemented in the catalogue for discovery, retrieval and management.
- OGC Catalogue Services – OWL Application Profile of CSW (0.3.0) (URL 23) – this application profile is proposed in the discussion paper and has not yet become the implementation standard. It uses OWL as information model.

Bai et al. (2009) proposed geospatial service taxonomies to represent the knowledge about the characteristics of geospatial services with the aim of promoting the global sharing of and interoperability among geospatial service instances. It is a hierarchical taxonomy consisting of six layers: service category, service type, version, profile, binding and uniform resource name. In this way the problem of various application profiles may be solved and increased interoperability achieved.

There are several proprietary and open source solutions that implement some of the OGC catalogue application profiles. They also provide the ability to create metadata for the geospatial resources according to various information models:

- *GeoNetwork opensource* (URL 24) is a web based geographic metadata catalogue application. It implements the ISO19115/19139 Geographic Metadata, Z39.50, CSW 2.0.2 and OGC WMS standards among others.
- *Deegree Web Catalogue Service* (CSW) (URL 25) is a software package that implements the OGC Catalogue Service Implementation Specification 2.0.2 and ISO 19115/19119 Application Profile 1.0.0.
- *ESRI ArcCatalog* (URL 26) is part of ArcGIS development environment for geographic information systems. ArcCatalog is used for cataloging all GIS resources within an organization and provides basic information about each of them. It also allows the creation and update of metadata for each GIS resource according to ISO 19115.
- *ERDAS APOLLO Catalog* (URL 27) offers a CSW compliant view on the content of the ERDAS APOLLO Catalog. The preferred OGC registry information model is based on the ebXML registry information model, ebRIM Application Profile for CSW.

3. Using OGC Catalogue Services in Geodetic-Cadastral Information Systems

OGC services are becoming more and more popular in organizations that are dealing with spatial data, such as real estate cadastre, urban planning, environmental protection agencies, agriculture stations etc. These organizations recognized the benefit of OGC open architecture and seamless data distribution and integration and many are in the process of implementing services such as OGC WMS, WFS and WCS. However, the usefulness of catalogue services has not yet been entirely recognized. Many find describing their geospatial resources with metadata rather tedious and time consuming task. Therefore, catalogues often contain only basic metadata, such as name or spatial extent, which is not very informative in resource discovery, especially for external users. Therefore it is necessary to introduce documenting data into organization's practice. The other problem of using OGC catalogue services in spatial information systems such as Geodetic-cadastral information system is the problem of the semantics of data. In order to effectively use metadata catalogues, users need to see the details of the underlying data model.

3.1. The Role of Metadata Catalogues in Geodetic-Cadastral Information System

According to the Law on State Survey and Cadastre (URL 28) Geodetic-cadastral information system consists of subsystems that contain data and services for the basic geodetic works, real estate cadastre, the address register, register of spatial units, the register of geographical names, network utility cadastre, topographic-cartographic data and other data. It provides data collection, management, maintenance, access and use of sub-systems, on the principles of:

- keeping data in one referent place and facilitate access and exchange at various levels within an organization;
- creating and maintaining metadata about data and services to facilitate its discovery;
- defining data and services in compliance with national and international standards;
- allowing the users outside the organization to access information via web services;
- facilitating issuance of documents in electronic or other form;
- security of electronic business transactions for the use of data and services that require a fee;
- providing safeguards for information system in all phases of operation.

Digital geodetic plan is a spatial information system which consists of four basic components: data, software, hardware and users who provide the collection, processing, maintenance, analysis and distribution of content. It is a subsystem of geodetic information system, and can be implemented as a separate system. Digital geodetic plan consists of the following themes: geodetic reference, cadastral parcels, parts of cadastral parcels according to land use, buildings, names and textual descriptions, network utilities, spatial units, elevation model of terrain, topography, unclassified. Digital cadastral plan, Digital network utility plan and Digital topographic plan consist of the subset of the themes from Digital geodetic plan. Data are distributed using standard vector and raster formats. Metadata descriptions are necessary to facilitate data distribution.

