Implementation of a geocatalogue to assist location and recovery of open geographic data

Drausio Gomes dos Santos
is an employee of the Federal Court of Accounts – Brazil. He has an M.A. in Applied Computing from the University of Brasília (UnB) and a specialist degree in IT Management from the Federal University of the State of Pernambuco (Cin/UFPE) and a technological degree in Data Processing from the São Paulo Technology College (FATEC/SP) and a B.A. in Geography from the State University of São Paulo (USP).

Alexandre Zaghetto
has an M. A. and a PhD in Electric Engineering from the University of Brasília (UnB). He has a B.A. in Electric Engineering, focused on Electronics, from the Federal University of the State of Rio de Janeiro (UFRJ) He is also an adjunct professor of the Computer Science Department of UnB and leader of the Research Group on Biometric Systems

ABSTRACT

This paper approaches the creation of a geocatalogue, which uses semantics and ontologies to assist in the process of discovering geographic resources in services of an open spatial data infrastructure. Its architecture incorporates a non-relational graphs-oriented database and it is subject to implementation in cloud computing. To conceive it, the SERVUS systems development methodology, specialized in architecture oriented to geospatial services, was used.

Keywords: Software Engineering. Geocatalogue. SERVUS. SDI. Semantics. Open data.

1. INTRODUCTION

It has been a challenge to the public administration bodies to obtain a spatial view of the result of governmental actions in a simplified and updated way, as data are decentralized, produced by several sources and in different moments. The Government have taken some actions in order to standardize, optimize resources and integrate data, such as the creation of the National Spatial Data Infrastructure of Brazil (NSDI – BRA).

The NSDI-BRA is an initiative of creating an architecture for dissemination and regulation of use of geogra-

phic data within the public administration. We highlight one of the goals set in its conception, which is to

> "avoid the duplicity of actions and the waste of resources in obtaining aerospatial data by public administration bodies, through the dissemination of metadata related to those data available at the public entities and bodies of the federal, state, district and municipal spheres" (BRASIL, 2008).

Camboim (2013) highlights the governmental initiative of creating the National Open Data Infrastructure (NODI) and the Brazilian Open Data Portal as a strategy to adop the Linked Open Data. This author proposes an architecture capable of making data in the NSDI-BRA available in the NODI in an integrated manner. Such proposal of architecture stands out due to the use of semantic layers and ontologies by joining the semantic web techniques and geospatial data, entering the geosemantics field.

Architectures from the current SDI are, as a rule, service-oriented and they adopt technological patterns established by the Open Geospatial Consortium (OGC). Some of these patterns define the operation of geoservices, which allow the collection of cartographic maps, thematic maps, geographic data and metadata.

One of the essential compounds in an SDI architecture is the services catalogue. The SERVUS software development methodology (USLÄNDER, 2010), specially created to deal with the use and composition of geospatial services, proposes the use of a semantic geocatalogue to assist in locating geographic network connected resources.

To study this aspect, this document is divided in the following way: Section 2 presents some correlated works; Section 3 presents the theoretical foundation and addresses SDI and SERVUS methodology; Section 4 presents the proposed model, details of the geocatalogue and an example of use with methodology SERVUS and finally, Section 5 presents the conclusions.

2. CORRELATED WORKS

Andrade and Baptista (2011) followed a path that included semantic search, listing great amounts of information from several sources, establishment of metrics and generation of a range of resources available in a spatial data infrastructure. The authors assumed that the use of ontologies would be the way to make the searches more precise and to facilitate the automation process. As contributions, we have had the establishment of a combined metric and the creation of a semantic network.

Daltio and Carvalho (2012) proposed the creation of a framework for semantic recovery of spatial data. Such framework was based on a process of semantic annotation of the geographic resources and use of an ontology management service. In this paper, it is worth noting the importance given to the process of selection of the ontologies.

Gimenez, Tanaka and Baião (2013) present a proposal of semantic integration for the NSDI-BRA by using geo-ontologies of domain incorporated in semantic layers acting on the SDIs.

