

Dublin Core Usage for Describing Documents in Brazilian Government Digital Libraries

Diego José Macêdo
Instituto Brasileiro de
Informação em Ciência e
Tecnologia, Brasil
diegomacedo@ibict.br

Milton Shintaku
Instituto Brasileiro de
Informação em Ciência e
Tecnologia, Brasil
shintaku@ibict.br

Ronnie Fagundes de Brito
Instituto Brasileiro de
Informação em Ciência e
Tecnologia, Brasil
ronniebrito@ibict.br

Abstract

Digital libraries are increasingly common, being developed by government agencies to disseminate and preserve the documentation produced by its employees. This proposes a challenge in describing this type of documents, dealing official aspects in tools that are originally designed for bibliographic and scientific documents. In this sense, our objective is to verify how digital libraries, linked to the executive, legislative and judiciary Brazilian powers, are describing its documents collections. A study with descriptive and qualitative characteristics reveals the great adoption of DSpace software for creating these digital libraries and Dublin Core to describe the documents, showing DSpace and metadata schema adaptability for nonacademic document types. Thus, one contributes to the discussion on the use of Dublin Core to describe various types of documents on the Internet.

Keywords: Government digital library; Dublin Core; Government Agency.

1. Government Digital Libraries

With the change of the physical medium on paper for publication in electronic format, digital libraries have become the locus for preservation and access to documentation of an institution. With this, Brazilian governmental institutions created digital libraries in order to provide transparency to their activities, providing access to the full content of its documentation, creating a scenario where institutions use tools originally designed for the dissemination of scientific information in the dissemination of governmental information.

Many institutions have been using tools developed in free software, especially DSpace much by the support of the Brazilian Institute of Information Science and Technology (IBICT), which disseminates and supports this tool. This is also due to the government policy for free software adoption, which significantly changed the business of IT sectors, where development has been gradually replaced by adjustment of free tools.

This approach saves time and resources, since there is a large supply of free tools, with the most varied purposes. Some tools have a specific purpose and are being used for other purposes, such as DSpace, originally designed for academic repositories and used in other scenarios.

Another point to collaborate with the dissemination of government documents it related to the fulfillment of requisites defined by Law No. 527 of 18 November 2011, in which the agencies linked to the Brazilian government must make non-sensitive documents freely available. This law guarantees Brazilian population unrestricted access to governmental documents, regardless of support, encouraging the use of tools that support the digital distribution of documents, such as digital libraries.

Digital libraries are a dedicated tool for dissemination of scientific and technological documentation and have flow and structure aimed at managing these documents, which has well established forms of classification and cataloging. A challenge is posed to librarians, archivists and documentation developers in the description of processes of governmental documents in digital libraries.

In the bibliographical studies, there are few studies regarding government documentation, to the extent that many researchers classify them as archival documents. However, manuals, technical reports and other documents of institutional memory have bibliographic aspects, but not always receive adequate treatment in government agencies. So, these documents are not always disseminated, even with relevant information that could be reused.

In this context, the present study aims to analyze the use of Dublin Core metadata schema in the description of documents in digital libraries developed with DSpace and linked to Brazilian government agencies belonging to the executive, legislative and judicial branches. It analyzes their document's metadata and describes the strategies used for the representation of their collections.

2. Methodology

The study has descriptive characteristics, that is aimed to characterize populations or phenomena and suitable to describe scenarios (Gil, 2006). In line with the objective of analyzing the use of Dublin Core in the description of government documents, the research provides a survey of the Brazilian scenario, following the guidelines of descriptive research.

It has a predominantly qualitative approach, more appropriate to the social studies as stated by Richardson (2008). The depth of qualitative analysis is justified in so far that the study transcends usage verification. However, has collection of quantitative data, where quantitative data are analyzed qualitatively (Creswell 2007).

The research objects are the digital libraries linked to the government agency, in which the variables are the descriptive elements. Thus, the used elements and qualifiers of Dublin Core are accounted, so it is possible to compare and analyze the results.

3. Results

The study identified 13 digital libraries linked directly with Brazilian government agencies, all designed with DSpace, as shown in Table 1, providing more than 427,000 documents in full text. Thus, there are four libraries from the executive power, five from the judicial and four of the legislative branch. This reveals the interest of the Brazilian government agencies in the use of DSpace, which was developed primarily for the development of academic systems. The Digital Library of Housing (Biblioteca Digital da Habitação - HABI) from São Paulo is fully restricted, preventing outside access to their documents, so staying out of the research.

