

edpMGB - A Metadata Editor Built as SaaS for the Brazilian Geospatial Metadata Profile

Marcos Montanari, Vitor Dias, Eduardo Lourenço and Jugurta Lisboa-Filho
Department of Informatics, Federal University of Viçosa, Viçosa, MG, Brazil

Keywords: Spatial Data Infrastructure, INDE, Editor, Metadata, MGB Profile.

Abstract: Aiming to encourage the reuse and prevent wasting resources in the production of spatial data, in 2008, the Brazilian government edited Act 6,666, instituting the National Spatial Data Infrastructure (*Infraestrutura Nacional de Dados Espaciais - INDE*). To meet the need of geospatial data documentation, the INDE defined the Brazilian Geospatial Metadata (MGB) profile based on ISO 19115:2003. This paper describes the development of edpMGB, a metadata editor for the MGB profile. The editor is open-source and is being released in the cloud via Web following the Software as a Service (SaaS) model, so it can be accessed from any point on the Internet, requiring only a Web browser. This editor was developed on the scope of the research and development project (R&D) Geoportal Cemig – SDI-based corporate GIS (*Geoportal Cemig - SIG corporativo baseado em IDE*), whose objective is the implementation of a corporate SDI for the Minas Gerais Power Company (*Companhia Energética de Minas Gerais - Cemig*).

1 INTRODUCTION

Within Information and Communication Technology, information redundancy and lack of data standardization proves very common. The same data often ends up being produced, managed, used, and stored by several independent producers that use different formats and standards, who seek to meet exclusively the individual needs of specific users (Dornelles and Iescheck, 2013).

The greater production of spatial data requires documenting them so they can be reused. A piece of data immersed in its context becomes information, however, with no such documentation, it is virtually worthless information (Nebert, 2004).

In order to prevent actions in duplicity and wasted resources to obtain spatial data, the Brazilian government began, in 2003, studies aiming to integrate and reuse geospatial data produced by the different federal administration organs. In 2008, Act 6,666 of November 27th established the National Spatial Data Infrastructure (*Infraestrutura Nacional de Dados Espaciais - INDE*) (Brasil, 2008). INDE's goal is "to catalog, integrate, and harmonize the geospatial data produced and maintained by the different governmental institutions so as to facilitate their location, exploration, and access by any user connected to the Internet" (Concar, 2009).

To Leme (2006), geographic metadata correspond to the documentation of geographic data and are created according to standards. Such standards consist of a set of regulations that allow the geographic data to be described textually in a previously established manner (Leme, 2006). The most well-known geographic metadata standards were defined by the Federal Geographic Data Committee (FGDC) and by the International Organization of Standards (ISO). These institutions established international geographic metadata standards that meet the needs of different users, thus which comprises the variability in geographic information (Prado et al., 2010).

In order to meet the metadata standardization demands started by INDE, the National Cartography Committee (CONCAR) created the Brazilian Geospatial Metadata Profile (MGB Profile) based on norm ISO 19115:2003. A metadata profile is a basic set of elements that portray the characteristics of geospatial products of a given community and guarantees their identification (ISO, 2003).

Based on norm ISO 19139:2007, XML schemas were defined to materialize and code ISO 19115:2003 as a file. Since the MGB profile is based on norm ISO 19115:2003, its metadata must also follow the materialization standards defined in norm 19139:2007 to increase the interoperability among

the systems that use the profile as their base (Pascoal et al., 2013).

The present paper aims to describe the development of edpMGB, a metadata editor for the MGB profile. edpMGB is open-source software and is available on the cloud via the web following the model Software as a Service (SaaS). Thus, it can be accessed from anywhere and requires only a web browser (Weiss, 2007). Some features of this metadata editor are also available through web services, which allow other systems to use some of this tool's features.

This editor was developed in the context of the research and development (R&D) project “GeoPortal Cemig – SIG corporativo baseado em IDE” (Cemig GeoPortal – SDI-based corporate GIS) being developed to help implement a corporate SDI for Minas Gerais Power Company - Cemig (*Compania Energética de Minas Gerais*).

