OPEN ARCHIVES FORUM: PROJECT DELIVERABLE

Project Number:	IST-2001-32015
Project Title:	Open Archives Forum
Deliverable Type:	Public

Deliverable Number:	D4.3
Contractual Date of Delivery:	September 2002
Actual Date of Delivery:	13 March 2003
Title of Deliverable:	Workshop Report 2
Workpackage contributing to the Deliverable:	WP4
Nature of the Deliverable:	Report
URL:	http://www.oaforum.org/otherfiles/oaf_d42_workshop2.pdf
Author:	Donatella Castelli, Leona Carpenter, Susanne Dobratz, Birgit Mattaei
Contact Details:	Donatella Castelli, +39 050 3152902

Abstract	This report documents the Second Open Archives Workshop held in Lisbon (Portugal) on 5th -7th of December 2002.
Keywords	Open archives, Open Archives Forum, Open Archives Initiative, e-print archives, conventional archives, libraries, intellectual property rights.

Distribution List:	Public
Issue:	Version 1.0
Reference:	32015-DEL4.3-20030313
Total Number of Pages:	22

This report documents the Second Open Archives Workshop held in Lisbon (Spain) on 5th - 7th of December 2002. It contains a detailed description of the issues discussed and the results achieved. In particular: Section 2 contains an overview of the workshop; Section 3 reports the presentations given by the invited speaker and the discussions held within the break-out sessions; Section 4 summarises the workshop outcomes and actions; and finally Section 5, concludes by reporting the lessons learned.

1 GENERAL OVERVIEW

One of the aims of the Open Archive Forum project is to promote the open archive approach and to evaluate its impact in new application areas. The objective of this workshop was to make a contribution towards the achievement of this aim by exploring whether, and under what conditions, the open archive approach is viable for archival and library organisations. The workshop also intended to promote the establishment of new collaborative links aimed at building interoperable infrastructures or supporting the dissemination of both archival and library resources.

Key players in the two kinds of organisations were invited to present their view about the new possibilities that the open archives approach can provide and to report the current state of its adoption. The profile of the invited speakers was quite different. This greatly contributed to the provision of a wide overview of what is happening both in the Nationally and in the European funded projects.

Following the suggestions gained during the First OA-Forum Workshop, a tutorial on the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) was held the afternoon before the workshop for those people who were not familiar with this protocol. This created better conditions for the discussions held in the following days.

The workshop was organised with both presentations and breakout sessions in order to better implement the project objectives, i.e. support a discussion forum. Five break-out sessions were set up to discuss relevant issues and the invited specialists were asked to take into account these issues in their presentations. In addition, a meeting of the Organisational Issues Working group, that was spontaneously created at the previous OA-Forum workshop, took place where specific organisational topics were discussed.

The workshop had a very good attendance, with more than fifty registered participants attending, along with seven invited speakers, two tutorial lecturers and seven OA-Forum project workers. There were representatives from four EU funded projects. Fourteen European countries were represented: Austria, Belgium, Denmark, Finland, France, Germany, Italy, The Netherlands, Norway, Portugal, Spain, Sweden, Turkey, and the United Kingdom. There were also two people from Taiwan. The profile of the attendants was quite heterogeneous, there were librarians, archivists, researchers, technicians, project leaders, etc.

Our general impression is that, so far, we have achieved our goals. A great number of contacts for future collaborations have been established. We have been happy to have had the opportunity to set up collaborations between OAI experimenters and archivists. We have also appreciated that many people who attended the first Workshop also registered for the second one. This has certainly been a demonstration of an interest that they found in this initiative.

2 THE INVITED PRESENTATIONS

This section contains the abstracts of the presentations provided us by the invited speakers and brief comments written down by the project partners. A session is dedicated to each presentation. The slides of these presentations can be found on the OA-Forum project Website **(**http://www.oaforum.org/workshops/lisb_programme.php**).**

2.1 Open Access to Libraries

by Josè Borbinha (National Library of Portugal),
Hans-Jörg Lieder (Berlin State Library)

Theo van Veen (Koninklijke Bibliotheek)

This was a collaborative presentation where the three speakers discussed how the OAI-PMH approach has been used within three European projects: MALVINE, LEAF and TEL.

MALVINE and LEAF: Perspectives of the Open Archives Initiative Protocol for Metadata Harvesting in European Projects and Beyond
by *Hans-Jörg Lieder (Berlin State Library)*

MALVINE will provide access to distributed holdings of modern manuscripts kept in European libraries and archives. The service will be available as of 1.1.2003 at: http://www.malvine.org. The launch of MALVINE will not mark the end of a development but - hopefully - rather the beginning of a new phase of activities in the sector of modern manuscripts. The future emphasis of MALVINE will clearly be the maintenance and consolidation of the status quo and the integration of further data providers with a view on increasing the pan-European significance of search results.

At the time of launching MALVINE the OAI protocol for metadata harvesting will play no role in the network scenario. The presentation will describe how the MALVINE Consortium envisages the possibly future use of the protocol in a landscape of European institutions - be these great or small - providing data to a joint service.

