**Integrate external bibliographic services in DSpace submission process to make self-deposit easy and improve metadata quality and presence of full-text.**

*Andrea Bollini, Susanna Mornati*

CINECA

One of the most difficult challenges for Institutional Repositories is to find ways to encourage deposit from the researchers. When an IR is part of the Current Research Information System (CRIS) of a University there may be mandates from the University top management to assure that researchers are required, to deposit their publications in the Repository, primarily for evaluation purposes. However also in this ideal scenario there is an issue: Researchers see the Repository as an administrative fulfillment. To change this perception and make the IR be seen as a useful tool for research, it is very important to facilitate the deposit procedure and improve data quality as far as possible.

As many publications databases already exists around the world and most of them offer APIs to access and reuse data, we have decided to revise the DSpace submission process to benefit of existent information as much as possible. One of the best known commercial bibliographic databases is Scopus that offers API access as part of the standard contract grant permission to reuse these data in IRs [1].

CINECA has therefore developed a generic infrastructure that allows easy integration of such services in the DSpace submission process. At the moment Scopus, PubMed [2], arXiv [3] and CrossRef [4] APIs are supported.

The first step of the submission has been replaced by a form where the user can enter one or more unique identifiers (DOI, PubMed ID, arXiv ID, etc.) and/or a combination of title, authors and year of her publication. The system will query all configured sources and present all results to the user. At this point she can confirm and proceed with the deposit using one of the system proposals or forcing the creation of a new publication from scratch. The system will use information related to the user (authorization) and metadata available in the external systems to suggest the most appropriate collection where to deposit the item, but the user is free to choose a different collection.

When available, the system will use the DOI to merge data from different external data sources in a single richer record, this allows for example to get the abstract of a publication in PubMed and combine other metadata from the standard Scopus API that does not provide this information out-of-box.

All metadata retrieved from external sources are mapped to the item metadata using rules defined in mapping configuration files. Different rules, normalization and data manipulation can be defined on a target collection basis. In the following steps of item description, metadata retrieved from external data sources are marked with corresponding icons and possible inconsistencies between data sources are highlighted to allow fast comparison and replace/integration. The same facilities are provided during the publication approval/validation workflow, so that librarians can focus on manually entered data to assure better metadata quality to IR records.

Over the year these integrations will be extended, adding automatic scanning capability to DSpace over different data sources. In this way the IR will have an active role alerting users, via email or notification windows on login, for new publications. Users will just have to claim/unclaim their publications.

Last but not least the existing integration made by CINECA between DSpace and Sherpa ROMEO[5] will be extended. Now it provides contextual publisher’s copyright information about self archiving in the upload step. We intend to make it provide also post-deposit analysis and followup. The system will be able to produce reports about missing full-text, grouped by ROMEO colors for publishers policies. Such reports, together with IR metrics that show usage of items and fulltext, can be shared with researchers to encourage post-deposit upload of full-text. Reports can be sent automatically via e-mail. Another useful service will be offered to retrieve and upload full-text automatically from external sources when possible (arXiv, etc.)

CINECA’s enhancements are periodically released to the DSpace community in form of patches or new features for the DSpace-CRIS [6] module.

**References**

All URLs have been accessed on 1st March 2013

[1] Scopus Content Policies: <http://www.developers.elsevier.com/devcms/content-policies>

[2] Sayers E. E-utilities Quick Start. 2008 Dec 12 [Updated 2011 Dec 14]. In: Entrez Programming Utilities Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2010-. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK25500/>

[3] arXiv API documentation: <http://arxiv.org/help/api/index>

[4] CrossRef OpenURL Query Interface Documentation: <http://help.crossref.org/#using_the_open_url_resolver>

[5] SHERPA/RoMEO Application Programmers' Interface: <http://www.sherpa.ac.uk/romeo/api.html>

[6] DSpace-CRIS website: <http://cilea.github.com/dspace-cris>