The main reason for developing edpMGB was the need for a specific tool for this metadata standard focusing on the Brazilian technical audience that works with spatial information. Other editors are available in the market, such as CatMDEdit (Catmdeedit, 2015) and GeoNetwork (GeoNetwork, 2012). Although these are metadata editors, they are not specific for documenting metadata in the MGB profile. GeoNetwork, in particular, can be used alongside the editor described in the present study.

2 THEORETICAL FRAMEWORK

2.1 Cloud Computing

According to Armbrust et al. (2010), cloud computing refers to applications provided as online services as well as the hardware and software systems that provide such services. These systems favor the user's access to different applications using the internet, regardless of the platform used or where the user is. Therefore, all data processing and storage occurs at some online datacenter, as virtual servers.

Cloud computing takes place through different types of services. The main one, which will be approached in this study, is the concept of Software as a Service (SaaS) (Yang et al., 2011). This type of service has several resources that the end user can access via web browsers (Armbrust et al., 2010). The metadata editor presented in this study is an example of SaaS.

2.2 Web Services

According to the W3C (2004), a web service is a software system designed to support the interoperability among machines on a network. A web service is accessed through platform-independent protocols and data formats such as the Hypertext Transfer Protocol (HTTP), eXtensible Markup Language (XML), and Simple Object Access Protocol (SOAP). A web service's interface is accessible through standardized XML messages, i.e., in text format (Lopez-Pellicer, 2012).

The editor hereby presented makes available web services that can be accessed via SOAP by any application regardless of the technology used, with no need to access the system's interface. The services can be used by developers that want to use them in their applications.

2.3 Brazilian Geospatial Metadata Profile (MGB Profile)

According to Pascoal et al., (2013), the purpose of the MGB profile is to establish a national standard to structure geospatial metadata. It was created with the collaboration of representatives of Brazilian geospatial-data-producing organs. The profile was entirely based on the international norm ISO 19115 (ISO, 2003). However, this profile does not define how these metadata are structured in electronic files, which is managed by norm ISO 19139:2007. Based on Pascoal et al., (2013), norm ISO 19139 aims to materialize the concepts of ISO 19115:2003 into a file by coding these metadata in XML schemas. Before the MGB profile was created, the profiles previously established in several other countries were analyzed, all of which based on norm ISO 19115, such as the following profiles:

- MIG – *Metadados de Informação Geográfica* (Portugal);
- NEM – *Núcleo Español de Metadados* (Spain);
- NAP – North American Profile (USA/Canada);
- LAMP – Latin American Metadata Profile (proposed for Latin America).

The MGB profile has a simplified version called summarized profile, which represents the minimum elements the geospatial metadata must have. The summarized profile has 23 elements, illustrated in Table 1. The idea is to make the adoption of the full MGB profile, comprising 82 elements, optional while all data in Brazil must be documented with at least this summarized core (Concar, 2009).

When the MGB profile was being built, the ISO 19115 standard was analyzed in order to identify the section, entities, and elements that describe the characteristics of the geospatial data produced in Brazil. Thus, the profile defines a subset of elements available in ISO 19115, changes the cardinality of some elements, and defines some code lists (Pascoal et al., 2013).

The elements in the MGB profile are organized into the following sections (Concar, 2009):

- Section 1 (Identification) - provides basic information on the geospatial dataset such as title, date, person or institution responsible, and summary of the data;
- Section 2 (Identification of the geographic dataset – GDS) - refers to the information required to identify and evaluate a GDS. This section characterizes the type of spatial representation, scale, language, extension, etc.;
- Section 3 (Restriction information) - publicizes information regarding access and use restrictions and is made up of two entities, one regarding legal restrictions and the other, security restrictions;
- Section 4 (Quality) - allows the quality of a dataset to be evaluated by informing the hierarchical level, lineage, and report on the data;
- Section 5 (Maintenance information) - informs the maintenance and update frequency;
- Section 6 (Spatial representation information) - describes the mechanisms used to represent the spatial information (matrix or vector);

- Section 7 (Reference system) - information on the reference system, including the coordinates system and the geodesic referential of the spatial dataset;
- Section 8 (Content information) - describes the catalog of features and the content of the matrix data;
- Section 9 (Distribution) - reports information related to the distributor and to the alternatives to obtain geographic data;
- Section 10 (Metadata) - section responsible for information on its own metadata. It includes the person or institution responsible, creation date, norm used, etc.

