

Theses and Dissertations Digital Library: Ten years of Open Access and Open Archives in Brazil

Diego José Macedo
diegomacedo@ibict.br

Ronnie Fagundes de Brito
ronniebrito@ibict.br

Washington L. R. de Carvalho Segundo
washingtonsegundo@ibict.br

Milton Shintaku
shintaku@ibict.br

Instituto Brasileiro de Informação em Ciência e Tecnologia - Ibict
<http://www.ibict.br/>

February 3, 2014

Abstract

Open Access and Open Archives are approaches for scientific information dissemination that can be supported by different software platforms. These platforms need to be integrated if a repositories federation is desired. However this integration deals with different metadata schemes that need to be normalized and integrated in order to allow information systems to share data. The Brazilian Digital Library of Theses and Dissertations is an example of integration such systems and it is presented in this paper. Its underlying system's architecture is described, as well as the type of documents stored, the amount of participant repositories along time and its geographical distribution. Finally, the Brazilian Digital Library of Theses and Dissertations is shown as an example of integration of Brazilian scholar production in federated repositories.

Keywords: Repositories Integration, Interoperability, Theses and Dissertation Repositories, Brazilian Repositories, Discovery and Delivery Services.

1 Introduction

The Open Access and the Open Archives movement mingle in many cases, even if they have clear distinctions, sharing similarities and the same abbreviations (OA). Suber (2012) [3] considers Open Access the literature available in digital format and with no access restrictions. Open Archives in turn, can be defined as a set of implementations that enable interoperability, as reported Triska and Café (2001) [4] on the Santa Fe Convention in 1999. However, these concepts are not mutually exclusive and many of the systems that adopt Open Access also adopt Open Archives.

In this context, the Brazilian Digital Library of Theses and Dissertations (BDTD) is a federation of repositories exclusive for theses and dissertations. It is a joint effort of the Brazilian Institute of Information in Science and Technology and Brazilian universities with the aim of creating a network that enables interoperability of metadata, in order to provide a national portal with consolidated services [2].

2 BDTD Architecture

The BDTD consists of three sub-systems: the local repositories of theses and dissertations; the harvesting system; and the BDTD National Portal (Figure 1). The portal provides access to theses and dissertations, which were regarded as gray literature, due to difficulty of access to such documents, as highlighted by Schöpfel and Farace [1].

The local repositories of theses and dissertations are held by universities, who manage the systems, feeding it with theses and dissertations from these institutions. Local repositories implement a OAI-PMH protocol layer that responds to harvesting, making them data providers. The system collects metadata from local repositories, and also offers collected metadata to other harvesting systems, becoming both service provider and data provider. The national portal indexes the collected metadata to offer discovery and delivery services.

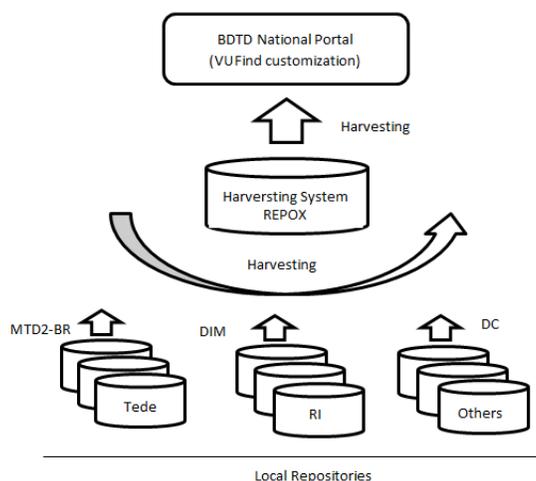


Figure 1: BDTD Architecture

Historically, BDTD was the first system in Brazil to implement the precepts of open archives. Launched in mid 2002, BDTD plays an important role in the dissemination of Open Access in educational institutions, enabling access to the full text of theses and dissertations.

3 Local Repositories of Theses and Dissertations

Local repositories are systems that operate via web for the deposit, access and preservation of theses and dissertations. In 2002, only three institutions made up the BDTD network, all with their own systems that implemented a OAI-PMH protocol layer. In 2006 Ibict developed and made available a system for the management of theses and dissertations called TEDE, which was widely accepted by the institutions. In 2009 Ibict initiated a program to encourage the creation of institutional repositories, adding theses and dissertations in DSpace-based systems.