TABLE 1 – List of analysed digital libraries

Branch	Agency Government	Library Name	Records	URL
Executive	Ministério do Planejamento, Orçamento e Gestão	SPI - Biblioteca Digital do Planejamento	494	http://bibspi.planejamento.gov.br
Executive	Ministério Público Federal - MPF	Biblioteca Digital do MPF	21.335	http://bibliotecadigital.mpf.mp.br/xmlui
Executive	Secretaria Geral da Presidência da República	Biblioteca Digital da Participação Social	395	https://biblioteca.participa.br/jspui
Executive	Prefeitura de São Paulo	Biblioteca HABI	11.322	http://biblioteca.habisp.inf.br
Judiciary	Tribunal Regional Federal da 1ª Região	Biblioteca Digital TRF1	44.977	http://www.trf1.jus.br/dspace

Judiciary	Superior Tribunal de Justiça - STJ	Biblioteca Digital Jurídica -STJ	76.124	http://bdjur.stj.jus.br
Judiciary	Tribunal de Contas do Município do Rio de Janeiro	Biblioteca Virtual em Controle Externo	150	http://bvce.tcm.rj.gov.br
Judiciary	Tribunal de Justiça do Estado do Ceará - TJCE	Biblioteca Digital Jurídica -TJCE	502	http://bdjur.tjce.jus.br/jspui/
Judiciary	Tribunal Superior do Trabalho - TST	Biblioteca Digital do Tribunal Superior do Trabalho	8.626	http://aplicacao.tst.jus.br/dspace
Legislative	Senado Federal	Biblioteca Digital do Senado Federal	262.210	http://www2.senado.leg.br/bdsf
Legislative	Câmara dos Deputados	Biblioteca Digital da Câmara dos Deputados	3.516	http://bd.camara.leg.br/bd/
Legislative	Câmara Legislativa do Distrito Federal	Biblioteca Digital da Câmara Legislativa do Distrito Federal	48	http://biblioteca.cl.df.gov.br/dspace
Legislative	Assembléia de Minas	Biblioteca Digital da ALMG	13.372	http://dspace.almg.gov.br/xmlui

A point to note is that out of the 13 libraries selected for analysis, only three provide interoperability via Open Archives Initiative - Protocol Metadata Harvesting (OAI-PMH), even if using DSpace software, a system where this option is very easy to implement. This indicates poor adherence to the precepts of open files, an political issue, as these institutions do not have the same concern with institutional visibility as academic institutions have.

The poor adherence to interoperability by government repositories can be explained by the absence of a federation to join all these repositories, such federation could offer services such as consolidated searches on government digital documents. Thus, it requires that repositories make available the OAI-PMH in order to establish interoperability, revealing certain isolation between government repositories.

Regarding the executive branch, it proves to be present in the various levels of public action, with repositories linked to the Presidency library up to city halls. Emphasis is on the Digital Library of Social Participation, created in 2014, linked to the General Secretariat of the Republic Presidency (Secretaria Geral da Presidência da República), focused on the dissemination of government documents on social participation in government actions. This digital library is linked to the higher Brazilian administrative level.

The judiciary has the highest amount of digital libraries, at the various hierarchical levels of power. Emphasis on the Digital Library Legal (Biblioteca Digital Jurídica), developed by the Superior Court of Justice, being the first nonacademic Brazilian institution to make use of DSpace for creating an information system, in operation since 2005. This library has stimulated the use of DSpace in other legal institutions, with support from the Brazilian Institute of Information Science and Technology (IBICT).

The Legislative, in turn, has digital libraries in the Senate and the House of Representatives, revealing the adherence of this tool by higher levels of legislative organs. In the Library of the Senate highlights the collection of articles in newspapers and magazines, in order to preserve its institution memory through this documentation. In the Digital Library of the Federal Chamber, highlights are to the historical documents of the Brazilian Republic. Together these two libraries provide over 250 thousand documents.

In digital libraries developed with DSpace, the classification process is presented in the form of organizing the collection in communities, sub-communities and collections. At this point, it appears that most digital libraries categorize documents by document type (six libraries), followed by the organ activities (two libraries) and to the organizational structure of the agency (one library). There are libraries that present joint document type/organizational structure categorizations (three libraries). There is no standardized form of collection organization, with only one digital library, the Biblioteca Digital da Participação Social, organized by thematic taxonomy of the organ.