2.4 Norm ISO 19139:2007

Norm ISO 19139:2007 defines a set of XML schemas for metadata defined in ISO 19115:2003. It aims to define a file format for geospatial metadata that follows ISO 19115:2003. These schemas enable structuring and validating the metadata's XML files in accordance with the norm (ISO, 2007). The XML format allows the metadata instances to go around the internet, including geospatial web services, as predicted in the specification of the Web Services Catalog (Nebert et al., 2007).

Maintained since 1998 by the W3C, XML is a flexible and simple mark-up language. Its main characteristics are being a text-based language, separating content from formatting, being simple and easily interpreted, and allowing the creation of

Table 1: Entities and elements of the metadata core of the summarized MGB profile.

Entity/Element	Condition	Entity/Element	Condition
1. Title	mandatory	13. Reference system	mandatory
2. Date	mandatory	14. Linage	optional
3. Responsible	mandatory	15. Online access	optional
4. Geographic extension	conditional	16. Metadata identifier	optional
5. Language	mandatory	17. Standard metadata name	optional
6. Character encoding	conditional	18. Metadata norm version	optional
7. Thematic category	mandatory	19. Metadata language	conditional
8. Spatial resolution	optional	20. Metadata character encoding	conditional
9. Summary	mandatory	21. Responsible for the metadata	mandatory
10. Distribution format	mandatory	22. Metadata date	mandatory
11. Time and altimetry extension	optional	23. Status	mandatory
12. Type of spatial representation	optional		

limitless tags, which facilitates online data exchange (Goldberg, 2009). An XML document can be considered well formatted if it matches what it prescribed in the norms (Bray et al., 2006). This document can be valid as long as it follows some norms described in its grammar. The XML Schema Definition (XSD) is an XML grammar format (Fallside and Walmsley, 2004). ISO 19139 is described as XML schemas built under the specification by Thompson et al. (2004).

3 RELATED WORKS

Filling out metadata is hard work, as is any other product cataloging.

The use of software to generate geographic metadata files in several formats that meet different norms has been proposed to improve geographic data documentation. Among these tools, the freeware ones such as CatMDEdit and GeoNetwork stand out, both featuring forms to fill out the metadata according to a pre-defined profile and enabling the automatic extraction of some characteristics from the dataset.

3.1 CatMDEdit

CatMDEdit (Catmdedit, 2015) is a tool to edit and visualize metadata in several standards that facilitates resource documentation, particularly geographic information. It is developed by the *Instituto Geográfico Nacional de España* (IGN) along with the Advanced Information Systems Group (IAAA) from the University of Zaragoza, with technical support from the GeoSpatiumLab (GSL). The tool was implemented in Java and has important features for metadata documentation, among which: The system is multiplatform (running on Linux and Windows); multilingual (Spanish, English, French, German, Polish, Portuguese, and Czech); open source, supporting the automated metadata file extraction and generation (Shapefile, DGN, ECW, FICC, GeoTIFF, GIF/GFW, JPG/JGW, PNG/PGW); and it converts and personalizes metadata standards to generate new metadata standards and profiles in order to serve all types of geographic data.

3.2 GeoNetwork

GeoNetwork (GeoNetwork, 2012) is a standardized decentralized environment based on a catalog system to facilitate geospatial data access, recovery,

update, and management. It provides a complete environment with metadata editor and catalog with search functions. It also carries an online interactive map viewer using web map service. It is currently used in countless SDI initiatives worldwide. Some of its main features include: (1) native support to the metadata standards ISO 19115, ISO 19139, FGDC, and Dublin Core, besides being able to configure a new metadata standard profile; (2) metadata synchronization among distributed catalogs; (3) user management and customized access control; (4) cataloging and access to several types of data and documents (upload/download); (5) interface with multilingual support; and (6) metadata importing in the ISO 19115 standard into a metadata profile configured in GeoNetwork. This system is free and open-source, which facilitates its evolution and customization by SDI developers.