The LEAF project (http://www.leaf-eu.org**)** develops a model architecture for establishing links between distributed authority records (personal names only) and providing access to them. The system will allow uploads of the distributed authorities to a central system and automatically link those authorities concerning the same entity. It will be shown how the OAI protocol plays a vital role in keeping the central repository up-to-date at any time.

The European Library: opportunities for new services
by *Theo van Veen (Koninklijke Bibliotheek)*

The European Library Project (TEL), sponsored by the European Commission, brings together 10 major European national libraries and library organisations to investigate the technical and policy issues involved in sharing digital resources. The objective of TEL is to set up a co-operative framework, which will lead to a system for access to the major national and deposit collections in European national libraries.

In this presentation I will discuss the development of a metadata model and the development of an interoperability testbed. This testbed will offer distributed searching in the national collections via Z39.50 alongside searching a central index of metadata harvested from other collections via the Open Archives Initiative protocol (OAI-PMH). Access to this central index is offered via SRU - a new protocol for search and retrieval based on http and XML initiated by the Z39.50 Implementers Group as a low barrier alternative to Z39.50.

The major challenges in the technical work of this project are related to the diversity of collections, languages and local services. From a user perspective TEL hopes to meet these challenges by lowering the barrier for users to access the different collections by offering metadata searches integrated in a central index, not just a menu of web sites and we hope to offer translations of search terms and facilitate searching on names by providing access to different name authority databases. From a provider perspective we hope to lower the barrier to participate in TEL by using simple protocols and by facilitating conversions. And from a library perspective we hope to share machine-readable metadata by developing a common metadata model and vocabularies with tools that allow for an ongoing evolution.

TEL will offer access to different services like multilingual services, name authority services and links to local services. One of the keys to meet the above challenges is integration: metadata generated by each service should be usable when accessing other services. This requires a common understanding of metadata, an easy way to carry metadata from one service to other services and an easy way to associate related metadata. It will be discussed how the TEL metadata development, resulting in a TEL Application Profile, dynamically generated links and integrated indexing of different types of metadata will contribute to fulfil those requirements. With name authorities as an example it will be demonstrated how this contributes to bringing new services within a "one mouse click distance". Integrated object metadata and name authority search will help the user in finding main entries rather than telling the user that his search resulted in no hits. A pan-European "Central Name Authority File", as being one of the results of the LEAF project can contribute in the realisation of these valuable services

Some remarks

These presentations demonstrate that the library community is willing to experiment with the potentiality of the OAI-PMH protocol since there is a real need of supporting interoperability. In particular, the presentations have outlined that libraries are increasingly seeing the possibilities for performance improvements in searching across distributed databases through metadata harvesting via OAI-PMH to build central indexes.

2.2 How Real Archivists can Learn to Love OAI

by George Mackenzie (The National Archives of Scotland)
Goran Kristiansson (The Regional Archives in Lund)

This talk looks at the potential for using the Open Archives Initiative Protocol for Metadata Harvesting as a simple means of disseminating and exchanging archive catalogues. The world of conventional archives is interested in exchanging metadata, and has widely adopted international data structure standards produced by the International Council on Archives. It has also shown interest in a system for encoding catalogues known as Encoded Archival Description, or EAD. Archive descriptions are complex and collection based, proceeding from the general (fonds or collection level) to the particular (item level). The talk briefly examines two implementations of OAI, the University of Illinois Project and the AIM25 project in the UK. It also considers a related hybrid implementation in Australia, and a planned use of the protocol in A2A, another UK project. It observes that OAI can be used for exchanging archive descriptions, but that there are problems. The first is difficulty in accurately reflecting linkages between levels of description. The second is the inconsistent application of EAD. The talk also looks briefly at alternative means archivists are using for exchanging metadata, particularly the Z39.50 protocol. The talk concludes that OAI will be used by conventional archives only if three conditions are fulfilled. First, archivists must be confident that compliant descriptions will respect archival principles, second, descriptions must be produced with little effort from existing systems, and third, archivists must believe that the wider OAI user base contains sufficient numbers of potential users. It suggests possible strategies in which archives would produce OAI compliant records for parts of their descriptions only.

Some remarks

The archival community has a long tradition; this is probably why innovation and agreements take a long time before being accepted. For example, the worldwide archival community has not yet reached an agreement on a standard resource description format. The OAI-PMH is not yet known within the conventional archival community, except for very few experimental projects. This presentation gave a very clear view of the state of these experimentations and outlined the motivations that might push archivists to move towards an open approach.

2.3 Open Archives and Intellectual Property- incompatible views? by Marc Bide (Rightscom)

This talk discusses the relationship between the Open Archives Initiative and Intellectual Property. There is considerable confusion over the nature of the Open Archives Initiative and the open access movement, which confuses much of the discussion surrounding OAI. So far as possible, we try to distinguish between these, although both are discussed.

Many of the issues that this raises have as much to do with commercial considerations as with legal ones, and it is inevitable that there should be some cross over between these different perspectives since "the content industry" is dependent on copyright for the security of its business model. Intellectual property is an essentially utilitarian concept, designed to maximise the value of creative effort for society as a whole as well as for individual creators.

