

Implementing citation management and report generation value-added services over OAI-PMH compliant repositories

Nikos Houssos, Christina Paschou, Ioanna – Ourania Stathopoulou, Konstantinos Stamatis,

Despina Hardouveli

National Documentation Centre / National Hellenic Research Foundation

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 literature¹, 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. In brief, this application harvests the digital repository's data and presents them outside the repository's framework, enabling citation management and configurable custom reporting, for example producing publication lists per researcher and institute, exactly in the format applied in the institute annual report. The application is in operation on top of the HELIOS 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.).

Main Functions

The main functions of the application are listed below.

- Retrieve bibliographic citations from an OAI-PMH compliant repository according to a combination of basic filtering criteria like author, title, publication type (e.g., journal or conference article) and publication year. 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.

- Transform of bibliographic citations among reference styles (e.g., from Harvard to Chicago). A variety of is supported, among others ACS - American Chemical Society, APS - American Physical Society, Chicago, Harvard, IEEE, AIP - American Institute of Physics, Springer for Medicine / Life Sciences, Springer for Physics, Elsevier for Chemical Physics Letters, Vancouver.
- 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 publicationslist.
- 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.

Architecture

The main goal of the application was to develop a system that would lie outside the repository infrastructure and be interoperable with any repository platform. In order to satisfy this demand, the application was built to depend only on data harvested through OAI-PMH. In fact in OAI-PMH terminology, it realizes a Service Provider that uses the interface that the Data Providers (repositories) provide in order to harvest their metadata. This Service Provider, that has been developed, consists of two modules, as depicted in Figure 1:

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.
2. 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.

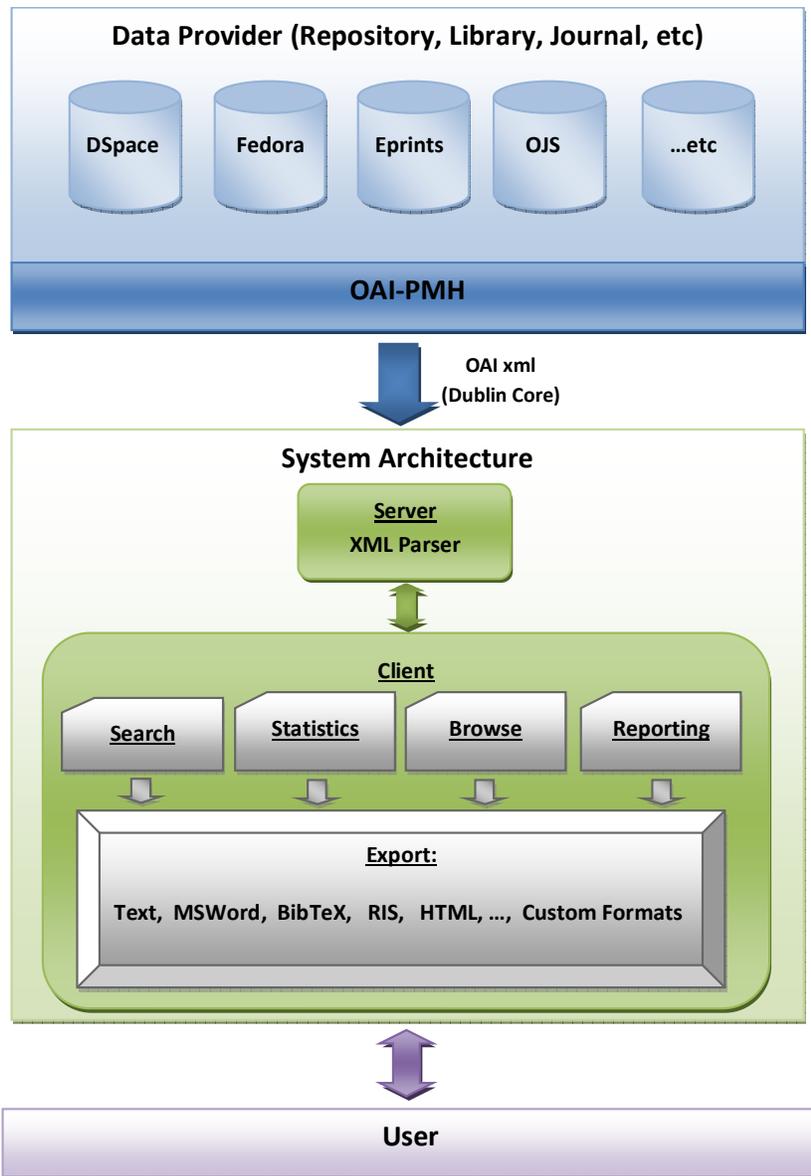


Figure 1. Architecture of value-added service implementation.

An example service – customised reporting for a research institute

In this example, we investigate the case of a research institute within NHRF, namely the Theoretical and Physical Chemistry Institute (TPCI), which had custom reporting requirements for the production of the institute’s annual report. In particular, the publication lists included in the institute’s annual report at any level of granularity (e.g., entire institute, research group within institute or individual researcher) should include a categorisation of publications into types. However, these types do not entirely correspond to values dc:type field in the repository (e.g., journal article, book, article in conference

proceedings, patent). Categorisation 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.

Therefore, in order to automatically produce reports from the application according to the institute’s needs publication are categorised automatically based on publication venue (e.g, journal or conference name) – not their type. In essence, we have created an authority file for publication venues, including not only alternative writings of their names but also the category in which they correspond in the TPCI classification. This authority file has been produced semi-automatically through clustering of publication venue names and subsequent correction of errors by EKT personnel. Based on this, each publication is placed in the appropriate category on the basis of the journal / conference it has been published.

To produce reports at the level of researcher we also have utilised the respective authority control for author names, which can be exploited by the user in the search function through an auto-complete mechanism. The corresponding researcher authority files, including only NHRF authors (not co-authors from other organisations) have been produced again 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 according to the MADS standard.

It is obvious from the above that practically it is impossible to provide successful services without good data quality and authority control within the repository; in fact, applying authority control and cleaning data is typically much more costly than developing the service itself.

Summary – conclusions

This document 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. The services are being used by researchers and their feedback is very positive; actually the existence of the services has spurred an increasing interest in the repository. 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, organisations 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.

References

1. (2008). A DRIVER’s Guide to European Repositories, Amsterdam University Press.