3. THEORETICAL FOUNDATION

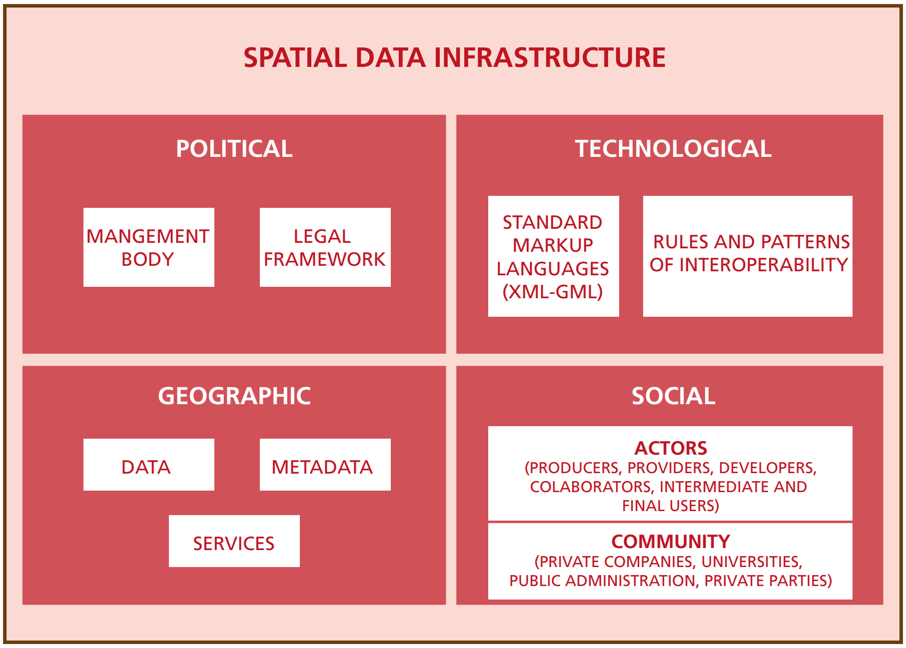
3.1 SPATIAL DATA INFRASTRUCTURE (SDI)

Bernabé-Proveda and López Vázquez (2012, p. 57-59) conceptualize Spatial Data Infrastructure (SDI) as an infrastructure necessary to access, share, exchange, match and analyze geographic data in a standardized and interoperable way. They have also considered the need of those data being available in a network, through a set of systems which use standard protocols and interfaces to promote the creation of applications that can be seen by users as a single system. The SDI can be classified as a structure of technological, geographic, social and political components as illustrated in Figure 1.

The technological component is represented by an architecture based on patterns of interoperability capable of sharing geographic data and information. The markup languages XML and GML have a key role in this component. The social component is represented by a set of actors, amongst them, data producers, service providers, users, software developers and those responsible for patterns and rules, besides a great community composed by private companies, government, universities and society as a whole. The geographic component is represented by data, their metadata and geoservices. The political component is represented by the persons and by the body responsible for establishing the regulation framework and its operating rules.

OGC Web Services Common Standard is a technological standard that proposes a common interface for a set of geoservices: Web Map Service (WMS), Web Feature Service (WFS) and Web Coverage Service (WCS). In this proposal, the intention is to standardize the ways of access, operations, mandatory and optional parameters, besides data structures, aiming at reducing the efforts of interoperability. One of the operations defined is the GetCapabilities, which recovers the metadata re-

Figure 1:
Components of an SDI



lated to the capabilities of the geoservices (GREENWOOD; WHITESIDE, 2010).

WMS consists in a service that produces dynamic maps, in which geographic information are organized in layers. The WFS service deals with geographic entities with discrete or vector data in GML format, representing attributes and geometries. WCS is the service that supports the recovery of spatial data as coverages.

A catalogue of services allows the location of data or geographic services through a range of operations, amongst them, the GetRecords operation, which recovers a set of records of metadata (OGC, 2016).