Even with minor variations, all repositories use the qualified Dublin Core metadata schema, despite DSpace's flexibility to use other schemes, which shows the adaptability of Dublin Core to describe a variety of document types. There was a wide variation in the use of metadata, not only on the amount used, but also on the elements and qualifiers. The Digital Library of the Regional Court of Ceará (Biblioteca Digital Jurídica - TJCE), for example, uses only 11 different metadata fields to describe the documents, while the Digital Library of the Superior Labor Court uses 43. This variation reveals little standardization in the description of the documents, as these two libraries are from the judiciary branch.

The most commonly used elements in all libraries are dc:contributor, dc:date, dc:identifier and dc:title (Figure 1). The Digital Library of the Superior Labor Court adds elements and qualifiers Electronic Thesis and Dissertation - Metadata Standard (ETD-MS), as it contains theses and dissertations in its collection. The Biblioteca Digital do Tribunal Superior do Trabalho created an element called dc:atos, to contain the identification of documents called "act", the only new element identified.

The research has revealed that dc.description is the metadata element used with more different qualifiers in these repositories. This can be explained because when there isn't a specific element to describe a digital object characteristic, many repositories' managers use the flexibility of this element on the description. Also, element dc.date is used the same way, as there are a lot of dates to describe a digital object, like creation date, submission date, publication date, and so on. Another point is about dc element identifier, usually a digital object has a unique identifier but in repositories there can be noticed two identifiers, URL and digital object own identifier.

This findings contrast in part with Alijani and Jowkar's (2008) research results, highlighting the differences between academic digital objects and governmental digital objects. In fact for academic digital objects, title element is very important, but in governmental documents the description is as important as the title, as far as in some cases governmental digital objects's title is sometimes irrelevant.