In the following, examples of metadata descriptions of cadastral plan in vector and raster format according to ISO 19115 standard will be shown. Metadata for a cadastral plan in vector format may include a description of the entire dataset or individual feature types or attributes. Metadata information about vector data will include several categories of information. Information about metadata includes name of the metadata set, language of metadata, metadata author, etc. Identification information includes title of the dataset, abstract, purpose, status, point of contact, category of data, keywords, etc. It is also necessary to document update frequency and maintenance of data, legal constraints, spatial resolution – scale and spatial and temporal extent. Data quality information contains information about data lineage and its completeness, logical consistency, thematic, positional and temporal accuracy. Metadata about vector data may also include reference to the application schema i.e. data model, feature catalogue that lists all the features in the dataset and portrayal catalogue which contains symbols for visualization of features in dataset and is used for data visualization on the maps. Listing 5 contains an extract of metadata description for cadastral plan represented as vector. Spatial representation info section contains information relevant for the vector spatial representation. Reference system info section contains an identifier of the coordinate reference system or its parameters. Identification info section contains the type of spatial representation, in this case vector, geographic bounding box, etc. Distribution section contains information about data formats for the distribution, online resources such as WMS, etc.

```
MD_Metadata
...
+spatialRepresentationInfo (digital representation of spatial data)
        MD_VectorSpatialRepresentation
        topologyLevel: 001 (code for data that contains only geometry,
        without topology)
        geometricObjects:
                MD_GeometricObjects
                geometricObjectType: 001 (complexes)
+referenceSystemInfo (description of spatial reference system)
        MD_CRS (Coordinate Reference System)
                RS_Identifier
                code: EPSG 31277
+identificationInfo (basic information about data)
        MD_DataIdentification
        …
        spatialRepresentationType: 001 (code for vector data)
        geographicBox:
                EX_GeographicBoundingBox (bouding box of data in
                degrees)
                westBoundingLongitude: 45.269
                eastBoundingLongitude: 45.283
                southBoundLatitude: 19.876
                northBoundLatitude: 19.905
                ...
+distributionInfo (distribution of data)
        MD_Distribution
        ...
                +distributorFormat:
                        MD_Format
```

Listing 5. *Metadata description of vector data.*

The similar metadata description may be given for raster data. The main difference is the spatial representation information which includes grid spatial representation and can be divided in georectified and georeferencable grid. Georectified grid is a grid whose cells are regularly spaced in a geographic or map coordinate system defined in the spatial referencing system so that any cell in the grid can be geolocated given its grid coordinate and the grid origin, cell spacing, and orientation. Georeferencable grid is a grid with cells irregularly spaced in any given geographic or map projection coordinate system, whose individual cells can be geolocated using geolocation information supplied with the data but cannot be geolocated from the grid properties alone. Spatial representation information includes properties of the grid such as number of dimensions, axis properties, cell geometry, availability of check points and transformation parameters, etc. The following listing shows extract of metadata for raster data in ISO 19115 format. Information about reference system, bounding box and distribution are similar as in previous example. The difference can be observed for spatial representation information which gives details of the grid.

This sort of information enables discovery and retrieval of data according to title, abstract, keywords, spatial and temporal extent, categories, themes, etc. It answers the “what, where, when, why, who, and how” questions about geospatial

```
MD_Metadata
...
+spatialRepresentationInfo (digital representation of spatial data)
       MD_Georectefied (descendent of MD_GridSpatialRepresentation)
       numberOfDimensions: 2
       axisDimensionsProperties:
              MD_Dimension
              dimensionName: 001 (code for the ordinate y)
              dimensionSize: 9449 (cell- pixel)
              resolution:
                     Measure (documented in ISO 19103)
              dimensionName: 002 (code for the abscissa x)
              dimensionSize: 14173
              resolution:
                     Measure (documented in ISO 19103)
       cellGeometry:002 (cell represent area)
       transformationParameterAvailability: 1 (transformation
       parameters are available)
       checkPointAvailability: 0 (check points for testing accuracy of
       georeferenced data)
       cornerPoints: (geographical coordinates of the corners of an
       image)
              GM_Point (documented in ISO 19107)
       pointInPixel: 001 (center)
...
+identificationInfo (basic information about data)
       MD_DataIdentification
              …
       spatialRepresentationType: 002 (code for grid data)
```

Listing 6. *Metadata description of raster data.*

resources. ISO 19115 also specifies Core metadata set which is a basic minimum number of metadata elements that should be maintained for a dataset in order to identify a dataset for catalogue purposes. It includes mandatory metadata elements as well as recommended optional elements which will increase interoperability, allowing users to understand the geographic data and the related metadata provided by either the producer or the distributor

Metadata catalogues may facilitate retrieval of the themes and features of the Digital geodetic plan. However, the retrieval of the data is only based on keyword-based search, and the part concerning the semantics of the data is still missing and the user is not able to see the details about underlying data model. Retrieval of the data should consider feature attributes which can be spatial, thematic, qualitative and temporal. Although application schema may be referenced in metadata set, the problems of heterogeneity of formats for its representation, as well the meaning of schema elements persist and therefore it is not suitable for the any kind of automatic processing.

