Implementing Citation Management and Report Generation Value-added Services over OAI-PMH Compliant Repositories

ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ
NATIONAL DOCUMENTATION CENTRE
www.ekt.gr

| Nikos Houssos (nhoussos@ekt.gr) | Christina-Eleni Paschou | Ioanna-Ourania Stathopoulou | Konstantinos Stamatis | Despina Hardouveli |

Abstract:

In this poster we present an application which harvests digital repository data and presents them outside the repository framework. Its main advantage is the ability of citation management by supporting a variety of citation formats (e.g. IEEE, ACS, Springer, RIS, BibTex etc.). In addition, it allows configurable custom reporting, facilitating the management and the promotion of the researchers' work. Currently, the application is in operation on top of a DSpace-based repository; however it has been designed and implemented to depend only on information retrieved via OAI-PMH, so that it can work with any OAI-PMH compliant repository platform (DSpace, EPrints, Fedora, etc.).

Introduction:

The National Documentation Center (EKT) has developed HELIOS (**http://helios-eie.ekt.gr**) - the institutional repository of the National Hellenic Research Foundation (NHRF) aiming at collecting the scientific work of its associate researchers. DSpace has been used as the repository platform in the implementation of HELIOS.

According to the repository literature, offering value-added services to researchers can be an important factor for repository take-up, able to significantly increase deposits through self-archiving. Therefore, in order to encourage the usage of HELIOS among the NHRF researchers, an application providing value-added services over the repository has been developed (**http://references.ekt.gr**). The architecture and the main functions and an example service of the application are presented in this poster.

Main Functions:

- Retrieve bibliographic citations from an OAI-PMH compliant repository according to a combination of basic filtering criteria like:
  - **author:** Since the HELIOS repository provides authority control for authors, a search can be done for a particular author retrieving all publications of the author in the repository independent of different writings of his/her name. This has been implemented using AJAX techniques for author name auto-complete in combination with getting dynamically author name information from an authorities web services that we have developed outside the repository, utilizing the MADS standard for authority file representation
  - **title**
  - **publication type** (e.g., journal or conference article)
  - **publication year**
- Transform of bibliographic citations among reference styles (e.g., Harvard, Chicago, ACS, APS, IEEE, AIP, Springer, Elsevier etc.)
- Export of the bibliographic citations in formats like BibTex format and RIS format that enable interoperability with online citation management services such as citeulike, connotea, mendeley and publication-slist
- Export of the resulting formatted entries in pdf, word and html files
- Generation of reports customized to the requirements of particular researchers and/or institutes. This requires cleaning and clustering of publication venue names (e.g., journals, conferences).

Application Layout:

Work in progress:

- Automatic parsing of citations from research articles and transformation into structured formats (e.g., BibTex, RIS)
- Formal specification of reference styles
- Automatic retrieval of citation information from sources like Web of Science and Scopus using the corresponding APIs.

Summary - Conclusions:

This poster provides an overview of the implementation of platform-independent specific value-added services over repositories, based only on the assumption of support of OAI-PMH. Besides the application development aspects, an important requirement for providing value-added services of this type is the high quality of the repository metadata and in particular the application of authority control to names of authors, organizations and publication venues. The availability of such mechanisms for authorities in the repository used in our case study was a key factor for the successful implementation of the services.

The Architecture:

Goal: The main goal of the application was

- the development of a system that would lie outside the repository infrastructure
- the interoperability with any repository platform.