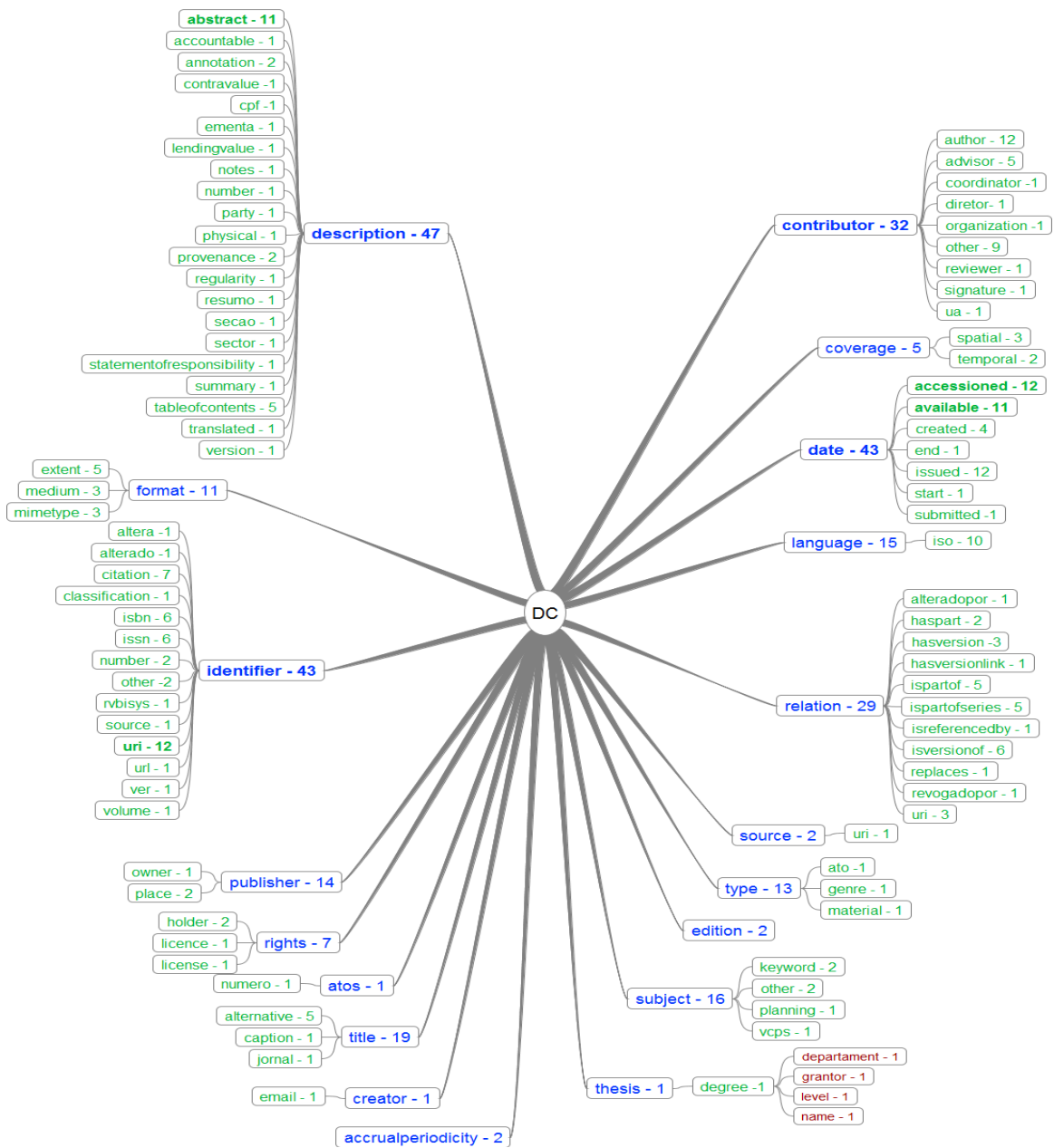


FIG. 1. Number of times that the elements and qualifiers are represented.

All digital libraries are using the author field (dc.contributor.author), title (dc.title) and date of publication (dc.date.issued), being these the most frequent, followed by summary (dc.description.abstract) and editor (dc.publisher), that does not appear in one library, the Biblioteca Digital TRF1. Another point is that 43 metadata fields are used by only one institution, the Tribunal Superior do Trabalho, indicating low standardization or specific needs to describe its documents.

The wide range of qualifiers can be highlighted in Figure 2, which presents the use of qualifiers per element. Noteworthy is the large number of qualifiers of elements dc:description, dc:identifier, dc:relation; dc:contributor and dc:date, noting that the description of government documents takes place in these elements.

