

The LA Referencia Software and the Brazilian Portal of Scientific Open Access Publications (oasisbr)

Washington Carvalho-Segundo, IBICT, washingtonsegundo@ibict.br;

Lautaro Matas, LA Referencia, lmatas@gmail.com;

Alberto Cabezas, LA Referencia, alberto.cabezas@redclara.net;

Bianca Amaro, IBICT, bianca@ibict.br; Gabriel Gomes, IBICT, gabrielgomes@ibict.br

Session Type • Presentation.

Abstract

After 8 years of the creation of the Brazilian Portal of Open Access Scientific Publications (oasisbr) this system just had added a small part of brazilian contingent in terms of open access scientific documents, just 440 thousand records. In the end of 2014, the IBICT team started a project of a technological update of oasisbr, with a replacement of its harvester, metadata validator, transformer, provider and meta-searcher. That project finished in July 2015, and the result was a new website, based on a platform provided by LA Referencia. The LA Referencia software brought a smooth metadata transformation and a validation UI, and harvesting statistics. After the technological update, the number of harvested open access sources was gradually growing which have added more than 1 million and 700 thousand documents between articles, theses, dissertations and others scientific documents.

Conference Themes

- Supporting Open Data and Open Science
- Repositories of high volume
- Integrating with the Wider Web and External Systems

Keywords— Meta-searchers; Harvesting systems; Institutional repositories; Repositories network; LA Referencia; oasisbr.

Audience: This text might be of interest of repository managers, developers, data producers and librarians.

1 Background

The popularization of Internet occurred in Brazil between the late 90s and the early 2000s, following a trend of emerging Latin American countries. This fact has created opportunity for greater access and explosion of demand for organization of information services provided by governmental agencies [4].

At the same time that poor population of Latin American have made their first access to the global network, the developed world had already started the discussion of the Open Access Movement, by declarations and manifests: Santa Fé (1999), Budapest (2002), Berlin (2003) and Bethesda (2003). The brazilian project of open access started just in 2005, also presented

by manifests, open letters and declarations, are examples: the Brazilian Manifest of Support to Open Access to Scientific Information (2005), the Salvador Declaration (2005), the Letter of São Paulo (2005), the Declaration of Florianópolis (2006) and the Memorandum of Understanding between Brazil and Portugal for the promotion of the Open Access to Scientific Publications (2009).

The need for a coordinated action that would join the local efforts of universities, research institutes, governmental organisms and foundations, with their electronic journals, digital libraries of theses and dissertations, and institutional repositories had been born. This network composition have been very like the Bjork list [1] of open access sources, which is: IRs providers, electronic theses and dissertations, scientific periodicals, research-area-specific archive (e-print) servers and author's personal home pages.

The Brazilian approach had been a creation of a metadata harvester and aggregator system which joint different open access providers all over the country, and culminated, in 2006, on the Brazilian Portal of Open Access Scientific Publications (oasisbr) [3]. The portal have been implemented by the Brazilian Institute of Information in Science and Technology (IBICT), with support from the Sponsorship National Agency of Studies and Projects (FINEP).

The system used some well established tools, like a harvester module with OAI-PMH (Open Archives Initiative - Protocol for Metadata Harvesting) and a meta-searcher above the IBM OmniFind platform. However, there had been a lot of problems such that invalid character encoding, deviations from the international metadata guidelines, redundant records, multi-standard metadata schema and difficulties with the search engine configuration.

2 Content

After 8 years of the creation of the Brazilian Portal of Open Access Scientific Publications (oasisbr) this system just had added a small part of Brazilian contingent in terms of open access scientific documents, just 440 thousand records [5] between articles, theses and other scientific publications. One has to note that the 440 thousand records also had included Portuguese publications that had been harvested from RCAAP portal [8] (in that time that portal had around 200 thousand documents).

In the end of 2014, the IBICT team started a project of a technological update of oasisbr, with a replacement of harvester, metadata validator, transformer, provider and meta-searcher. That project finished in July 2015, and the result was a new website, based on a platform provided by LA Referencia [6, 7]. This organization is the network of open access repositories from nine Latin American countries. It supports national open access strategies in Latin America through shared standards and a single discovery platform.