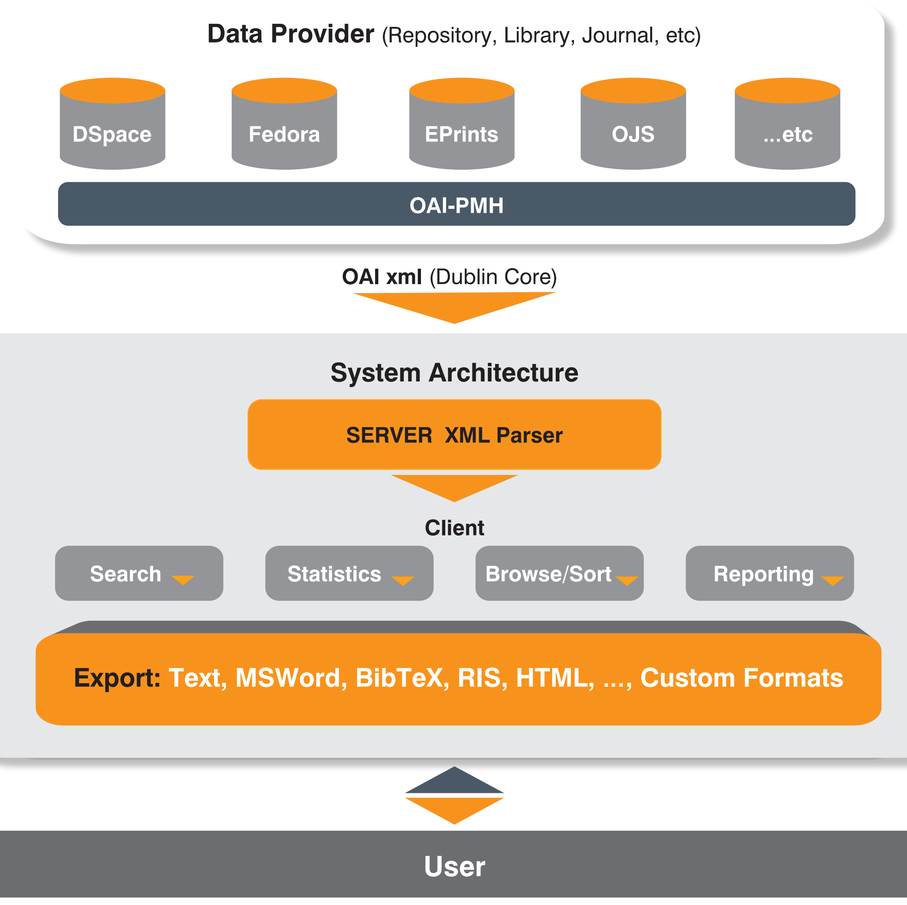
Solution: The application was built to depend only on data harvested through OAI-PMH

a Service Provider has been developed using the standard OAI-PMH interface exposed by Data Providers (repositories)

This Service Provider consists of two modules (Figure 1):

The server side module that communicates with the Data Providers (repositories) by sending them OAI requests (specific HTTP requests) and parses the incoming data (HTTP responses). This module handles also the implementation of all the required bibliographic references transformations and provides the relevant functionality through an appropriate API. In order to accomplish this task all the desired bibliographic reference styles have been modeled and appropriate structures have been created in the underlying application data model.

The client side module which is responsible for serving the resulting data to the users according to his/her preferences and generating exports and reports in various formats.



An Example Service

Target Group: A research institute within NHRF, namely the Theoretical and Physical Chemistry Institute (TPCI).

Problem: TPCI has custom reporting requirements for the production of the institute's annual report. The report includes categorization of publications into types which do not entirely correspond to values dc: type field in the repository (e.g., journal article, book, article in conference proceedings, patent). Categorization is based not only on publication type but on the quality of the publication venues. For example, articles in high-profile journals are classified in category 1 (termed "Papers in Referred Journals"), while papers in magazines of non-archival publications or in less competitive journals are classified in categories "Articles in Magazines" and "Other", respectively. Certain publications in very competitive conferences can also make it into category 1, instead of the "Articles in Referred conferences" class. The aim is to allow the researchers to automatically extract their publications categorized based on the aforementioned requirements.

Solution:

1. An authority file has been created which including not only alternative writings of their names but also the category in which they correspond in the TPCI classification. This way the publications are categorized automatically based on publication venue (e.g, journal of conference name) – not their type. This authority file has been produced semi-automatically through clustering of publication venue names and subsequent correction of errors by EKT personnel.
2. An auto-complete mechanism has been developed to help the researcher during search. This has been achieved by creating researcher authority files, including only NHRF authors (not co-authors from other organizations), which again has been produced through a semi-automatic manner, involving a first automated step performing clustering on author names and a round of human personnel correcting potential errors. A specific desktop software application has been developed to enable this manual editing of authority files, which are represented and stored in compliance to the MADS standard.
3. The search results can be exported to pdf or word format, categorized based on the researchers' requirements

Design by Dimitra Pelekanou | EKT 2010