What differentiates edpMGB from the editors mentioned in this section is that the former is a SaaS system that does not need to be installed on the user's machine, besides having a simplified interface that helps users document metadata. Moreover, it is the only editor specific for the MGB profile and its interface has the local advantage of being in Portuguese.

4 edpMGB - MGB PROFILE METADATA EDITOR

Metadata documentation in the MGB profile is an important task for them to be shared and reused. Up until now, no other editor specific for the MGC profile has been created, hence, each metadata set is specified according to the preferences of their authors (Pascoal et al., 2013).

edpMGB is a web application developed with the Google Web toolkit (GWT), a framework developed by Google for web-applications. The GWT used the java programming language to develop applications and the tool itself converts the JavaScript code so that the application can be interpreted by any web browser regardless of the platform the user is running.

Through edpMGB, the user can create, edit, and save metadata as a .xml file following international standards so it can be used in several geospatial tools. Before the XML file is generated, it must be validated according to the MGB profile rules.

One of the components of *GeoPortal Cemig* is the geospatial metadata catalog. Therefore, the metadata are documented through edpMGB, thus integrating SDI-Cemig to the INDE.

Figure 1 illustrates edpMGB’s home screen. The left-hand side has the navigation tree separated into sections and profile elements. The center-right area features the screen with the MGB profile fields divided into ten panels that represent the sections, which may be accessed using the green arrows to the right or to the left.

Figure 2 (a) shows the rules of enforcement, occurrence, and type of value of an element. The icon shown in Figure 2 (b) shows an element’s detailed information.

The bottom part of the home screen features the editor buttons panel. The button “Abrir” (Open) displays the dialog box where the user can load an XML file to be edited. Each element of the profile will be loaded in its respective text box.

The button “Limpar” (Clear) clears the open

element’s text box, while “Limpar Tudo” (Clear All) clears all text boxes of the metadata being edited.

The button “Validar” (Validate) performs one of the system’s main features, which is to validate whether the metadata is in accordance with the XML schema of the MGB profile. When the button is clicked, the system verifies the data input and may display, for instance, a dialog box as shown in Figure 3 to inform that the metadata does not respect the MGB profile’s rules. The dialog box displays an error log alerting the user and showing which fields are not in accordance with the rules. The user may keep editing the metadata or store it in his or her machine even if it is not validated for the MGB profile. When a non-validated metadata is generated, it is tagged informing it does not conform to the MGB profile.



Figure 1: edpMGB’s home screen.

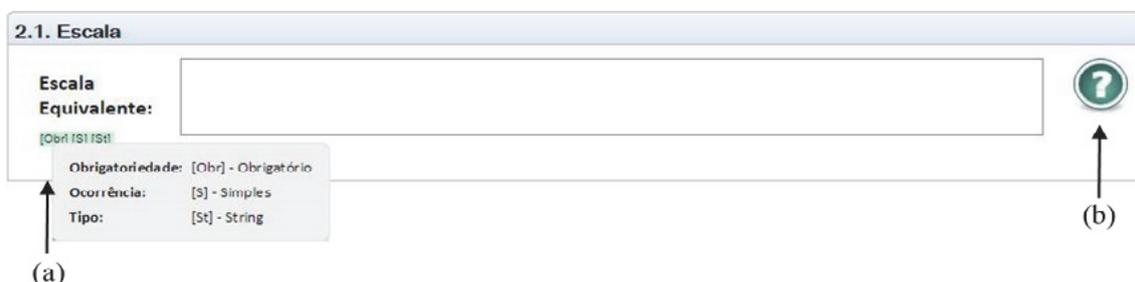


Figure 2: Warning dialog about mandatory elements.