Record orientation of catalogues as in ISO 19115, is a clear user / client paradigm but it is hard to maintain and limited for complex metadata relationships. A registry model makes catalogs easier and more flexible to maintain, but it is rather complex when exposed to the clients. ebRIM allows the classification of data and services into categories which only partially solves the problem of semantics by introducing taxonomy, but non-taxonomic relationships are hard to maintain. Possible solution for the problem is the introduction of formal ontologies, namely OWL, a semantic markup language for the web. This approach is described in the next section.

3.2. Introducing Semantics into OGC Catalogues

Ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. It is used to reason about the properties of that domain, and may be used to describe the domain. Its role is to provide a shared vocabulary within a certain domain such as the real estate cadastre. The Land Administration Domain Model (LADM) specified in ISO 19152 standard (URL 29) provides a base for building ontologies in real estate cadastre domain. ISO 19152 is an international standard from ISO 19100 series of standards. Since the land administration is a large field, the focus of this standard is on that part of land administration that is interested in rights, responsibilities and restrictions affecting land, and the geometrical (spatial) components thereof. LADM contains a reference model that has two goals. First goal is to avoid reinventing and re-implementing the same functionality over and over again. Therefore, this standard offers basis for data model that can be expanded and adapted in order to develop an accurate data model. Second goal is to enable involved parties, both within one country and between different countries, to communicate, based on the shared vocabulary (that is, an ontology) implied by the model.

Based on the degree of generality, ontologies can be divided into three levels (Guarino 1998): top-level ontologies, domain ontologies, and application ontologies. Top-level ontologies describe the general concepts independent of domain, for example, object or event. Domain ontologies describe the concepts in a generic domain. Application ontologies are related to a specific domain or task that is intended for use in one application rather than across many applications. In geospatial domain ontology architecture has been provided in (Zhao et al. 2009). According to this approach ontologies are divided into six layers as shown in Fig. 7.

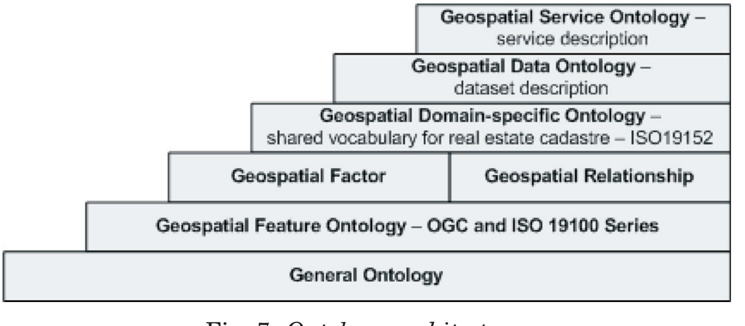

Fig. 7. *Ontology architecture.*

General ontology is the core upper level vocabulary representing common human consensus reality that all other ontologies must reference and it is domain independent. Geospatial feature ontology provides the core geospatial vocabulary and structure, and forms the ontological foundation of geospatial information and should be coordinated with geospatial standards such as ISO 19100 series and OGC. Geospatial factor and relationship ontologies describe location, unit conversion factors and geospatial and logical relationships between geospatial features. Geospatial domain specific ontology represents the specific concepts in one domain such as land administration. Geospatial data ontology provides a dataset description. Geospatial service ontology semantically describes the service using OWL-S (URL 30). Considering the traditional division into upper, domain and application ontologies, general ontology corresponds to upper ontology, next three layers correspond to domain ontologies while the part that relates to information and services belongs to application ontologies.

In order to build ontologies an open source ontology editor Protégé (URL 31) may be used. It has capabilities to build ontology in the OWL language and visualize taxonomies of OWL ontologies. It is the tool for OWL-based ontology development and inference; it is extensible via plug-ins (Knublauch et al. 2004). Protégé has its own internal representation mechanism for ontologies and knowledge bases, based on a metamodel, which is comparable to object-oriented and frame-based systems.