Intellectual property is governed by national law, drawn up in accordance with international conventions and treaties.

National law has two distinctively different traditions: the continental European tradition is based on the *"droit d'auteur"*, the inalienable right of the creator over the creation; the Anglo-Saxon tradition is more explicitly commercial, seeing copyright as predominantly a property right, something that can be traded. These differences in outlook sometimes lead to substantially different attitudes to Intellectual Property issues, although the difference in their practical impact is relatively limited.

Copyright provides creators with an exclusive right to control the copying and publication of their work for a limited period of time. This right may be assigned or licensed to others.

Moral rights provide additional rights to creators, including the right to be identified as the creator of a work, and the right to object to derogatory treatment of the work; moral rights carry significantly different weight in different national legislative frameworks.

All copyright legislation includes certain limited exceptions. These must (under international treaty) pass a "three step test" which ensures that the exceptions do not unduly interfere with the normal exercise of the creator's rights. Exceptions are normally limited by a test of "substantiality" which cannot be objectively measured. Additional rights exist in many countries to protect databases that may not be protected by copyright law because they exhibit insufficient creativity.

The development of the global network has not altered the law of copyright - all existing legislation applies equally to content on the network as elsewhere. However, new legislation has been necessary to reflect changed circumstances, creating new exceptions to copyright and new protections for copyright owners.

Because of the ease with which intellectual property can be copied and distributed over the network, some owners of intellectual property rights believe that the law is not able to provide sufficient protection and are seeking to develop and implement technical measures to protect their content.

Some believe that there can be no effective technological measures for the protection of copyright, and that other ways will have to be found to compensate owners for casual copying. In some countries, this includes the introduction of levies on either copying equipment or media.

The existence of the network is also encouraging the development of new ways of licensing intellectual property, based on the "open software" model. These licences selectively assert creators' rights under copyright law, but permit users wide licence to copy and distribute without payment.

Individual items of metadata may not be protected by copyright, to the extent that they are simply facts intrinsic to the resources that they describe. However, metadata records which include elements of significant creativity - including abstracts - may be "works" in their own right and protected by copyright. Collections of metadata may be covered by database right, even if the metadata records themselves are not covered by copyright.

The resources described by the metadata are likely themselves also to be subject to copyright protection, unless they have passed into the public domain because of their age. Our focus in this report is on academic journal articles, since these are the main subject of current Open Archives activity.

Although it might be assumed that academic institutions would in general own copyright in journal articles (the normal rule for employers whose employees create intellectual property), it is custom and practice, and often explicit in employment contracts, that academics retain rights in journal articles. We believe this is highly unlikely to change to any great extent in the foreseeable future.

Journal publishers have traditionally insisted on a full assignment of rights in articles that they publish. However, many are now content to accept an exclusive licence to publish. However, an exclusive licence may be just as restrictive as an assignment.

Many journal publishers do not seek (at least at the present time) to restrict authors from posting copies of journal articles (either before or after formal publication) to the eprint archives. However, authors should ensure that they have an explicit understanding of the rights and the contractual situation, which may be complex.

Those who manage eprint servers are publishers, and will need to be aware of their responsibilities as such. This implies that they should ensure that they received proper warranties that an author has the right to post an eprint of a paper.

Non-textual resources are more complicated than text resources from the point of view of rights clearance and ownership; the owners of the rights in these resources are often much

more rigorous about their enforcement. Repositories that include non-textual materials will have to be very careful to ensure that they do not infringe any rights.

It is clear that authors' attitudes to questions of intellectual property and Open Archives are substantially coloured by the value that they seek from publication (which is not directly monetary). Their behaviour indicates that, even in those disciplines where Open Archives have been long established, formal publication in the peer-reviewed literature remains essential. This is always likely to mean giving up some rights over the content.

Publishers' current attitudes to the Open Archives Initiative have been much affected by the confusion between the Open Archives Initiative and the open access movement. It is hardly surprising that publishers show little enthusiasm for what is often openly portrayed to them as an attempt to undermine their business.

It would be equally unsurprising if academic institutions did not favour a mechanism which might make the acquisition of journals content less expensive (or indeed anything else). This is the other side of the coin. However, they will have to take on considerable responsibilities if they are themselves to become publishers on a large scale.

We recommend that those involved as data providers and service providers in the OAI model should develop mechanisms to make explicit their understanding of the use to which harvested metadata will be put. To this end, we recommend that metadata harvested under the OAI protocol should include information about the permitted uses of the metadata itself and the rights and permissions status of the resource which it describes. We believe that those operating eprint archives - or any other online resource repository - will need to take their responsibility as publishers seriously. This will include developing "notice and takedown" procedures for dealing with situations when notice is given of alleged infringements of copyright.

There is ultimately no conflict between Open Archives and Intellectual Property - but Open Archives exist within the framework of Intellectual Property law, and would be advised to recognise this in the way that they operate.