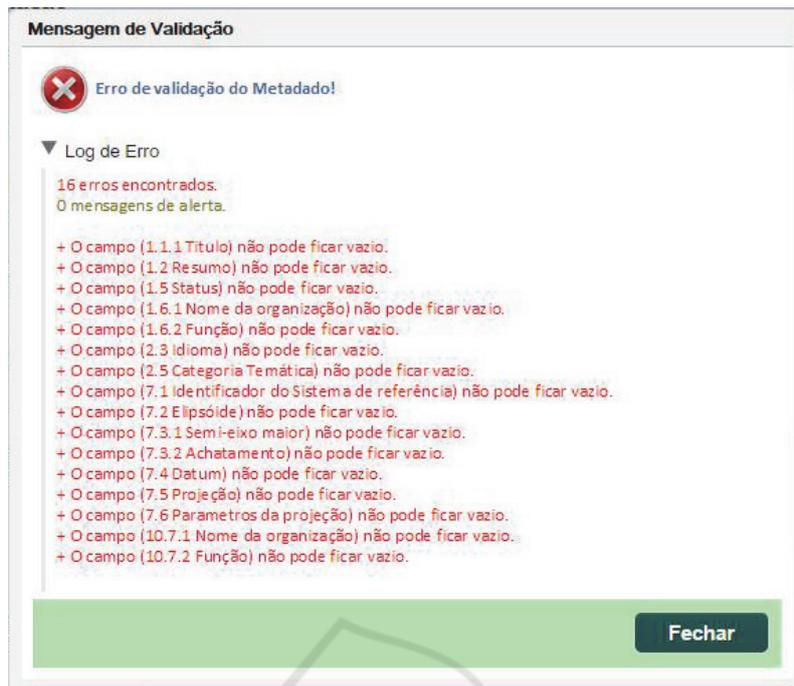


Figure 3: Result with errors found by the validation service.