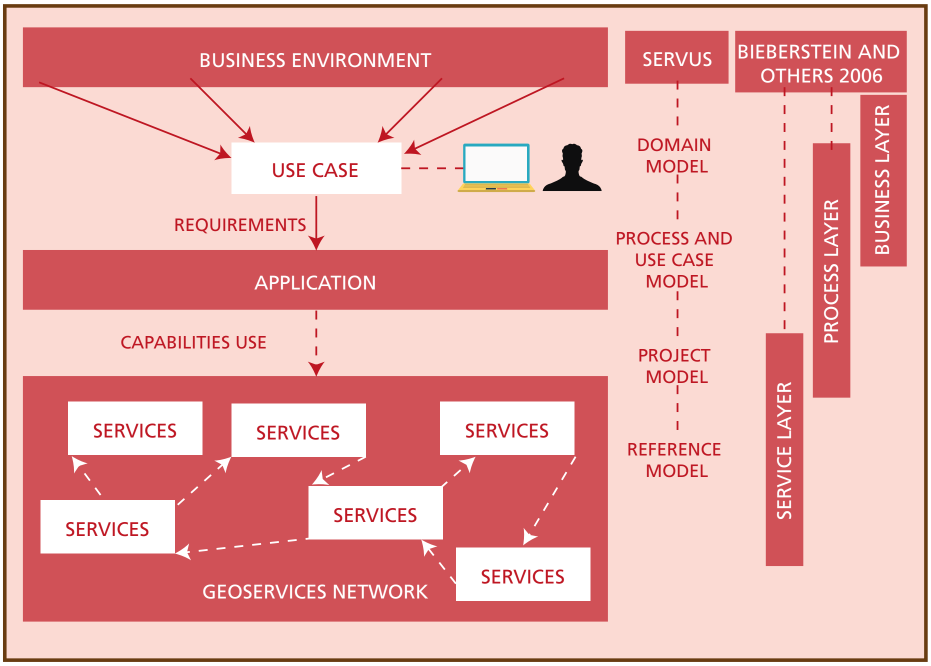
3.2 SERVUS DEVELOPMENT METHODOLOGY

The acronym SERVUS originates from two software engineering terms: "SERvice" and "USe case". The central problem such methodology proposes to solve is – considering a set of use cases which meet the business requirements and demand capacity of resources in a network of geospatial services – how to discover and associate required resources to offered resources, composing an application that supplies quality, functional and informational requirements. Two other elements compose this scenery: semantic descriptions in a network of semantic resources, and side conditions to the development of environmental systems (processes of discovery of resources, the use of OGC patterns, resources matching) (USLANDER, 2010).

According to Uslander (2010, p.90), "SERVUS understands the environmental information systems project as an interactive discovery and matching activity: available capabilities are discovered and associated to the users requirements formulated as use cases".

A modeling language, a development process and a reference architecture compose methodology. During the development process, artifacts are created, such as a domain model, a process/use case model and a project model. The main model is the project model and it expresses the requirements and capabilities in the form of resources. The generated models are based on the abstraction layers proposed by Bieberstein (2006). Figure 2 displays the mapping between requirements and capabilities, the models produced and the relation with the abstraction layers.

Figure 2:
SERVUS model hierarchy



Source: Adapted from Uslander (2010, p.92)

The domain model of the SERVUS methodology represents the thematic domain of the problem to be solved. It formally defines the part of the world that constitutes the speech universe between the user and the system designer, that is, it comprises the shared knowledge on the application domain. Typically, such shared knowledge are represented by the specification of an ontology (USLANDER, 2010).

The reference model is composed by an architectural framework, responsible for orientation and rules on how to specify the system, and by a conceptual model. The SERVUS conceptual model is a metamodel in accordance with the precepts of Model Driven Architecture (MDA) and Meta-Object Facility (MOF). The main metaclasses are feature, interface, service and resource. The conceptual model is composed by three subsets of metaclasses. The first one is associated to the service view, the second one to the information view and the third one to the resource view. The project model may be seen as a composition of the three models: requirements (REQ's), capabilities (CAP's) and mapping between the two first models (REQ2CAP). Figure 3 displays a general table of models and activities of the project stage.

Below there is a description of the main activities in the development process.