Some remarks

This presentation excited a great deal of interest from workshop participants, one of whom later commented that it was the first non-boring and really comprehensible presentation on the subject he had experienced. The issues raised made clear the importance of policies and processes for dealing with IPR issues within repositories, including the IPR in the harvestable metadata.

2.4 Various European OAI-PMH compliant services

by Donatella Castelli (ISTI-CNR)
Theo Van Veen (Koninklijke Bibliotheek)
Heinrich Stamerjohanns (University of Oldenburg)

This session, held by volunteers, was organized during the workshop to cover the absence of invited presentation by Simon Warner that was cancelled in cause of bad snow and ice storms in Eastern US.

"The Cyclades prototype"

by Donatella Castelli (ISTI-CNR)

Cyclades (http://www.ercim.org/cyclades) is a Vth Framework EU funded project that has developed an open collaborative virtual archive service environment supporting both single scholars as well as scholarly communities in carrying out their work. In particular, it provides functionality to access large, heterogeneous, multidisciplinary archives distributed over the Web that implement the OAI-PMH protocol and it supports remote collaboration among the members of communities of interest. The first prototype is now ready and it will be soon available for testing.

"TEL advanced services"

by Theo Van Veen (Koninklijke Bibliotheek)

The European Library (TEL) enables integrated searching in descriptions of collections and digitised and printed objects via a central portal. To define common terms for metadata the concept of application profiles as emerged within the Dublin Core Metadata Initiative has been adopted. The choices with respect to which terms are being used and how the Dublin Core elements are refined with qualifiers and encoding schemes is based on the required functionality and services that will be offered to the user.

The presentation shows how different services require the availability of specific metadata elements and on the other hand how the presence of certain metadata elements can trigger the TEL portal to offer specific services. Different examples of such combinations as used in a test-interface for The European Library are shown.

It is shown that the presence of a title or an abstract will trigger the portal to offer the user a choice to have the contents of that field translated from the current language to other languages. The presence of an identifier field with the URL as encoding scheme will trigger the portal to offer the user the possibility to link to that service. Other examples are the record-id to link to the original record, a base-URL to generate a new search target from the metadata record, link to OpenURL resolution in case of an identifier with OpenURL as encoding scheme and so on.

It is argued that the TEL application profile, which describes the characteristics of terms used by TEL, is based mainly on which functionality and services are required.

"Distributed Open Archives"
by Heinrich Stamerjohanns
University of Oldenburg
This presentation introduced the PhysDoc (Physics Documents Worldwide) and the "Open Archives: Distributed services for physicists and graduate students" (OAD) projects. PhysDoc aims at offering lists of links to document sources, such as preprints, research

reports, annual reports, and list of publications of worldwide distributed physics institutions and individual physicists, ordered by continent, country and town.
OAD is an international project sponsored by the National Science Foundation (NSF) and the Deutsche Forschungsgemeinschaft(DFG). This project improves distributed digital library services to support scholarly communication. It focuses on two types of users: graduate students and physicists.
In particular, the presentation focused on the discussion of the different approaches to the interoperability of structured data in heterogeneous environment and on the issue of low barrier framework.

Some remarks

The workshop organizers were really surprised at the number of participants that were willing to present their work in this session. Because of this, the extra presentations were added to the afternoon session.

This extra session provided a good overview of the advances that have been made with the implementation of OAI-MPH compliant services in the last months. The audience had a clear perception that these services are becoming concrete and that the availability of services is going to stimulate the adoption of the open archive approach by many other organizations.

2.5 Design of The PORTA EUROPA Portal (PEP) Pilot Project

by Marco Pirri (University of Florence)

This talk concerns the conception of an OAI compliant service that can manage three different digital historical archives maintained by the European University Institute (EUI) in Florence. This situation requires careful consideration of interoperability issues related to uniform naming, metadata formats, document models and access protocols for the different data sources.

In this talk we will present the design approach for the digital archives federation services to be developed in the Porta Europa Portal (PEP) Pilot Project. The PEP pilot project specialised portal should provide high quality information, selected according to the criteria of originality, accuracy, credibility together with the cultural and political pluralism derived from the EUI's profile. The information in Porta Europa will be: relevant, reliable, searchable and retrievable.

To test the feasibility and the impact of the PEP project the EUI committed itself to the development of a PEP prototype concerning historic topics. To this extent, among the various available digital historical archives three of them were chosen for the implementation of the pilot. Our approach in solving problems of standardization and interoperability in the PEP pilot project is based on two main issues:

- Metadata standard (Dublin Core)
- Protocols (OAI-PMH)

The PEP (Porta Europa Portal) project refers to the integration of three digital libraries related to European history topics: Voices on Europe, Virtual Library and Biblio library catalogue.

Each of these data source is characterized by:

- a collection of data objects (books, journals, documents, multimedia objects etc.) available locally or through the network
- a collection of metadata structures

- a collection of services (access methods, management functions, logging/statistics, etc.)
- a domain focus (topic)
- a community of users

Of course the need of integrating the three data sources comes from the topic (European history) and users community that are common to all three archives.