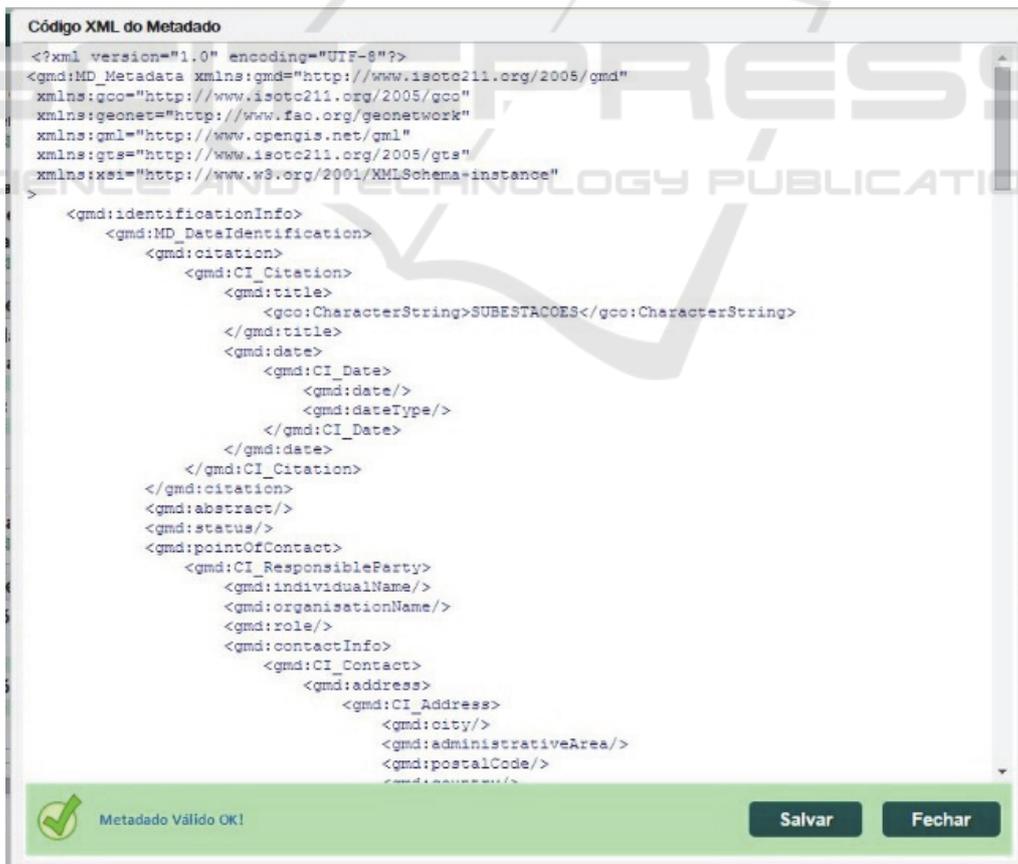


Figure 4: XML file generation screen.

A `generateXMLScript()` method receives as parameter the elements of the MGB profile, which are validated by the method `validateMGB()`, then the former outputs the XML script as a string.

The method `validateMGB()` validates the metadata according to the MGB profile, receiving its elements as parameter and outputting a list with the error messages found in the metadata or an empty list in case it respects the profile.

To Pascoal et al., (2013), most metadata provided by national data producers do not fully respect the profile's rules, which is a big issue since it compromises the interoperability among the systems that use the same profile. Nonetheless, the impossibility of saving a non-validated metadata may cause problems to users, perhaps due to the lack of information on the metadata elements. Hence, the user has the option of saving the metadata even if it does not conform to the MGB profile and, if needed, load it again in the tool for further editing. The "XML" button saves the metadata as a .xml file (Figure 4) that contains all information input in the fields. The .xml file is saved in the user's computer.

5 CONCLUSIONS

The present paper presented the development of a geographic metadata editor that follows the concept of Software as a Service (SaaS) and is available to any user with internet access. `edpMGB` was developed specifically to create geospatial metadata in accordance with the *Perfil de Metadados Geoespaciais do Brasil* (MGB profile). Besides using software developed in his or her native tongue, the user has easy access to the system with no need to install it locally. `edpMGB` may also be used to change metadata sets created with other editors (e.g., Geonetwork).

The feature of validating whether the metadata is in accordance with the standard defined by the MGB profile helps produce higher quality, more complete and correct metadata. However, the user may save the documents still incomplete, in which case the metadata receives a tag of non-conformity with the MGB profile.

Since the XML validation and generation were also developed as web services, other developers will be able to remotely create applications (e.g., metadata catalog manager) that use the services implemented in the tool through SOAP, which broadens the tool's reach.

As proved by Pascoal et al. (2013), most metadata sets currently available in the INDE do not

conform to the MGB profile. Therefore, this metadata editor with its XML schema conformity validation service is an important contribution to INDE's evolution.

This editor was developed in the context of the research and development (R&D) project "*GeoPortal Cemig*" being developed to help implement a corporate SDI for Cemig - Minas Gerais Power Company. One of the components of *GeoPortal Cemig* is the geospatial metadata catalog, whose metadata are documented via `edpMGB`, which integrates SDI-Cemig to the INDE.

Finally, being free open-source software, `edpMGB` may also be adapted to other geospatial metadata standards and/or profiles.

As future works, other features that will be implemented include: the development of a web service to index terms used in the metadata's documentation so that they can be recovered semantically from a vocabulary specific to the power sector, structured through a hyperbolic tree; integration of the editor with automated extraction modules of the rectangle involving the geospatial data; treatment of strongly related metadata collections; etc.

ACKNOWLEDGEMENTS

This project was partially funded by the Brazilian research promotion agencies Fapemig and CAPES, along with *Companhia Energética de Minas Gerais* - Cemig.