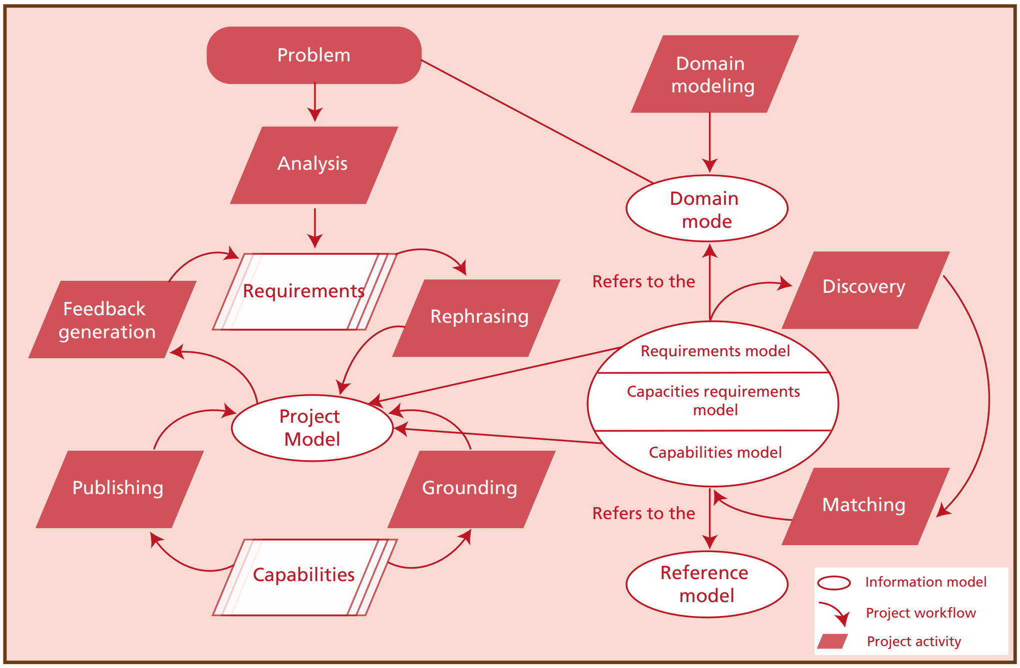
a. ***Publishing***: manual or automatic activity of searching for capabilities in the geospatial services network and of incorporation to the capabilities model. Capabilities are converted into the offered resources, that is, those available in the geoservices network;

b. ***Rephrasing***: activity responsible for converting requirements to extract a set of necessary resources for holding the use cases. Such resources may be associated to the concepts of ontology and then uploaded in the semantic resources network, composing a requirement model;

c. ***Discovery***: activity responsible for selecting, from the package of capabilities arisen in the publishing stage, a set which meets the required resources arisen in the rephrasing stage. For each required resource there may be a set of resources offered which meets the needs;

d. ***Matching***: activity responsible for associating requirements to capabilities, that is, required resources to offered resources. It evaluates among the candidate capabilities, catalogued in the discovery stage, the most appropriate ones to the requirements. The final result of

Figure 3:
Models and Project activities



Source: Adapted from Usländer (2010, p. 99) this stage is the requirements/capacities mapping model;

e. ***Grounding***: activity which provides a new capability in the geospatial services network.

SERVUS proposes the creation of a semantic catalogue, which composes an implementation architecture and a project environment. The semantic catalogue role is to communicate with the network geoservices and to process searches through resources. It works as a semantic extension that allows searches based on ontologies as well as the evaluation of the semantic proximity to the achieved results. Regarding the support to the methodology, the semantic catalogue accounts for activities of harvesting, which comprises the collection of services metadata, and of publishing, the publication of resources available to a network of semantic resources.

4. PROPOSED MODEL

This paper was developed in stages, as follows: literature review with studies on the SDIs, SOA development methodology, SERVUS methodology, graphs-oriented data banks, geocatalogue and semantic search. Another stage was a survey of requirements based on the geocatalogues studied - mainly the one proposed by the SERVUS methodology - on the needs of the environment stakeholders and on the current technological infrastructure of the body. The next stages were definition of high-level architecture, considering, among other aspects, data sources, SDI architectures and NSDI--BRA geoservices and development of prototype, including the definition of development languages, components, software libraries and the development environment. Then the stages of implementation with definition of environment and prototype installation; elaboration of a scenery to use the geocatalogue typical of environment auditing, prototype validation and finally evaluation of results (SANTOS, 2016).