Voices on Europe; **(**http://wwwarc.iue.it/webpub/Welcome.html**)** Voices on Europe is an archive containing the electronic audio version and electronic transcriptions about a hundred of interviews given by outstanding politician and historians.

WWW-VL (Virtual Library) on European History Integration; **(**http://vlib.iue.it/history/index.html) The Virtual Library (VL) is the Web oldest catalogue, conceived by Tim Berners-Lee. Unlike commercial catalogues, it is run by a loose confederation of volunteers, who compile pages of relevant links for specific areas in which they are expert. The EUI Library Web site contains the complete list of VLs belonging to the WWW VL History Project in the University of Lawrence/Kansas (USA) and mirrored at the European University Institute's Library.

Biblio (the EUI historical archives); (http://www.iue.it/LIB/Catalogue/**)** This is the library catalogue containing more than 250.000 bibliographic records. Access to resources is supported by INNOPAC, well known Library Automation System.

The PEP Pilot Project is being developed according to the following steps:

Analysis of the three data resources; in this part we first understand the current situation of the resources and we identify the main issues involved in each case. Each resource is characterised by different issues that are elicited and therefore faced. This phase end with a detailed description of the metadata formats, document models and access protocols for each of the data sources. The analysis revealed the strong points and the weakness of each digital library setting the basis for the definition of a common document description model.

Definition of the federation architecture (figure 1); the architecture of our federation service is structured in three layers: the data source layer where all information is stored with autonomy of representation and access interfaces, the adapter layer were special adapters have to be implemented to provide uniform access and transform the data source specific model into the global model of the federated system, and the federation layer which is responsible for global data integration using an on purpose database and is the OAI data provider and the User interface that will be the OAI service provider.

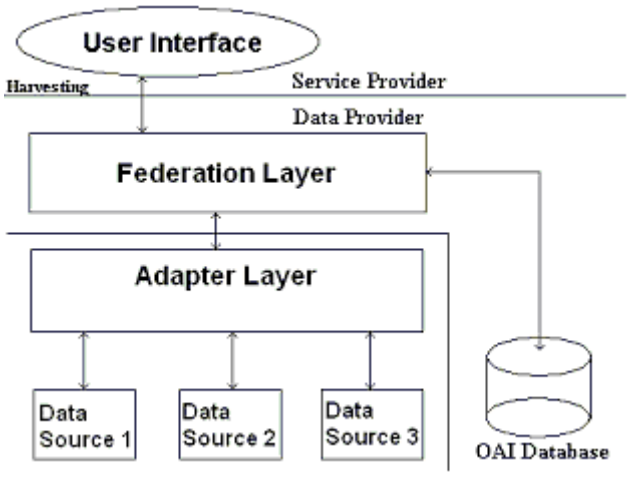

Figure 1. PEP Architecture

Data Source Layer: these are the archives (digital libraries) whose integration we deal with: Voices on Europe, Virtual Library and Biblio library catalogue.

Adapter Layer: this layer provides uniform access to the information, hiding the differences in the data models and query interfaces. Here the metadata are translated from the source specific model into the global model of the federated system. The development of this work is the adoption of the Web services technical framework where a standardized mechanism would be used to describe, locate and communicate with each digital library. The main operation of this layer is the "extraction of data". This operation has to be automatic so that each interface has to be implemented specifically for the resource. As instance SQL queries could be used to extract data from some Data sources (Voices on Europe, Virtual Library) and some external tools such as Innopac tools could be used for the catalogue.

Federation Layer and User Interface: in this two layers is implemented the OAI-PMH, in details:

- Data Provider (The Federation layer)
- Service Provider (The User Interface)

Moreover the Federation layer has to describe Metadata of the three different resources in a common standard to allow in a second step to store them in a unique database. To this extent a common metadata format (Meta Resource Card - MRC) must be devised for the three resources. To effectively address the interoperability issue, the Meta Resource Card should follow the unqualified Dublin Core Standard to define the common fields. In the Federation Layer are implemented interoperability functions, the OAI compliant Data Provider, that is the core of pilot project. The User Interface will be OAI compliant Service Provider and it will use OAI harvesting to extract data. In a first period externally implemented interfaces such as Arc could be used as the Service Provider.

Some remarks

This presentation provided a good example of the motivations that may push a real organization to adopt an open archive approach. The European University Institute (EUI) in Florence is a very big institution with more than six hundred PhD students and researchers from all over the world. This institute has collected many important cultural resources in the past years. It has decided to improve their potential value by opening its repositories to a wider community through a portal.

This presentation has also illustrated a non-standard use of the OAI-PMH. In this application it has been employed as a means for solving an interoperability problem among the heterogeneous repositories that are made available through the Porta Europa Portal rather than to render them harvestable by third party services.

2.6 Innovations through Open Access: Short Presentations

One of the aims of the Open Archives Forum Workshop is to stimulate discussion among the participants and to disseminate information about open archives applications. We decided to reserve a time slot for short presentations in order to achieve this specific aim.

"The Belgian Union Catalogue Project"

by Elisabeth Mazur

This presentation introduced the Belgian Union Catalogue project and it discussed the issues related to the adoption of the OAI-PMH within this project.

