

Federated Networks of Open Access Repositories in Mexico and Latin America

Rosalina Vázquez Tapia, alinavn@uaslp.mx

Autonomous University of San Luis Potosí

Antonio Felipe Razo Rodríguez, antrazo@gmail.com

University Corporation for the Development of Internet CUDI, Mexico
Iberoamericana University Puebla

Abstract

Open Access to scientific literature through repositories has grown significantly in recent years, increasingly favoring the creation of federated networks at national or regional level. In November 2012, nine countries in Latin America signed an agreement to develop the Federated Network of Institutional Repositories of Scientific Publications, LA-Referencia (<http://www.lareferencia.info>).

From a framework of agreements, member countries developed their national network under a common interoperable infrastructure. The participation of Mexico is represented by the Mexican Network of Institutional Repositories – REMERI (<http://www.remeri.org.mx>), developed in 2012 by a group of six universities with public funding. To date (February 2015), a total of 108 Institutional Repositories of 61 Mexican Institutions of Higher Education have joined *REMERI*, with more than 430,000 documents. In this proposal, we describe some strategies adopted for the standardization of the repositories, technical requirements for interoperability between federated networks, technological developments for harvesting, indexing, normalization, search and retrieval of digital documents and finally some recommendations for maintenance and long-term sustainability.

Conference Themes

- Re-using Repository Content

Keywords

Open Access, Institutional Repositories, Federated Networks, Interoperability

Audience

Repository managers, developers, data producers, librarians and researchers.

OR2015 | 10th International Conference on Open Repositories

June 8-11, 2015, Indianapolis, Indiana, USA

Background

The Mexican Network of Institutional Repositories (REMERI) was created in 2012 by a group of six universities with public funding from Mexican National Council of Science and Technology (CONACYT). The first goal of this project was the creation of a common interoperable infrastructure of Mexican digital repositories for interconnection with federated networks. During the first stage we developed our own tools for harvesting, indexing and querying metadata records from digital repositories using interoperable international guidelines based on DRIVER.

Actually, REMERI is being developed by a General Coordinator, a Technical Manager and supporting staff and it is funded by the University Corporation for Internet Development in Mexico (CUDI for its acronym in Spanish). REMERI offers the following services: a Directory of Institutional Repositories, Institutional Scientific Production Indicators, an INDIXE of Mexican Open Access Journals, an INDIXE of Mexican Theses and Dissertations and an INDIXE of Documentary Heritage (in development).

Presentation content

The Mexican Network of Institutional Repositories, REMERI, is a federated network of institutional and thematic repositories of Mexican Institutions of Higher Education and Research Centers, collects, integrates and promotes open access scientific, academic and documentary production. REMERI is a nationwide aggregator repository containing references (metadata) to digital repositories and open access journals of Mexico .

Technical Infrastructure of REMERI, architecture and functionality of INDIXE System

The technology platform of REMERI consists of a web portal and a harvester-aggregator named INDIXE, developed specifically for the project with XML technology (XQuery / XMLDB) using an open source platform (eXist / Tomcat).

INDIXE was developed in-house, a solution was needed to handle metadata inconsistencies (incorrect or missing elements), partial or incorrect implementation of the Open Archives Initiative Protocol and also there was a requirement to implement a DRIVER compliant data provider.

INDIXE stores and process the metadata in the XML format. Metadata harvesting, normalization, integration, search and retrieval is implemented using the XQuery language. The database (eXist) indexes the collection with Lucene using a combination of vector-space and boolean models. The solution is scalable, compact, efficient, and multiplatform. INDIXE is ideal for thematic, regional or national repositories (aggregators). INDIXE system architecture is shown in Figure 1.

INDIXE implements the following services and tools:

1. OAI-PMH validator for metadata providers
2. Collection harvester
3. Metadata normalization and integrator
4. Metadata search and retrieval
5. OAI-PMH data provider for the collection
6. DRIVER compliant data provider for LA-Referencia

