

Promoting and Supporting Open Archives in Europe: The Open Archives Forum Project

*Donatella Castelli
IEI-CNR
Pisa (Italy)
castelli@iei.pi.cnr.it*

Abstract

This paper presents the Open Archives Forum, an EC 5th Framework Accompanying Measure for supporting the dissemination of information about European activities that adopt an open-archives based approach. OA-Forum aims at building a framework where European and national initiatives that use this approach can share their experiences and co-ordinate the development of software tools and infrastructures. Special attention is dedicated to those initiatives, which are implementing or using the Open Archives Initiative Metadata Harvesting Protocol, a technical solution proposed within the context of the scientific pre-print archives, but the forum is also paying attention to other solutions that are emerging in different areas of interest.

1. Introduction

The “open archives” approach is gaining popularity in several communities such as the cultural heritage, scientific and learning ones. The technology adopted makes it possible to ensure cross-archive access, and low cost dissemination and access to content that is currently “hidden” to the wider public at a very low cost. If the full potential of this approach is fulfilled, it will have a significant impact on the building of new service provision. It is essential that European data and service providers be prepared to participate in this innovation.

The building of a cross-archive access service from distributed archives has been firstly made feasible by the Z39.50 community [1]. The Z39.50 protocol allows complying systems to search on multiple distributed databases in a homogeneous manner, thus permitting to view the selected information sources as a unique library. Although a Z39.50 Profile for Access to Digital Collections has been developed, the Z39.50 Protocol has been mainly used for bibliographic databases.

Recently, the scientific communities that publish their pre-prints on electronic archives have raised again the problem of building an infrastructure for supporting cross-archive services. In a meeting held in Santa Fe (New Mexico) in July 1999, they set up the Open Archives Initiative (OAI) [2] whose objective is to develop and promote interoperability solutions that aim to facilitate the efficient dissemination of content. OAI proposed a simple protocol, the Open Archives Initiative Metadata Harvesting Protocol (OAI-PMH) that, once implemented by an archive, allows third parties services to harvest its records. At the present, more than sixty large archives have adhered to this initiative implementing the suggested protocol and more than ten organizations have implemented cross-archive services relying on this protocol.

Other general solutions are currently being experimented by different communities, mainly scientific communities [3]. These solutions often reflect the different requirements exposed by these

communities, in particular the services they need and the kind of information stored in their archives e.g. large numeric/scientific datasets, medical and satellite images, moving images and sound, molecular structures and engineering records.

A lot of work is also underway in the educational environment [4,5]. The opening of the archives offers to the teachers the possibility of training their students by giving them access to a great amount of different information. This offers unexpected possibilities and it is going to strongly revolutionize education at all levels.

The Open Archives Forum (OA-Forum) [6], is a recently EU 5th Framework Accompanying Measure (IST-2001-32015)¹ which has been set up for disseminating the open archives-based approach and for supporting the co-ordination and co-operation among those European and national initiatives that adopt, or are willing to adopt, this approach. OA-Forum aims to build a framework where both data providers (archives) and service providers can share their experiences and co-ordinate the development of software tools and infrastructures. A special attention will be initially dedicated to those initiatives, which are implementing or using the OAI-PMH protocol but the plan is to be able to extend the scope also to the other application areas where the need of cross-archives services is emerging.

This paper presents the OA-Forum project by describing its objectives, its context and the activity planned to create a framework for supporting a wider co-operation among the actors that are involved, or are willing to be involved, in the open archive approach.

The rest of the paper is organized as follow. Section 2 briefly introduces the OAI, which is the initiative that inspired the OA-Forum project proposal; Section 3 presents the project activities; and Section 4 describes the next steps of the project.

2. The Open Archives Initiative

There are currently several e-print archives that serve large communities of scientists. The most widely know is the Los Alamos E-Print Archives, called ArXiv.org [7], which at the present maintains non peer-reviewed research papers, not only in the physic research areas but also in the mathematics, non-linear systems and computer science. This archive stores more than one hundred and fifty thousands documents, it has a monthly submission rate of two thousand documents and typically process more than one hundred thousand connections per-day. Other significant archives are NDLTD [8], a digital library that maintains electronic theses and dissertations published in more than one hundred different countries and ETRDL [9], a European collection of computer science research reports.

Each of these archives have been set up by a specific community which use them as an instrument for supporting collaboration among its members. Their services address the peculiar needs of their own specific community and are accessible through ad-hoc interfaces. This specificity prevent the exploitation of the full power of these archives since it precludes the possibility of serving cross-domain communities and, when there are more archives, also limit their full utilization even within the same community. The Open Archives initiative (OAI) was set up to discuss and solve this problem of interoperability between author self-archiving solutions.