"Cultural Heritage Services"

by Muriel Foulonneau (Relais Culture Europe)

The French government is funding research and the setting up of OAI-PMH compliant repository services. Some test-beds have been planned for 2003 for considering organizational issues.

In particular, Relais Culture Europe participates to the Minerva Project. It is in charge with the UK of a workpackage on interoperability in European policies for digitisation. It is considering a discussion with all governments for setting up OAI-PMH services for gathering digitised collections descriptions from various European countries.

Relais Culture Europe also participates in the EMII-DCF project where it focus on the possibilities for re-using and repurposing digital cultural content for European researchers. Organisational issues raised by OAI-PMH protocol are quite interesting in that scope.

"OA-Forum Technical Validation Questionnaire"

by Birgit Matthaei

We started a first Technical Validation Questionnaire in preparation for the first OA-Forum workshop in Pisa. The objective was an overview on status, experiences and future plans relating to the OAI implementations of the workshop participants. At this time exclusively participants of this workshop were involved.

In Pisa we raised a high interest on the results of this small survey and got the feedback that it will be a good idea to collect experiences of a broader spectrum and to learn more about starting points of planned implementations. The impact of this interest based on fundamental questions such as: Is there a large common ground and thus good conditions to cooperate and learn from each other or are requirements so individual that necessarily many further isolated solutions will be developed? Do the existing instruments for implementation fulfil the requirements or should tools and protocols correspond better to the needs of different communities?

Thus in the second questionnaire we added or changed some questions and expanded it.. This long-term questionnaire will be in use until autumn 2003.

So far thirty-three repositories participated. Eleven of them do not have any OAI implementations yet but the possibility is being considered. The distribution of repositories covers all of Europe; and beyond that there are also two participants from overseas. Altogether more than a third of the returned questionnaires come from Germany or UK. If we look at the attendance of different communities it is remarkable that nearly half of the respondents came from Libraries or Archives.

The presentation gave a first summary of the information the participants gave about used software, implementation costs, interoperability, experiences and expectations in different communities and in different countries.

3 THE BREAK-OUT SESSIONS

We decided to increase the number of breakout sessions with respect to the first workshop in order to reduce the number of participants to each session and create the conditions for a better discussion. The topics of the breakout sessions were chosen taking into account both the themes of the workshop and the profile of the participants.

3.1 To what extent can the complex relationships and hierarchies in archives descriptions be disseminated through the OAI-PMH?

This session was facilitated and reported by George Mackenzie (UK). He began by proposing the following questions to the audience:

1. Is it possible to expose collection descriptions via the OAI protocol?
2. Can the dissemination of name authority files be realised with the OAI-PMH?
3. How can we attract relevant audiences / new users using OAI?

The main discussion points were the following:

- OAI is not restricted to the transmission of (unqualified) Dublin Core

- OAI community should make clearer that the support of DC is mandatory only in order to guarantee a minimum of interoperability between data providers and service providers. In contrast to that fact it should be clear that DC actually has no mandatory fields. Consequently it is not a big deal for data providers to support DC.

- Archives (and other repositories) should not concentrate too much on converting there rich metadata formats to unqualified DC

- Archives use EAD as a standard metadata format. Two problems occur when trying to convert it into DC:

    - EAD is sometimes used in different ways / it is interpreted differently (community has to solve this problem)

    - EAD has up to now only a DTD description (XSD has to be created in order to make it useable via the OAI-PMH, should be no a big deal)

- Selective harvesting is not very well usable with the OAI protocol. Example: If you have a very big repository (several million items) and have defined several overlapping set hierarchies, it is not possible to combine two or more of these sets in a request, e.g. want to have only records which describe dissertations and describe documents from medicine and are available as full texts. Possible solutions within the protocol:

    - Create very detailed set hierarchy, e.g. a medicine set within the dissertations set etc.

- Issue only ListIdentifiers requests first for all interesting sets, then look for the identifiers that occur in all the sets (e.g. in the dissertation set *and* within the medicine set *and* within the full texts set) on the service provider side. Finally retrieve the records from the data provider using the GetRecord request type with each found identifier

- Allow dynamic set identifiers, e.g. set=dissertation!medicine!fulltexts – meaning in fact a construct like set=dissertation&set=medicine&set=fulltexts problem: protocol does not allow for data provider to specify the syntax of such dynamic set identifiers (ListSets cannot list all possible combinations)

- Collection level description: is possible in principle (with OAI descriptions of all kinds of resources can be exchanged). Would be reasonable to allow exposing metadata for sets by extending the protocol. Within the OAI-PMH only the identifier and a name can be assigned to a set. A possibility within the protocol definition would be to introduce a kind of application profile where the set information can be exposed interpreting it as an item (define a special set "sets" containing all sets as items, e.g. the set "dissertations" would have an identifier and metadata information encoded in DC and other formats). This solution would also partly solve the ontology problem of set definitions (thesis vs. dissertations).

3.2 Digital Rights Management

Elisabeth Gadd of Loughborough University facilitated the "Digital Library Management" breakout session.