Cadastral domain ontology should be developed in accordance with ISO 19152 standard. This ontology should be related to Geospatial feature ontology developed in accordance with ISO 19100 series of standards. Cadastral domain ontology is the part of Geospatial Domain-specific Ontology and represents a shared vocabulary for the domain of the real estate cadastre. Extension of this ontology according to LADM profile for the national cadastre is at the application level when the goal is the integration of international cadastral data, although within the state it may be considered to be at the domain level. This ontology structure is shown on Fig. 8. Therefore, ontology for any national cadastre should be developed according to data model for the real estate cadastre in that country and should be aligned to LADM ontology. The example has been given in (Bošković et al. 2010).

Fig. 8. *Ontology dependency.*

Examples of the OWL classes from the LADM and the national cadastre ontologies have been given in the following listing. Listing 3.a shows the description of the OWL class *LA_SubParcel* in N3 notation. This class is related with the class *LA_Parcel* via existential and universal restriction on property *isPartOfParcel*, which specifies that the subparcel must be related to particular parcel i.e. if subparcel exists it must be the part of a particular parcel. Listing 3.b shows the OWL class *Building*. This class is the subclass of the OWL class *LA_Building* and it is related to *PartOfParcel*. This relation indicates that a building must be placed at exactly one part of parcel. OWL classes with the prefix LA_ belong to the LADM ontology, whereas other classes belong to the national cadastre ontology.

```
iso19152:LA_SubParcel
  a owl:Class ;
  rdfs:subClassOf
iso19152:LA_SpatialUnit ;
  rdfs:subClassOf
    [ a owl:Restriction ;
      owl:allValuesFrom
iso19152:LA_Parcel ;
      owl:onProperty
ladm:isPartOfParcel
    ] ;
  rdfs:subClassOf
   [ a owl:Restriction ;
     owl:onProperty
```

```
rec:Building
  a owl:Class ;
  rdfs:subClassOf
iso19152:LA_Building ;
  rdfs:subClassOf
    [ a owl:Restriction ;
      owl:cardinality
"1"^^xsd:int ;
      owl:onProperty
rec:isBuildingOf
    ] .
  rdfs:subClassOf
    [ a owl:Restriction ;
      owl:onProperty
```

Listing 7. *a) OWL Class LA_SubParcel, b) OWL Class Building.*

Ordinary OGC catalogues can be enhanced with such semantic information, which will enable users to discover and retrieve data based on semantics of the data model as well. This can be achieved by incorporating semantic annotations into metadata as described in OGC discussion paper (URL 32). This discussion paper proposes semantic annotation at three different levels: geospatial service metadata, data models and process descriptions, and actual data instances in the database. In order to process such information new catalogue clients need to be developed, but the existing clients can be used for processing ordinary metadata ignoring semantic extensions. A proposal how to semantically enhance OGC catalogues has been given in (Lutz and Klien 2006).

4. Conclusions

One of the essential components for the construction of a spatial data infrastructure at a regional, national or global level is the geospatial catalogue service. But, for the catalogue to be a useful component, it must enable access to geospatial metadata independently of the nature of search client applications. Client applications do not need to be developed by the same company or same technology that implemented the server. This is achieved by OGC Catalogue specification which various vendors must comply in order to achieve interoperability and make possible this enterprise and technological independence. This paper reviews General catalogue model, various information models that can be implemented, different protocol bindings, among which the most common is HTTP protocol binding which enables web interface and application profiles that combines various information models and protocol bindings. The analysis of the usage of metadata catalogue services in geodetic information systems has been given. The problem of the semantics of data has been discussed and the proposal for a possible solution for improvement based on ontologies has been made. It is necessary to develop these ontologies in accordance with OGC and ISO 19100 series of standards and data model.

References

Baader, F., McGuinness, D., Nardi, D., Patel-Schneider, P. F. (2002): Description Logic Handbook: Theory, Implementation and Applications, Cambridge University Press, Cambridge, UK.

Bai, Y., Di, L., Wei, Y. (2009): A taxonomy of geospatial services for global service discovery and interoperability, Computers and Geosciences, 35, 783–790.

Bošković, D., Ristić, A., Govedarica, M., Pržulj, Đ. (2010): Ontology Development for Land Administration, Proceedings of IEEE International Symposium on Intelligent Systems and Informatics (SISY), Subotica, 437–442.