LA Referencia also harvests scholarly articles and theses & dissertations from national nodes (in the case of Brazil the national node is oasisbr), which, in turn, harvest from repositories at universities and research institutions in each country. LA Referencia is an initiative that came out of the technical and organizational agreements between public science and technology organizations (National Ministries and Science & Technology Departments) with RedCLARA (the organization that manages the high speed network in Latin America) to provide a common network for open access publications.

The LA Referencia software brought a smooth metadata transformation/validation user interface and harvesting statistics. In particular, the validation system allow local administrators to have a diagnosis of the metadata quality of their sources. And the integration with VuFind software created a responsive search interface based on a CSS Bootstrap framework, above of an efficient search engine, that is the Apache-Solr application.

After the technological update, the number of harvested sources was gradually growing

and today one has 370 institutions that hosts 1 scientific national repository, 69 institutional repositories, 114 electronic theses and dissertations databases and 712 electronic journals. These add a total of **1 million and 700 thousand documents**, whose the distribution by document type is showed in the chart of Figure 1.

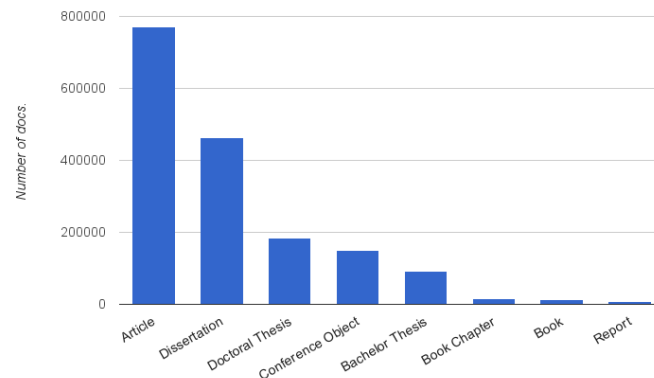


Figure 1: Distribution by document type

Moreover, it had been possible to get harvesting statistics, between Jun/2015 and Aug/2015, and so one can measure the growing of the number of documents, which was relatively fast in the beginning, because many institutions that could not participate due technical problems in the harvesting system became new tributaries of the network. So, until January of 2016 the increment of the number of aggregated documents was soft. Between January and February of 2016 there was a rising gap of more than 200 thousand documents. This was due to the entry of hundred electronic journals in the network, with a comprehensive review in the harvesting sources. After that, the growing have been kepted in a gradual manner as shows Figure 2.

The gradual growing was related with the contributions of the brazilian sources and the RCAAP (Open Access Scientific Repository of Portugal) as shows the chart of Figure 2. Particularly, the sequence “OTHERS” in Figure 2 represents mostly the contribution of the electronic journals. As explained in the previous paragraph, one can observe a raising gap in that sequence between January and February of 2016.

The new software also brought an access statistics mechanism which allowed the plot of the total of visited pages in the period of June of 2015 to November of 2016 for the top most accessed institutions, as displayed in Figure 3.