Three issues were discussed in this session:

1.Rights of the Metadata (Data Providers – Service Providers)

- should be explicit within the metadata possibly in the "about" field;
- possibly a set of standard statements that data providers can choose from should be defined;
- it is important to consider what enhancements should be allowed to metadata and what the IPR status of enhanced metadata is.2. Rights of the Documents/Content (Authors – Data Provider)
- who assigns rights metadata? Authors or Data Providers?
- Should the metadata be human readable or machine readable or both?
- machine: should it be use in a technical protection measure?3. OAI and IPR/DRMOAI is looking at projects like:

- Creative Commons
  - suggested four different machine readable rights statements
  - legal licences are attached to those
- RoMEO
  - suggest rights metadata solution for eprints
  - possibly Creative Commons
  - possibly Open Digital Rights Language

3.3 The role of libraries in open archives

Elisabeth Gadd of Loughborough University facilitated this breakout session.

The breakout session investigated a number of issues:

1. Opportunities
   - enabling interoperability;
   - integrate fragmented libraries (e.g. within one university)
   - enhance role of central libraries within universities

- provide integrated library service to users (students, researches, ...), in one single access point
- co-operation between library and computing centres (necessary but not always realised)

2. Metadata
   - not always created centrally
   - central (university) library should define minimum standard to enable basic library information exchange. They should also formulate guidelines for the creation of metadata (for librarians mainly)
   - problem: people who are creating metadata often are not librarians
3. Right management
   - libraries should be involved in the creation of rights metadata ("meta metadata");
   - right statements on documents and metadata should be archived ;
   - example of different handling: worldwide availability of information vs. availability within the campus
   - special case: article dissertations (similar to master theses in Germany) where the single articles have been already published by another publisher but should be hosted (re-published) by the document server as well.
4. Sustainability
   - problem: solutions often developed by project staff
   - projects should lead over new tasks and know-how to long term staff
   - project time should be used to reach the "critical mass" of documents/open archived resources
5. Awareness raising
   - encourage institutions and authors (mainly researchers)
   - marketing
   - enhance credibility of new solutions and the open archives approach as such
   - use political coalitions (e.g. in Florence / Italy it has been made obligatory to publish dissertations electronically)
   - libraries should be an advocate of open archives
6. "major" concerns

- problem: integration of new solutions in existing library workflows / processes

3.4 Common services for libraries and archives

Josè Borbinha of the National Science of Portugal facilitated and reported this breakout session.

Certainly the integration of libraries and archives can open a range of new service possibilities that goes far beyond the simple search. However, before making these services real a number of problems has still to be solved.

Common service for libraries and archives must take into account the heterogeneity of the two kinds of organizations. The table below compares some characteristics of the two organisations:

Services	
Library is/has	*Archive is/has*
Bibliographic descriptions	Archival descriptions
Authority descriptions	Authority descriptions (including contextual desc. of creation)
Multicopy contents	Unique documents arranged in groups and subgroups
Possibly digital access to the resource	Possibly digital access to the resource

There are aspects like authority descriptions that are already quite closed, other like bibliographic/archival descriptions that are still very far from each another.

Progresses in the interoperability between the two types of organisations will be only achieved if convincing business models will push it. These models also depends from the request of common services. The general impression is that something is moving in this direction.

3.5 Subject interoperability

The breakout session on subject interoperability was facilitated by Paul Child. It began by posing the following questions:

1. Is it necessary to have interoperability between the subject terms used in different repositories harvested via OAI-PMH?
2. Is it possible to have interoperability between the subject terms used in different repositories harvested via OAI-PMH?
3. If necessary and possible, then at what level?
4. If necessary and possible, then how much interoperability?

A number of subject interoperability issues were discussed. One challenge to overcome is the inconsistency introduced by the varying levels of granularity of subject indexing across repositories. Data providers and service providers may choose differing levels of granularity in line with their assessment of the requirements of their own target user communities or the size of their collections of resources or metadata. Subject interoperability could be seen as one of the major enhancements, or added value, that a service provider could offer. Language differences were seen to be a serious barrier to interoperability. Disambiguation between subject areas also presents difficulties for subject interoperability. Quality of choice of subject terms was also seen as an issue, particularly where metadata is created by self-archiving authors, because of the difficulty of achieving consistency among those assigning subject terms.

Various elements of a possible solution or solutions were suggested, including automatic or semi-automatic classification and cross-walks or mapping among subject schemes. The primary suggestion was that just as DC is the lowest common denominator for metadata description providing low-barrier interoperability across description, OAI needs a multilingual lowest common denominator for subjects. Perhaps this could be used for harvesting sets, and the service provider could then choose whether or not to enhance the subject descriptions as part of its service. In discussion, it emerged that many participants believe that a community of a manageable size (defined either geographically or by discipline) was required in order to agree on a common subject scheme. Everyone agreed that a subject approach is necessary in order to reveal hidden resources.

Several actions were agreed. The Organisation Issues working group (see below) intends to compile guidelines for best practice in the subject approach for open archives, and this will be further discussed via the OA-Forum "info" mailing list. In addition, examples and links will be collected and posted on the OA-Forum web site.