The architecture chosen to implement the geocatalogue prototype was a web application in three layers. The presentation layer was developed in HTML5 by using JavaScript libraries: Bootstrap 3.3.5, jQuery 2.1.0, Google Maps 3.0 and the bootstrap-slider.js component. The business layer was implemented in Java version 1.7 and the application wrapped in web archive format – WAR. The requests of the presentation layer are made through the http protocol and they access a REST (Representational State Transfer) standard server, which returns documents in a JSON (JavaScript Object Notation) format. The business layer was hosted in an Amazon EC2 cloud service. The persistence layer has been implemented through an

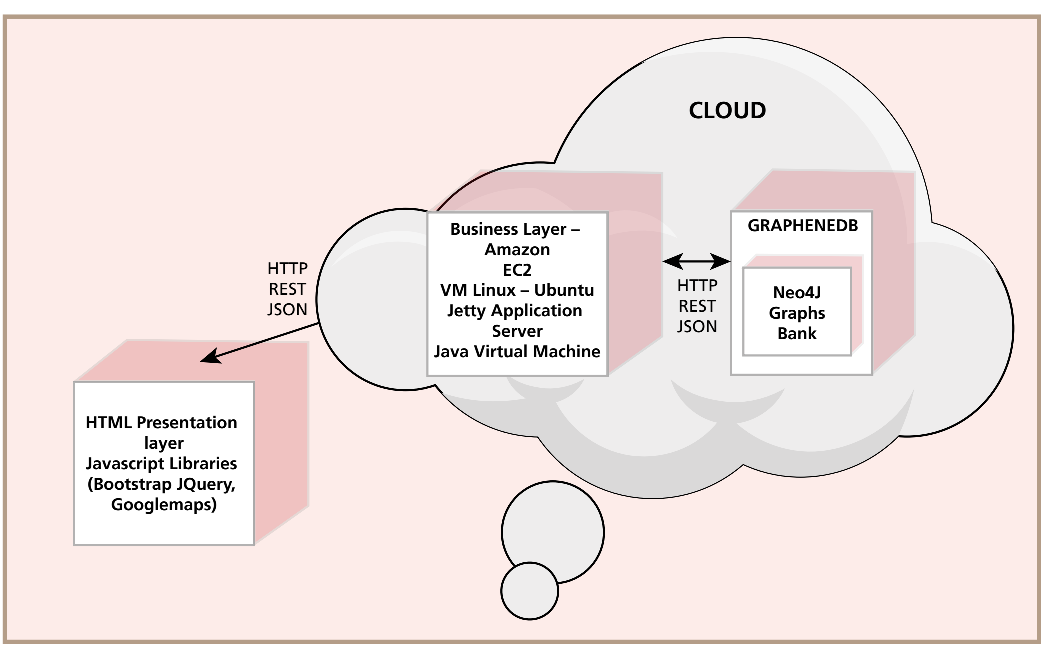
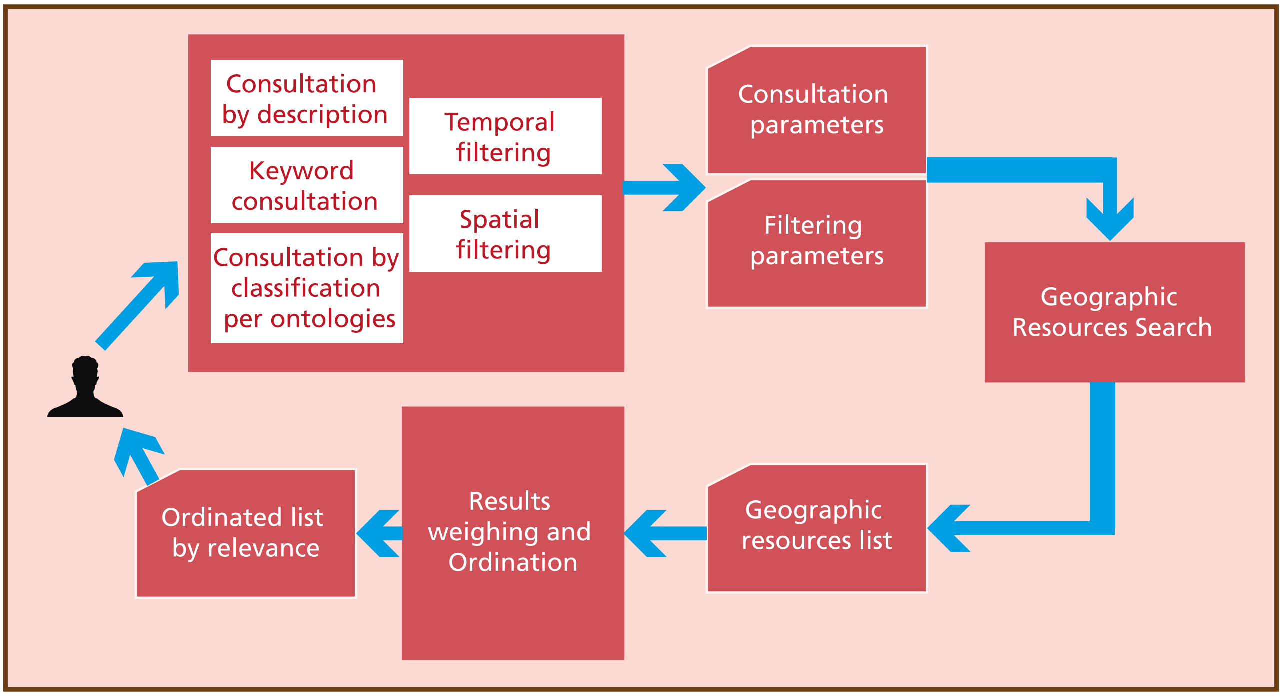