OAI originated by a meeting held on July 1999. The Council of Library and Information Resources (CLIR), the Digital Library Federation (DLF), the Scholarly Publishing & Academic Resources Coalition (SPARC), the Association of Research Libraries (ARL) and the Los Alamos National

¹ The participants to the Open Archives Forum project are: UKOLN-University of Bath, Bath (United Kingdom); Humboldt University, Berlin (Germany); Istituto della Elaborazione della Informazione-CNR, Pisa (Italy).

Laboratory (LANL) sponsored the meeting. The participants were representatives of different scholarly e-print archives. The aim of the meeting was to explore the co-operation among these archives as a way to contribute in a concrete manner to the transformation of the scholarly communication. The participants to the meeting wanted a very simple, low-barrier to entry interface that could shift implementation complexity and operational processing load away from the repositories. They also wanted to minimize the interdependency between the quality of application service as viewed by the user and the behavior of repositories that supplied data to the application services [10]. Thus instead of a distributed search model, as that adopted by Z39.50, they chose a “harvesting -based” approach. According to this approach service providers take the responsibility for collecting the metadata stored in the repositories and normalizing these metadata to support functions as federated searches. To support this approach the OAI designed a protocol, known as “Open Archives Initiative Metadata Harvesting Protocol”(OA-PMH) that defines the set of requests that each archive must respond to. In designing this protocol a lot of effort was put in defining requests that could be implemented by each repository with a few days person effort. This requirement was deemed essential for a wide adoption of the interoperability solution.

A lot of discussion about the OAI and its solution for interoperability had taken place during the past years. Many other groups that have similar problems to those faced by the e-print community, i.e. libraries, museum, commercial journal publisher, and communities of scholars who need to share distributed resources, had discussed the possibility of adopting the OAI approach and had raised their own specific requirements. Several archives had implemented the protocol and a number of services had been built on top of it. As a result of this experimentation a number of open questions has been identified. Some of them will get an answer in the revised version of the protocol that will be officially released in June 2002, others, which still need considerable research and experimentation have been left out at the moment. These include very important issues like property issues for metadata, registries, granularity, and metadata schemes.

3. The Open Archives Forum

The Open Archives Forum (OA-Forum) project proposal was motivated by the increasing discussion about the open archives approach raised by the OAI. The aim of the project is to set up a framework for discussing the issues related to open archive approach in general, i.e. not necessarily related to the OAI technical proposal, and for sharing experiences and solutions. The ultimate goal of the project is to promote this approach in all the areas where its application can provide a framework for the construction of more powerful user services.

The activity of the project is articulated in a number of tasks. Each of them focuses on a specific issue related to the open archives-based approach. The main tasks are the following:

- *Creation of a repository of information about activities related to the open archive approach*
The different European experiences in exposing the archive data to the outside services and in implementing services that use these data have been collected, described and registered on a Web-accessible database that supports public search. The database maintains descriptive information about each experiences, like the application or research project, the tools used, the people involved and, when possible, existing documentation about it. Particular attention is dedicated to the collection of developed services, like document delivery, searching and browsing harvesting, summarization, linking, etc. The purpose of this database is to provide a picture of the European open archive research and developments that can be exploited in setting up new projects or reusing software solutions. The OA-Forum consortium is currently exploring the possibility of sharing this database with other organizations, e.g. OAI, in order to create a more complete source of information.

- *Validation of the European experience in implementing and using the OAI-PMH protocol and other similar approaches to interoperability*

The OA-Forum project monitors the on-going activities that implement mechanisms for supporting the interoperability among open archives in order to highlight limitations, difficulties and emerging software needs. In particular, it records evaluations of these mechanisms in terms of the potential they offer for building new services and the resources they need. It also collects requirements from implementers and users both in terms of functionality and supporting tools. This information is made public available on the Web site of the project and can be used by anyone who is approaching the development of new open archives and services.

- *Review and analysis of a variety of possible business models for the implementation of open archives*

The added value that the open archives model might offer in different application communities is being explored. This information will lead to an understanding of the issues necessary to ensure that European organizations benefit from the potential of the open archive technology. Particular attention is dedicated to the issues relating to Intellectual Property Rights, terms and conditions of use, and quality assurance.

- *Dissemination of the open archives approach*

Four European workshops will be organized over the two years life of the project to increase awareness of the value to the user of open archives, encourage the sharing of experiences and propagation of good practice in opening-up archives. They will be targeted towards different domains, e.g. cultural institutions, research organizations, public sectors, community services and commercial organizations, and towards different classes of users e.g. potential implementers, service providers, archivists. Particular attention will be dedicated to involve communities that may be potentially interested in open archives but which do not have yet developed the technology necessary to do it.