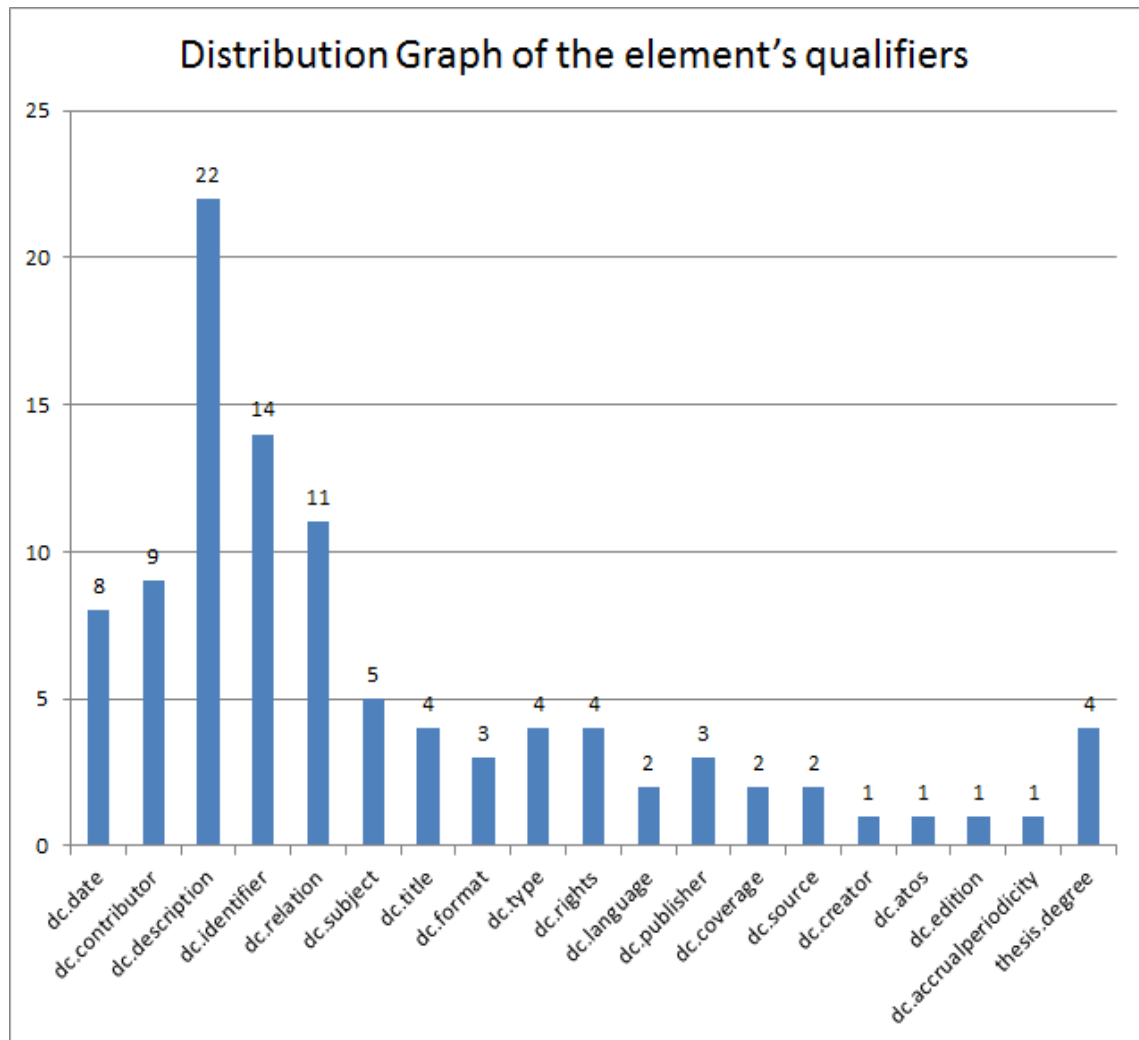


FIG. 2. Distribution Graph of Element's Qualifiers

As for the qualifiers of elements can be highlighted:

- The use of dc:description to describe the characteristics of documents that do not have elements provided in DC, in many cases, creating qualifiers to the description;
- As government documents have specific identifications, the large amount of dc:identifier qualifiers;
- The use dc:relation to indicate the various types of relationship between government documents;
- Government documents have lots of contributors, so lots of qualifiers;
- Dates are important in government documents, so the large number of qualifiers in dc:date.

This shows specific requirements of government documents in front of traditional academic digital libraries, even though in a few cases certain discrepancy in the understanding of elements, qualifiers and its content. However, as interoperability is not a concern on these libraries, this is not a big problem.

3. Final Remarks

The results revealed interest of government institutions on DSpace in the construction of government digital libraries, in part by the action of IBICT for the dissemination and support offered to user's community, even for non-academic institutions.

Also, the study found that government libraries do not use taxonomies related to its area to organize documents, preferring to use document types or organizational structure. As not offering interoperability there is not a concern with standardization of metadata fields, making use of wide variation due to the documentary specificity.

Also arise perspectives for the study of government documents classification in digital libraries and repositories, in order to facilitate its organization and retrieval, using government related taxonomies for example, and supporting the organization of this type of documents on the web.

The use of the dc:description elements can be observed to adapt Dublin Core to describe the government documents, revealing the flexibility of this metadata schema for describing a varied documentary typology. This point may be evidence of the need for studies for the proposal of more specific elements or qualifiers for these type of documents in the context of Brazilian government.

In addition, an analysis of the users of government repositories, their expectations, experiences and requirements regarding what they seek in the repositories can guide the planning and preparation of metadata application profile.

An increased number of libraries, archives or other initiatives on the Internet using the Dublin Core to describe the documents present challenges and opportunities studies. In Brazil, this is a promising scenario as a recommendation, all government documents must be accessible ads defined by the Information Access Act, No. 12,527, from November 18, 2011.

References

- Alijania, A. S.; & JOWKAR, A. (2009). Dublin Core metadata element set usage in national libraries' web sites. *The Electronic Library*, 27(3), 441-447.
- Creswell, J. W (2007). *Projeto de pesquisa: método qualitativo, quantitativo e misto*. Tradução de Luciana de Oliveira da Rocha. 2. ed. Porto Alegre: Artmed.
- Gil, A. C (2006). *Como elaborar um projeto de pesquisa*. 4. ed. São Paulo: Atlas.
- Richardson, J. R (2008). *Pesquisa social: métodos e técnicas*. 3. ed. São Paulo: Atlas.