OR2015 | 10th International Conference on Open Repositories

June 8-11, 2015, Indianapolis, Indiana, USA

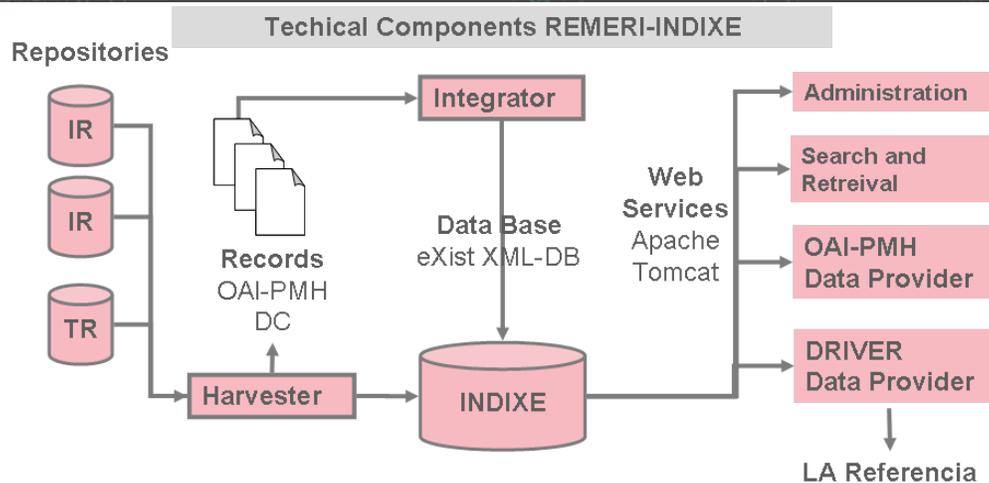


Figure 1. Technical components diagram

Interoperability

The experience in this project has allowed us to identify common problems in different elements of repositories Dublin Core metadata, such as:

dc : identifier

It is common in the case of the DSpace to show the identifier with a handle when this is not active, other repositories use the server IP or the term "localhost".

dc : type

It is common to find the type "other" or records without type. Some times we have to analyze the collections in order to assign the correct type.

dc : date

It is common to find more than one occurrence for dates, we try to identify the record publishing date.

dc : publisher

In the case of theses and dissertations, when the publisher is not mentioned, the institution provides the title which is assigned to the metadata.

Conclusion

In the case of *LA- Referencia*, REMERI was part of the technical tests since October 2012. REMERI is the national network that incorporates the largest amount of Spanish records to LA-Referencia with 111,637 (second only after Brasil in total records) It totally complies with LA-Referencia requirements based on DRIVER interoperability guidelines. Furthermore, REMERI was the first national network to meet specific technical metadata guidelines, including the institution and repository names for each record.

Now (May,2015), REMERI integrates information from: 61 Mexican institutions and research centers with 108 institutional and thematic repositories that contain more than 430,000 documents including research papers, bachelor, master and doctoral theses, videos and presentations mostly in Spanish.

OR2015 | 10th International Conference on Open Repositories

June 8-11, 2015, Indianapolis, Indiana, USA

In the Mexican context, the main challenges at the national level for the development and consolidation of open Access digital repositories, we can mention the following:

- Promote the correct implementation of metadata providers
- Standardize types, dates, identifiers, publisher, and author names according to La-Referencia new guidelines (OpenAIRE)
- Promote the creation of new Institutional Repositories
- Promote the creation of Open Access Mandates
- Integrate REMERI with the National Repository of Mexican Science (in development)
- Collaborate with CONACYT in training for technical and management skills.

References

Abadal, E. (2012) . Open access to science . Editorial UOC. Retrieved February 18, 2014 , from : <http://diposit.ub.edu/dspace/bitstream/2445/24542/1/262142.pdf>

Alonso, J., I. Subirats and Martinez , ML (2008) . AOSIS Report on Open Access . Gijón: Professional Association of Information Specialists . Retrieved on December 9, 2013 , from : <http://eprints.rclis.org/12507/>

Bustos G. A., and P. A. Fernández (2008) . Guidelines for creating institutional repositories at universities and higher education organizations . Retrieved on December 13, 2013 , from : <http://repository.urosario.edu.co/handle/10336/223>

Budapest Open Access Initiative (2002). Retrieved January 28, 2014 , from : <http://biblioteca.upc.es/rebiun/BOAI.pdf>

Crow, R. (2002) . The Case for Institutional Repositories : A SPARC Position Paper. Retrieved December 3, 2013 , from : http://works.bepress.com/ir_research/7

Lynch, C. (2003). Institutional Repositories : Essential Infrastructure for Scholarship in the Digital Age ARL : A Bimonthly Report, no. 226 (February 2003). Retrieved November 30, 2013 , from : <http://www.arl.org/storage/documents/publications/arl-br-226.pdf>

Suber (2012) . Open Access. Cambridge , Massachusetts : The MIT Press Essential Knowledge Series. London, England.