3.6 Meeting of the Organisational Issues Working Group

The Organisational Issues (OI!) working group formed as a result of discussion in a breakout session of the first OA-Forum workshop in Pisa, May 2002. The aim of the working group is to develop use scenarios and best-practice guidelines relating to organisational issues for open archives. It agreed to conduct discussions to flesh out the details of the work it would undertake (using the OA-Forum public mailing list that was set up following the workshop), and to meet again at further OA-Forum workshops. The main outcomes of the meeting in Lisbon were

1. Email task groups set up as follows:
   - Metadata Issues - William Nixon leads
   - Business Models - Paul Child leads
   - Role for CMS - Dennis Nicholson leads
2. Agreement to draft guidelines for best practice for discussion at a further meeting during the 3rd OA-Forum workshop in Berlin (March 2003)
3. Agreement to ask 'leaders' of separate breakouts (see above) on Subject issues and IPR to set up similar task groups to those detailed above.

4 OUTCOMES AND ACTIONS

Many thanks are due to the National Library of Portugal – for the venue and bringing us to Lisbon, for the ideas and enthusiasm and all the hospitality and hard work that contributed so much to the success of this workshop. Special thanks to Josè Borbinha and his colleagues Eulalia and Madelena. We also thank our invited speakers, especially those who stepped in

with presentations on their projects and initiatives when bad weather in the USA triggered a last-minute change of programme.

Workshop participants shared knowledge and experience and in some cases initiated alliances for joint work. They represented both archival organisations and libraries, and both organisation-based and subject-domain-based repositories and services. Across these differences areas of common interest were identified, with the following themes emerging across presentations, breakouts and informal discussions as of key importance: users, multiliguality, sustainability of repositories and services, IPR, open standards, the applicability of OAI as a solution in particular communities and types of organisation, and need for guidance on best practice. With regard to users, many questions were raised. Do we know who our users are, and do we understand their requirements? Could the open archives approach bring new users to our collections? Do we understand users who are creators of the resources in our repositories? The difficulties of multilinguality in the European context relate not only to users groups, but also to resources and metadata (including subject terms and classification schemes) in a wide variety of languages.

Issues of sustainability of repositories and services often focus on how to make the transition from project funding to business models suitable to ongoing service provision. The importance of open standards was frequently emphasised but there are clearly outstanding concerns. OAI does seem to provide low-barrier interoperability, but is not on its own a sufficient basis for providing interoperability-based services. Service providers will have to find ways of adding value beyond simple searching of metadata, and may find a requirement still exists for distributed searching, perhaps implementing SRU and Z39.50 in addition to OAI-PMH in some cases. The question of the scope of OAI and how it might be extended in the future arose, and some people suggested a possible role in or for knowledge management.

The need for guidance on best practice was acknowledged, and the challenge this represents in an area where there is as yet little practice on which to draw. However, more and more repositories and even services are emerging. There had been real progress in projects and service development in the six months since the first OA-Forum workshop. Workshop participants will continue to share knowledge and experience across Europe through forged alliances, through the OA-Forum and other mailing lists, and through the OA-Forum information source. All participants are encouraged to register their services, repositories and projects with the information source (**www.oaforum.org/oaf_db/index.php**). The work of the Organisational Issues (OI!) working group will continue, with the immediate aim of drafting best-practice guidelines in various areas as outlined above. All are welcome to participate in the work of the OI! working group in various one or more of these areas, via the OA-Forum mailing list and further meetings at future OA-Forum workshops.

In addition, two expert reports commissioned by the OA-Forum were first presented at this workshop. The presentations on the reports, one on Intellectual Property Rights (IPR) and OAI and the other on archival organisations and OAI, are described above in the section on invited presentations. The full reports will be published on the OA-Forum web site once they have been through a final, post-workshop revision process, and their publication will be publicised through the OA-Forum “info” mailing list and other appropriate mailing lists.

5 LESSON LEARNED

About the diffusion and use of OAI-PMH:

- The workshop revealed that there is now in Europe a strong on-going activity in the development of OAI- compliant services.
- Many communities are experimenting with the OAI-PMH as a means for intra-community interoperability. In many case they are interested in exporting their own metadata format and they do not understand why they are forced to expose also a Dublin Core description of their resources even if they do not use it.

About the Workshop:

- The high percentage of workshop participants that attended the tutorial (thirty out of the fifty registered to the workshop) demonstrated that there is a strong need for introductions to the OAI-PMH, especially if the workshop is addressed to communities different from the e-print one.
- Social events and long coffee breaks are useful. At this second workshop there has been a lot of informal discussion and new collaborative links have been established.
- A well organised workshop environment is also very useful. The possibility of having easy access to computer facilities, made it possible for many people to show their systems to the others during the breaks.
- It is very important that breakout sessions be not too crowded and that the topic be very focused. Reaching a concrete result in one hour discussion when the participants have heterogeneous backgrounds is difficult.
- It is also very important to stimulate the setting up of working groups on particular subjects because the members of these groups then continue their exchange of ideas outside the context of the workshop.