Bulatović, V., Ninkov, T., Sušić, Z. (2010): Open Geospatial Consortium Web Services in Complex Distribution Systems, Geodetski list, 1, 13–29.

Erl, T. (2005): Service-Oriented Architecture: Concepts, Technology, and Design, Prentice Hall, New Yersey, USA.

Guarino, N. (1998): Formal ontology and information systems, Proceedings of the First International Conference on Formal Ontologies in Information Systems, FOIS'98, Trento, Italy, 3–15.

Knublauch, H., Fergerson, R. W., Noy, N. F., Musen, M. A. (2004): The Protégé OWL Plugin: an open development environment for semantic web applications, Lecture Notes in Computer Science 3298, Springer, 229–243.

Lutz, M., Klien, E. (2006): Ontology-based retrieval of geographic information, International Journal of Geographical Information Science, 20, 233–260.

Maguire, D. J., Longley, P. A. (2005): The emergence of geoportals and their role in spatial data infrastructures, Computers, Environment and Urban Systems, 29, 3–14.

Nebert, D. (2004): Developing Spatial Data Infrastructures: The SDI Cookbook, available at: http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf, (20.09.2010.).

Nogueras-Iso, J., Zarazaga-Soria, F. J., Bejar, R., Alvarez, P. J., Muro-Medrano, P. R. (2005): OGC Catalog Services: a key element for the development of Spatial Data Infrastructures, Computers and Geosciences, 31, 199–209.

Staab, S., Studer, R. (2009): Handbook on Ontologies, Springer, Berlin, Germany.

Tait, M. G. (2005): Implementing geoportals: applications of distributed GIS, Computers, Environment and Urban Systems, 29, 33–47.

Zhao, P., Di, L., Yang, W., Yu, G., Yue, P. (2009): Geospatial Semantic Web: Critical Issues, in: Karimi, H.A., Handbook of Research on Geoinformatics, Information Science Reference, 178–189.

URL 1: Open Geospatial Consortium,
http://www.opengeospatial.org, (21.09.2010.).

URL 2: OpenGIS Catalogue Service Implementation Specification 2.0.2, 2007.,
http://www.opengeospatial.org/standards/cat, (21.09.2010.).

URL 3: ISO 15836:2003 – Information and documentation – The Dublin Core metadata element set,
http://www.iso.org/iso/catalogue_detail.htm?csnumber=37629, (21.09.2010.).

URL 4: ISO 19115:2003 – Geographic information – Metadata,
http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020, (21.09.2010.).

URL 5: OASIS ebXML Registry Information Model Version 3.0,
http://www.oasis-open.org/committees/download.php/13591/
/docs.oasis-open.orgregrepv3.0specsregrep-rim-3.0-os.pdf, (21.09.2010.).

URL 6: ISO/TS 19139:2007 – Geographic information –- Metadata –- XML schema implementation,
http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557, (21.09.2010.).

URL 7: Extensible Markup Language (XML),
http://www.w3.org/XML, (21.09.2010.).

URL 8: OMG Unified Modeling Language,
http://www.uml.org, (21.09.2010.).

URL 9: OWL Web Ontology Language Overview,
http://www.w3.org/TR/owl-features, (21.09.2010.).

URL 10: World Wide Web Consortium (W3C),
http://www.w3.org, (21.09.2010.).

URL 11: Resource Description Framework (RDF),
http://www.w3.org/RDF, (21.09.2010.).

URL 12: RDF Vocabulary Description Language 1.0: RDF Schema,
http://www.w3.org/TR/rdf-schem, (22.09.2010.).

URL 13: HTTP – Hypertext Transfer Protocol,
http://www.w3.org/Protocols, (22.09.2010.).

URL 14: ANSI/NISO Z39.50-2003,
http://www.loc.gov/z3950/agency/Z39-50-2003.pdf, (22.09.2010.).

URL 15: OMG Common Object Request Broker Architecture,
http://www.corba.org, (22.09.2010.).

URL 16: Object Management Group,
http://www.omg.org, (22.09.2010.).

URL 17: OpenGIS Filter Encoding Implementation Specification,
http://www.opengeospatial.org/standards/filter, (22.09.2010.).