The OA-Forum project started its activity on October 2001. The first six month of the project has been mainly dedicated to the set up to the instruments required to support its activities (Web site, database) and to acquire an initial panoramic of the European activity around open archives. In particular, this first part of the project has been focussed mainly on those initiatives that are related to the OAI approach. The project initial investigation outlined an emerging need of across-archives services in several sectors all around Europe. As a result of this preliminary investigation we found that the situation is very different from North to South Europe. The northern countries, e.g. Germany, United Kingdom, Denmark, are deeply involved in the discussion about the open archives approach. They have implemented the OAI-PMH and participate actively in the experimentation set up by the OAI. On the contrary, the Mediterranean and the Eastern Europe, i.e. Italy, France, Greece, Portugal, Spain, Czech Republic, Hungary, etc. with very few exception [11, 12, 13] are just now starting to discuss about the open archives approach and the OAI. One of the reasons for such mismatch is certainly related to the presence, in northern countries, (absence in all the others,) of a National programme for supporting the digitalisation and the creation of digital libraries. The National programmes have in many cases favoured the construction of large digitalised archives and have feed the research and experimentation of new solutions to their problems. As a result, in these countries there are many large digitalised archives and the problem of making them cross accessible has been widely discussed. More details about the European map on the open archives activity in Europe can be found visiting the OA-Forum Web site [6].

3. OA-Forum future activity

The collection of information about initiatives and projects related to the open archive approach will continue. The scope of the monitoring activity, which started by focusing on the OAI, will be

enlarged to consider other communities and other technical approaches. This will imply to face with different type of archives and different services. The OA-Forum database will be updated and, according to the requests received, its design may be re-thought to add further information. The organizational and validation activities will widen their scope by evaluating and comparing the different requirements and solutions.

A first step towards the dissemination of the open archives approach to other communities will be done at the First Open Archives Workshop, titled “Creating a European Forum for Open Archives Activities” which will be held on 13th-14th of May 2002 at IEI-CNR in Pisa (Italy). This two-day workshop seeks to bring together researchers, technical implementers and project managers who are experimenting, or are willing to experiment, the open archives approach. The goal of the Workshop is to set up the basis of a European forum for sharing experiences and solutions and encouraging networking among projects. The first day of the Workshop will be dedicated to an overview of the OAI activity in Europe. Invited speakers from different European countries will illustrate their experiences in opening their e-print archives and libraries according the OAI-PMH protocol and in implementing cross-archive services based on it. The second day will see the involvement of other communities, in particular those that use scientific data archives. The co-existence of these different communities will permit to discuss and compare the incentives, technical and organisational challenges faces by each community. It will also offer the possibility of understanding which aspects might be generalised to serve other realities.

The project plans to continue this process of enlarging the set of communities in the future, moving from those that are strictly related to the scientific world to those more closed to the real life, such as cultural, educational, public administration, etc.

The project will also provide reviews prepared by experts on main topics related to the open archives approach. An expert review report on Property Rights has already been planned. In the world proposed by the OAI-PMH metadata resources becomes migratory and recombinant. This creates the problem of expressing and guaranteeing metadata ownership and permissible uses and reuses. The representation and handling of property rights is one of the open questions for most of the archives that are planning to open themselves to cross-archive services.

Both the results of the Projects Workshops and the Experts Reviews reports will be make public available on the project Website (<http://www.oaforum.org>).

References

- [1] Z39.50 Maintenance Agency, <http://lcweb.loc.gov/z3950/agency>
- [2] Open Archives Initiative, <http://www.openarchives.org>
- [3] Astrophysical Virtual Observatory, <http://www.eso.org/projects/avo/intro.html>
- [4] Re-inventing Scholarly Information Dissemination and Use, <http://elib.cs.berkeley.edu/>
- [5] National Science, Mathematics, Engineering, and technologies Education Digital Library, <http://www.nsd1.nsf.gov/tour1.html>
- [6] Open Archives Forum, <http://www.oaforum.org>
- [7] arXiv.org e-Print archive, <http://arXiv.org>
- [8] Networked Digital Library of Theses and Dissertations, <http://www.ndltd.org>
- [9] Ercim Technical Reference Digital Library, <http://dienst.iei.pi.cnr.it/>
- [10] Clifford A. Lynch, “Metadata Harvesting and the Open Archives Initiative”, ARL Bimonthly Report 217, <http://www.arl.org/newsltr/217/mph.html>
- [11] Cyclades: An Open Collaborative Virtual Archive Environment, <http://www.ercim.org/cyclades/>
- [12] Torii, an Open Portal over Open archives, High Energy Physics Libraries Webzine, <http://library.cern.ch/HEPLW/4/papers/4/>
- [13] Budapest Open Access Initiative, <http://www.spros.org/openaccess/read.shtml>