

REPOX – A Framework for Metadata Interchange

Diogo Reis^{1,2}, Nuno Freire^{1,2}, Hugo Manguinhas^{1,2}, and Gilberto Pedrosa^{1,2}

¹ INESC-ID, Rua Alves Redol 9, Apartado 13069,
1000-029 Lisboa, Portugal

² IST – Department of Information Science and Engineering, Instituto Superior Técnico,
Lisbon Technical University, Portugal
{diogo.menareis,nuno.freire,hugo.manguinhas,
gilberto.pedrosa}@ist.utl.pt

Abstract. This demonstration presents an XML framework for metadata interchange. REPOX has two goals: to be a means for libraries and other cultural institutions to provide OAI-PMH access to their metadata records, independently of their original format, with a tool that is easy to install, use and deploy; and to be used as an aggregator of OAI-PMH Data Sources. The records are stored internally in XML and there is a metadata transformation service that allows for translation to desired formats. This demonstration will show the usage scenarios, technologies and current results.

1 Introduction

Across libraries, we find many heterogeneous Library Management Systems (LMS) that use diverse metadata schemas to represent bibliographic data. Libraries face the need to make their bibliographic databases available by OAI-PMH to some European initiatives associated with digital libraries, like Europeana¹ and The European Library². Because many vendors of LMS don't support OAI-PMH, libraries need to implement custom made solutions, often using open source software which requires some technical expertise, often not found in the libraries staff.

REPOX [1] is a tool that can be deployed to have OAI-PMH access to bibliographic databases, requiring little technical knowledge, having a fast start process (installation and configuration). REPOX is focused on having support for the specific schemas used in the libraries. It also contains a metadata transformation facility so they can provide their records through OAI-PMH in the desired schemas.

2 REPOX

The supported record ingesting processes are: a file system folder containing the records; and an OAI-PMH source. REPOX supports any XML schema, but it has built in support for schemas frequently used in libraries, such as MarcXchange, Dublin Core

¹ <http://dev.europeana.eu/>

² <http://www.theeuropeanlibrary.org>

and the metadata profiles of Europeana and The European Library. ISO2709, a non XML encoding, is also supported since it is the main schema used by libraries to export/import data from the LMS. REPOX supports many product specific variations of this standard and several character encodings, including those that are library specific.

The record identifiers used in REPOX can be associated in two ways: generated by REPOX or extracted from each record using an XPath expression. The advantage of using extracted identifiers is that it is possible to update just the changes because the records can be recognized by the identifier.

The implementation is completely in JAVA. There is an installer for Windows and for Unix/Linux. The Web server used is Jetty, because it does not require a separate installation. For the same reason, the database is Derby, embedded in REPOX.

To expose the records by OAI-PMH, it is required at least the metadata schema `oai_dc`. REPOX has two approaches to solve this: sending an XSLT transformation from the local schema to `oai_dc`, or creating a transformation to `oai_dc` with a visual tool in a web page (using a JavaScript library).

4 Results and Future Work

REPOX is being deployed or tested in the national libraries of Portugal, Hungary, Slovenia, Poland, Russia and Spain in the context of TELplus³. In project FUMAGABA⁴, it has been tested by the national libraries of Albania, Bulgaria and successfully deployed at Ukraine. As an aggregator of metadata collections, REPOX was used in project DIGMAP to fulfill the requirements for submission of metadata records and their retrieval by the other DIGMAP services [2].

Current work focus on the use of REPOX as an aggregator, scalable to hundreds of data sources with several millions of metadata records. To accomplish that, the database API will be adapted for other available databases that can handle large scale reads and writes.

REPOX will be used in the infrastructure of Europeana and also in the project EuropeanaLocal⁵ starting in 2009.

References

1. Freire, N., Manguinhas, H., Borbinha, J.L.: Metadata spaces: The concept and a case with REPOX. In: Sugimoto, S., Hunter, J., Rauber, A., Morishima, A. (eds.) ICADL 2006. LNCS, vol. 4312, pp. 293–302. Springer, Heidelberg (2006)
2. Martins, B., Manguinhas, H., Borbinha, J.L.: Extracting and Exploring Semantic Geographical Information from Textual Resources. In: Proceedings of the Second IEEE International Conference on Semantic Computing (ICSC) (August 2008)

³ <http://www.theeuropeanlibrary.org/telplus/>

⁴ <http://www.theeuropeanlibrary.org/portal/organisation/cooperation/fumagaba/>

⁵ <http://www.europeanlocal.eu/>