URL 18: ISO/IEC TR 10000-1:1998, Information technology –- Framework and taxonomy of International Standardized Profiles – Part 1: General principles and documentation framework,
http://www.iso.org/iso/catalogue_detail.htm?csnumber=30726, (22.09.2010.).

URL 19: OpenGIS Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile (1.0.0),
http://portal.opengeospatial.org/files/?artifact_id=21460, (22.09.2010.).

URL 20: CSW-ebRIM Registry Service – Part 1: ebRIM profile of CSW (1.0.1),
http://portal.opengeospatial.org/files/?artifact_id=31137, (22.09.2010.).

URL 21: Catalogue Services Standard 2.0 Extension Package for ebRIM Application Profile: Earth Observation Products,
http://www.opengeospatial.org/standards/cat2eoext4ebrim, (22.09.2010.).

URL 22: GML 3.1.1 Application Schema for EO products (1.0.0),
http://portal.opengeospatial.org/files/?artifact_id=31065, (22.09.2010.).

URL 23: OGC Catalogue Services – OWL Application Profile of CSW (0.3.0),
http://portal.opengeospatial.org/files/?artifact_id=32620, (22.09.2010.).

URL 24: GeoNetwork opensource,
http://geonetwork-opensource.org, (22.09.2010.).

URL 25: deegree – Free Software for Spatial Data Infrastructures,
http://www.deegree.org, (22.09.2010.).

URL 26: An overview of ArcCatalog,
http://webhelp.esri.com/arcgisdesktop/9.2/
/index.cfm?TopicName=an_overview_of_arccatalog, (22.09.2010.).
URL 27: ERDAS Apollo,
http://www.erdas.com/tabid/84/currentid/1850/default.aspx, (22.09.2010.).
URL 28: Zakon o državnom premeru i katastru,
http://www.rgz.gov.rs/web_preuzimanje_datotetka.asp?FileID=321, (22.09.2010.).
URL 29: ISO/CD 19152, Geographic information – Land Administration Domain Model (LADM),
http://www.iso.org/iso/catalogue_detail.htm?csnumber=51206, (22.09.2010.).
URL 30: OWL-S: Semantic Markup for Web Services,
http://www.w3.org/Submission/OWL-S, (22.09.2010.).
URL 31: Protégé,
http://protege.stanford.edu, (22.09.2010.).
URL 32: Semantic annotations in OGC standards,
http://portal.opengeospatial.org/files/?artifact_id=34916, (22.09.2010.).

Katalog metapodataka u prostornim informacijskim sustavima

SAŽETAK. U radu je prikazan kratki pregled usluga kataloga metapodataka Otvorenoga geoprostornog konzorcija (OGC) koji ima ključnu ulogu u otkrivanju geoprostornih izvora informacija u Prostornoj infrastrukturi podataka (SDI). Pojam Prostorne infrastrukture podataka obuhvaća skup tehnologija, strategija i institucionalnih sporazuma koji osiguravaju lakši pristup geoprostornim podacima. SDI je pogodan za upotrebu pri otkrivanju geoprostornih podataka, evaluaciji, kao i različitim primjenama unutar vladinog, komercijalnog i neprofitabilnog sektora, akademskih institucija, itd. Usluge kataloga metapodataka specificirane su u OGC Specifikaciji ostvarivanja usluga kataloga. Dio specifikacije koji se odnosi na web-sučelje koje podržava pohranu, učitavanje i upravljanje podacima koji se odnose na web-servise naziva se Katalоški servis za web (CSW). Katalozi metapodataka su agenti za usluge koji predstavljaju ključnu komponentu u arhitekturi namijenjenoj uslugama, a koja upravlja zajedničkim izvorima i olakšava otkrivanje izvora unutar otvorenoga distribuiranog sustava. OGC usluge stekle su značajnu popularnost u proteklim godinama, a broj organizacija koje ih koriste se povećao. Međutim, puni potencijal kataloga metapodataka još nije dostignut, ne samo zbog nedostatka odgovarajuće dokumentacije podataka u obliku standardiziranih metapodataka, već i zbog nedostatka semantike podataka. Napravljena je analiza korištenja usluga kataloga metapodataka u geodetskim informacijskim sustavima te je dan prijedlog za moguća rješenja u svrhu poboljšanja.

Ključne riječi: *OGC katalog, CSW, prostorni informacijski sustavi, metapodaci, semantika.*

Prihvaćeno: 2010-11-05