Figure 4: Implementation architecture

Figure 5:
Catalogue search model



NoSQL graphs data bank, the Neo4J 2.3.1. Hosting of the persistence layer was held in the GrapheneDB web page, which implements the Neo4J banks in a cloud environment. The communication between the business and persistence layers also takes place through REST interfaces and in JSON format. Figure 4 displays the implemented architecture.

Consultations can be made through description (free text), keywords and also through ontologies terms. For each one of the previous items, sentences or terms are divided into tokens which will be the terms used in the search. Filtering may be held by geographic coordinates (spatial filter) and by year (temporal). A weight is attributed to each ordered pair, composed by the term and the searched field, to be the base for procedure to generate a ranking. Figure 5 displays the geocatalogue search model.

The effective result of a search depends on the metadata quality and on the business value attributed to each field. To deal with this aspect

Table 1:
Catalogue search model

Search terms	Field with localized term	Weight	Maximum weight
Description: "vegetation" – (weight 8)	Name – (weight 8)	64	216
VCE ontology (root term): "vegetation" – (weight 6)	Description – (weight 6)	36	162
Keyword: "agricultural" – (weight 5)	Abstract – (weight 5)	25	135
	Total weight	125	513
	Relevance index calculated (125/513)	0,24	

a weighting scheme of the search fields and of the metadata fields was elaborated and a relevance index that acts as the basis for the process of generating a results ranking was established. Table 1 displays a sample of the relevance index calculation.

The records of resources are recovered through REST requests to the remote data bank, composed by the Cypher language commands. The main types of interaction are consultations of ontologies, creation of relationships between the resources and the generation geographic resources ranking. Figure 6 displays the ranking in a list format with the corresponding links to the resource preview, download and metadata preview.