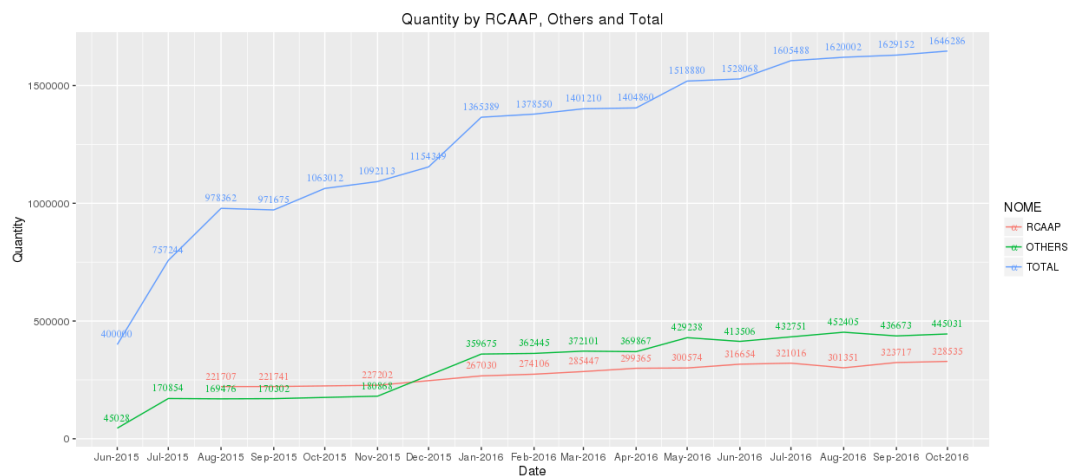


Figure 2: Evolution of the total number of documents in oasisbr

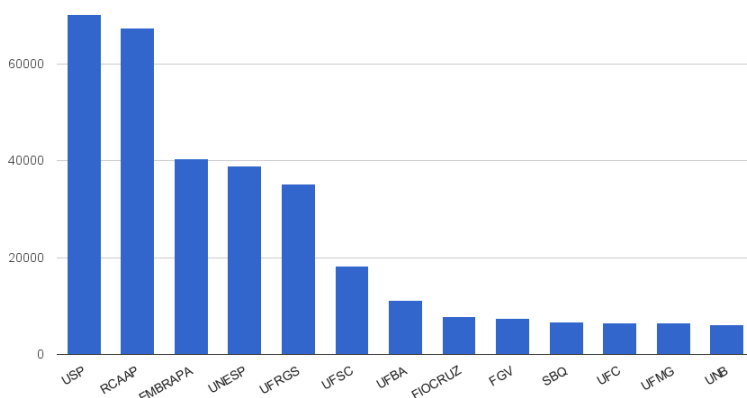


Figure 3: Number of visits for the top most accessed institutions in oasisbr

3 Conclusion and future work

The existence of a single national portal of open access scientific publications enhances the availability and visibility of the regional scientific production and strengthens the link between countries that has Portuguese as native language, by the mirroring between RCAAP and oasisbr. Another result of having a well-established national portal is the enriching of Latin American on the open access engagement by coordinated actions of the LA Referencia network.

By technological point of view, the software package provided by LA Referencia, was an elegant answer to the demand of building a robust harvester and meta-searcher. It created a stable and well-coordinated platform that resulted, in the case of oasisbr, on a growing to more than a four times the original number of open access indexed records.

LA Referencia and IBICT together have already started a project of improvements in the LA Referencia software, that will include, between other things, an extension of the metadata standard to better fit the description of theses and dissertations, a full text indexing process and a record deduplication mechanism.

References

- [1] BJORK, Bo-Christer, *Open access to scientific publications - an analysis of the barriers to change?*, Information Research, Vol. 9, No. 2, January 2004.
- [2] BOAI. *Budapest Open Access Initiative*. Available in: <<http://www.budapestopenaccessinitiative.org>>. Accessed in : 2016.
- [3] BRAZIL. IBICT. *Brazilian Portal of Scientific Open Access Publications (oasisbr)*. Available in: <<http://oasisbr.ibict.br>>. Accessed in : 2016.
- [4] NDOU, Valentine, *E-government for developing countries: opportunities and challenges*, EJISDC, Vol. 18, No. 1, Pages 1-24.
- [5] SILVA, Ana P., ALENCAR, Lucidio L. and SOUZA, Jaqueline S., *Iniciativas de Acesso Aberto à Informação Científica: Uma abordagem sobre o Portal Brasileiro de Acesso Aberto à Informação Científica - oasisbr*, In: SNBU, Belo Horizonte, 2014.
- [6] CHILE. RedCLARA. *LA Referencia: Red de repositorios de acceso abierto a la ciencia*. Available in: <<http://lareferencia.redclara.net>>. Accessed in : 2016.
- [7] CHILE. RedCLARA. *LA Referencia Software*. Available in: <<https://github.com/lareferencia/main>>. Accessed in : 2016.
- [8] PORTUGAL. UMinho/FCT. *Open Access Scientific Repository of Portugal (RCAAP)*. Available in: <<http://www.rcaap.pt>>. Accessed in : 2016.