In this context, the BDTD currently consists of 104 institutions, where 83 institutions using the TEDE, 14 institutions with DSpace-based repositories and 7 institutions with proprietary systems or being collected from the On-line Public Access Catalog (OPACS). Thus, a total of 104

local repositories, with more than 274,600 documents: 73,000 Doctoral Theses and 201,600 Master Theses.

Local repositories provide metadata in multiple schemes, such as BAT and MTD2-BR (Metadata for Theses and Dissertation - Brazilian), with 226,514 registrations, DC (Dublin Core) with 10,783 records, DIM (DSpace Intermediate Metadata) with 18,990 records, MARCXML, with 18,313 records. Thus, BDTD allows the adherence of various types of systems, operating with different metadata schemes.

Geographically, the repositories are present in most Brazilian states, only the states of Tocantins, Acre, Piauí, Mato Grosso and Amapá has no representatives in the network. There are the southern and southeastern regions, with almost 70% of the repositories, according to the distribution of universities in Brazil, which has some concentration in these regions (Figure 2).

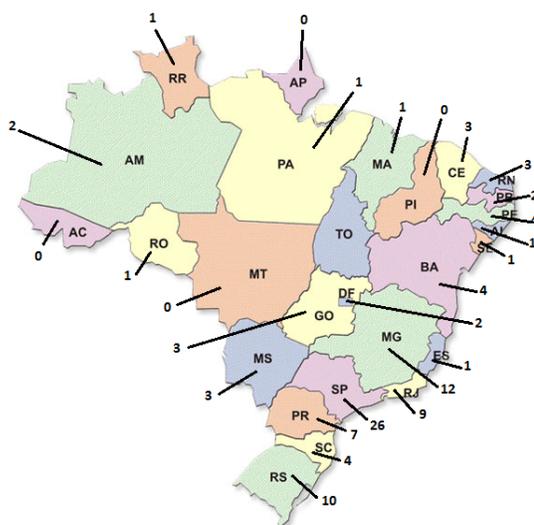


Figure 2: Repositories Concentration

4 Harvesting System

The harvester system was developed based on REPOX software with various settings to meet the needs of BDTD. Due to the possible different schemes of metadata, crosswalks were developed to turn various metadata schemes into a simpler new scheme. This scheme is based on the Qualified Dublin Core (QDC) with the addition of specific fields to describe Theses and Dissertations.

The system can also respond to other harvesting systems, acting as an aggregator: collecting local repositories and delivering data to other service providers requests. Therefore, crosswalks were implemented to respond to various forms of metadata, such as for DC, ETD-MS (Electronic

Theses and Dissertation - Metadata Schema) and others.

Another highlight is the normalization of data implemented by the harvesting system, presented as a major problem for the generation of indicators and retrieval based on metadata. Thus, the system maps the contents of metadata to normalize more restricted fields, like `dc.type`, that indicates the type of document, which in this case can only be thesis or dissertation, eliminating all possible variations, promoting standardization of the data that supports recovery.

In this sense, the harvesting system not only collects but also provides services related to the metadata and its contents. Thus, delivers to the national portal, metadata normalized on a single schema.

5 BDTD National Portal

The BDTD National Portal is an implementation done with the VUfind software customization, with added functionalities and graphical modifications. The portal offers consolidated services like search, refinement by faceting, navigation through the contents of the collection, among other forms of recovery offered by web scale discovery software. So, allows the search and recovery of theses and dissertations from 126 universities.

The portal also supplies indicators about its collections, as it receives the normalized metadata. In this sense, BDTD can be an important source of Brazilian research information giving a panorama about the masters and doctoral programs.

6 Conclusion

The BDTD is an example of how a federation of repositories can be implemented in a system with different softwares. Seamlessly integrating data from independent systems, based on the precepts of Open Archives and Open Access, enabling access to the full text of theses and dissertations. There is a concentration of documents from the southeastern and south of Brazil, therefore this federation of repositories can provide an integration between regions less and more developed, converging different technological infrastructures in an unique portal.

References

- [1] D. J. FARACE and J SCHÖPFEL. Grey Literature in Library and Information Studies. *Walter de Gruyter*, 2010.
- [2] S. B. SOUTHWICK. The Brazilian electronic theses and dissertations digital library: providing open access for scholarly information. *Ciência da Informação*, 35(2), 2006.
- [3] P. SUBER. *Open Access*. MIT Press, 2012.
- [4] R. TRISKA and L. CAFE. Arquivos abertos: subprojeto da Biblioteca Digital Brasileira. *Ci. Inf.*, 20(3), 2001. In: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652001000300012&lng=en&nrm=iso Accessed: 2014-15-01.