For validation of the use of the geocatalogue, stages of the SERVUS methodology were used. We used as a baseline a problem related to an environmental audit that we present as follows: how to obtain the necessary data, in Brazil, to raise the locations of environmental protection areas, by using the available geoservices as a basis in the network

Figure 6:
Geocatalogue search screen

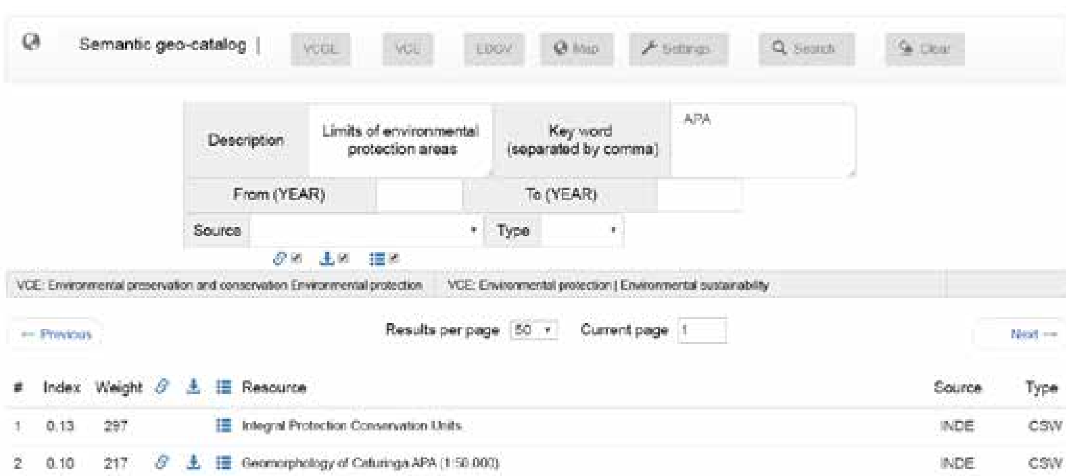


Figure 7:
Environmental protection areas imported to QGIS tool

of national spatial data infrastructure of public institutions and bodies?

The result was the location and preview of those areas. Figure 7 presents the environmental protection areas located by the geocatalogue, whose data were downloaded from WFS services and imported to the Quantum GIS tool.

5. CONCLUSION

This paper approached aspects of the implementation of a geocatalogue that allows the use of ontologies as research parameters. Consultations are carried out in the metadata of geoservices of spatial data infrastructures. We addressed implementation aspects such as architecture and data models. Regarding architecture, the geocatalogue was implemented as a three-layer web application that can be implemented in cloud environments. The data model was implemented in a graphs-oriented data bank.

The geocatalogue proved to be feasible as a tool to search for geographic resources and to support the SERVUS methodology, allowing operation in the publishing and matching stages. We can consider the integrated use of Cypher language in the resources network, ontologies and metamodel manipulation in non-relational graphs-oriented database, as contributions. The establishment of parametrization mechanisms for consultations, through parameters weighting and metadata fields, and the establishment of a relevance index, are contributions to the search field in metadata bases. For future works, we can consider automation of the harvesting process and integration with the corporate spatial data infrastructure, in such a way that the geocatalogue can perform together with the geoportals.