REFERENCES

- Armbrust, M.; Fox, A.; Griffith, R.; Joseph, A. D.; Katz, R.; Konwinski, A.; Zaharia, M., 2010. A View of Cloud Computing. *Communications of the ACM*, 53(4), pp. 50-58.
- Brasil. Decreto Presidencial nº 6.666, de 27 de novembro de 2008. Available at: http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2008/Decreto/D6666.htm. (in Portuguese)
- Bray, T.; Paoli, J.; Sperberg-McQueen, C. M.; Maler, E.; Yergeau, F., 2006. Extensible Markup Language (XML) 1.0. 4th ed. *World Wide Web Consortium (W3C)*.
- Catmdedit. CatMDEdit OpenSource Project. 2012. Available at: <http://catmdedit.sourceforge.net/index.html>.
- Concar. *Perfil de Metadados Geoespaciais do Brasil*, 2009. Comissão Nacional de Cartografia. Available at: http://www.concar.ibge.gov.br/arquivo/Perfil_MGB_F

- inal_v1_homologado.pdf. (in Portuguese)
- Dornelles, M. A.; Iescheck, A. L., 2013. Análise da aplicabilidade da Infraestrutura Nacional de Dados Espaciais (INDE) para dados vetoriais em escalas grandes. *Bol. Ciênc. Geod.*, sec. Artigos, v. 19, n. 4, p. 667-686. (in Portuguese)
- Fallside, D. C.; Walmsley, P., 2004. XML Schema Part 0: Primer. 2nd ed. *World Wide Web Consortium (W3C)*.
- GeoNetwork, 2012. "GeoNetwork User Manual: Release 2.6.4". Available at: <<http://geonetwork-opensource.org/manuals/2.6.4/eng/users/GeoNetworkUserManual.pdf>>.
- Goldberg, K. H., 2009. XML – Guia prático visual. Rio de Janeiro: *Alta Books*.
- ISO. ISO 19115:2003. Geographic information – Metadata. International Organization for Standardization (ISO).
- ISO. ISO 19139:2007. Geographic information - Metadata - XML schema implementation. International Organization for Standardization (ISO).
- Leme, L. A. P. P., 2006. Uma arquitetura de software para catalogação automática de dados geográficos. *Dissertação (Mestrado em Informática)* - PUC-Rio, Rio de Janeiro. 120 p. (in Portuguese)
- Lopez-Pellicer, F. J. et al., 2012. Towards an active directory of geospatial web services. In: *Bridging the Geographic Information Sciences*. Springer, p. 63–79.
- Nebert, D.D., 2004. Developing spatial data infrastructures: the SDI Cookbook, version 2.0. *GSDI-Technical Working Group*.
- Nebert, D.D.; Whiteside, A.; Vretanos, P. A., 2007. Open GIS® Catalogue Services Specification. Version 2.0.2, OGC 07-006r1. *Open Geospatial Consortium (OGC)*.
- Pascoal, A. P., Carvalho, R. B., e Xavier, E. M. A., 2013. Materialização do Perfil de Metadados Geoespaciais do Brasil em esquema XML derivado da ISO 19139. *XVI Simpósio Brasileiro de Sensoriamento Remoto*. (in Portuguese)
- Prado, B. R.; Hayakawa, E. H.; Bertani, T. C.; Silva, G. B. S.; Pereira, G.; Shimabukuro, Y. E., 2010. Padrões para metadados geográficos digitais: modelo ISO19115:2003 e modelo FGDC. *Revista Brasileira de Cartografia*, n. 62, v. 1, p. 33-41. (in Portuguese)
- Thompson, H. S.; Beech, D.; Maloney, M.; Mendelsohn, N., 2004. XML Schema Part 1: Structures. 2nd ed. *World Wide Web Consortium (W3C)*.
- W3C - World Wide Web Consortium. *Web Services Architecture*. <http://www.w3.org/TR/2004/NOTE-ws-arch-20040211/>. 2004.
- Weiss, A., 2007. Computing in the clouds. *netWorker*, 11(4), p. 16-25.
- Yang, C.; Goodchild, M.; Huang, Q.; Nebert, D.; Raskin, R.; Xu, Y.; Fay, D., 2011. Spatial cloud computing: how can the geospatial sciences use and help shape cloud computing?. *International Journal of Digital Earth*, v. 4, n. 4, p. 305-329.