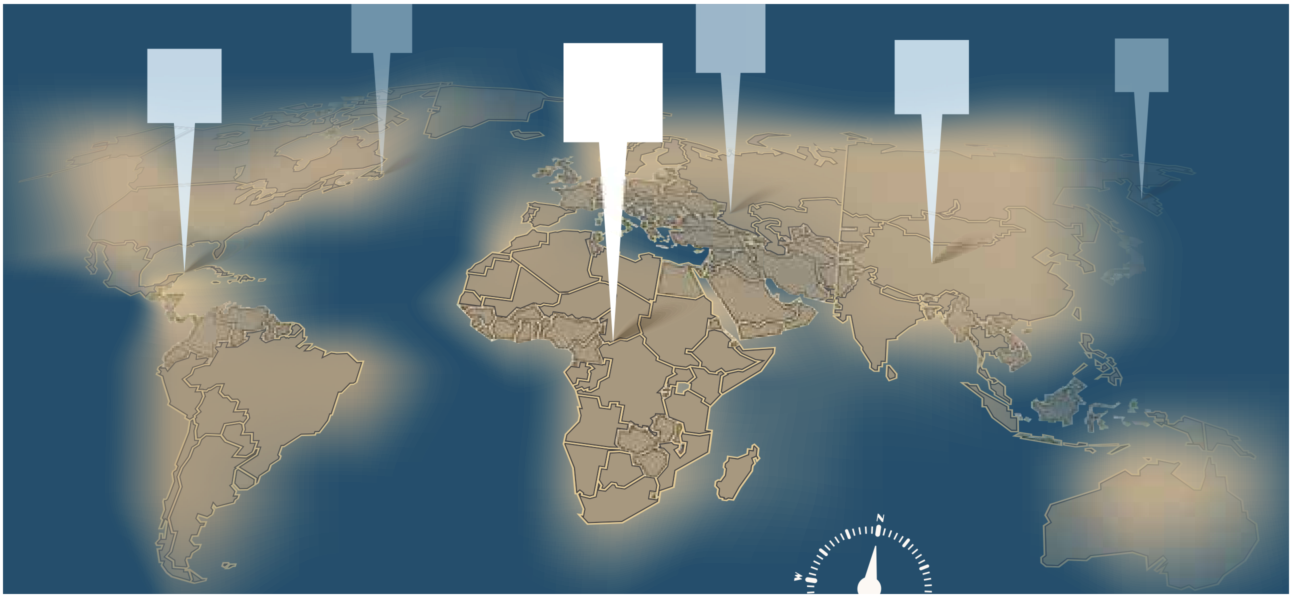
REFERENCES

ANDRADE, F. G.; BAPTISTA, C. S. Using semantic similarity to improve information discovery in spatial data infrastructures. Journal of Information and Data Management, v. 2, n. 2, p. 181, 2011.

BERNABÉ-POVEDA, M. Á.; LÓPEZ-VÁZQUEZ, C. M. (Eds.). Fundamentos de las infraestructuras de datos espaciales (IDE). Madrid: Universidad Politécnica de Madrid, 2012. 593p.

BIEBERSTEIN, N. Service-oriented architecture compass: business value, planning, and enterprise roadmap. Indiana: FT Press, 2006.

BRASIL. Decreto n. 6.666, de 27 de novembro de 2008. Institui, no âmbito do Poder Executivo Federal, a Infraestrutura Nacional de Dados Espaciais – INDE, e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm>. Acesso em: 22 out. 2016.



CAMBOIM, S. F. Arquitetura para integração de dados interligados abertos à INDE-BR. 2013. 140f. Tese (Doutorado em Ciências Geodésicas) – Universidade Federal do Paraná, Curitiba, 2013.

DALTIO, J.; CARVALHO, C. A. Um framework para recuperação semântica de dados espaciais. In: BRAZILIAN SYMPOSIUM ON GEOINFORMATICS, 13, 2012, Campos do Jordão. Proceedings... São José dos Campos: MCTI/INPE, 2012. p. 60-65.

GIMENEZ, P. J. A.; TANAKA, A. K; BAIÃO, F. Uma proposta de integração semântica para a Infraestrutura Nacional de Dados Espaciais usando geo-ontologias. In: WORKSHOP DE COMPUTAÇÃO APLICADA AO GOVERNO ELETRÔNICO (WCGE), 2013, João Pessoa. Anais... João Pessoa: UFPB, 2013. v. 5. p. 25-32.

GREENWOOD, J.; WHITESIDE, A. OGC Web Services Common Standard. 2010. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.248.6351&rep=rep1&type=pdf>. Acesso em: 22 out. 2016.

OGC. Catalogue Services 3.0 Specification: HTTP Protocol Binding. 2010. Disponível em: <http://docs.opengeospatial.org/is/12-176r7/12-176r7.html>. Acesso em: 22 out. 2016.

SANTOS, D. Implementação de um geocatálogo utilizando banco de dados orientado a grafos para apoio à metodologia SERVUS. 2016. 82f. Dissertação (Mestrado em Computação Aplicada) – Universidade de Brasília, Brasília, 2016. Disponível em: <http://alexandre.zaghetto.com/wp-content/uploads/2016/07/Disserta%C3%A7%C3%A3o_Vers%C3%A3oFinalParaEdi%C3%A7%C3%A3o.pdf >. Acesso em: 22 out. 2016.

USLÄNDER, T. Service-oriented design of environmental information systems. Vol. 5. Karlsruhe: KIT Scientific, 2010.