

## OPEN ARCHIVES FORUM: PROJECT DELIVERABLE

<b>Project Number:</b>	IST-2001-32015
<b>Project Title:</b>	Open Archives Forum
<b>Deliverable Type:</b>	Public

<b>Deliverable Number:</b>	D4.6
<b>Contractual Date of Delivery:</b>	July 2003
<b>Actual Date of Delivery:</b>	January 2004
<b>Title of Deliverable:</b>	Workshop Report 4
<b>Workpackage contributing to the Deliverable:</b>	WP4
<b>Nature of the Deliverable:</b>	Report
<b>URL:</b>	<a href="http://www.oaforum.org/otherfiles/oaf_d46_workshop4.pdf">http://www.oaforum.org/otherfiles/oaf_d46_workshop4.pdf</a>
<b>Author:</b>	Donatella Castelli, Philip Hunter, Birgit Matthaehi
<b>Contact Details:</b>	Donatella Castelli, +39 050 3152902

<b>Abstract</b>	This report documents the Fourth Open Archives Workshop held in Bath (UK) on 4 <sup>th</sup> -5 <sup>th</sup> of September 2003.
<b>Keywords</b>	Open archives, Open Archives Forum, Open Archives Initiative, e-print archives, multimedia archives, intellectual property rights.

<b>Distribution List:</b>	Public
<b>Issue:</b>	Version 1.0
<b>Reference:</b>	32015-DEL4.6-20040130
<b>Total Number of Pages:</b>	15

---

<b>1</b>	<b>GENERAL OVERVIEW .....</b>	<b>3</b>
<b>2</b>	<b>THE INVITED PRESENTATIONS .....</b>	<b>3</b>
2.1	OPEN ARCHIVES FORUM - TECHNICAL VALIDATION .....	4
2.2	INTRODUCTION TO THE NEW OA-FORUM ONLINE OAI-PMH TUTORIAL.....	5
2.3	IMLS NLG COLLECTION REGISTRY AND ITEM-LEVEL METADATA REPOSITORY PROJECT AT THE UNIVERSITY OF ILLINOIS .....	5
2.4	OPEN ARCHIVES INITIATIVES PROTOCOL FOR METADATA HARVESTING - PRACTICES FOR THE CULTURAL HERITAGE SECTOR .....	6
2.5	THE ROMEO PROJECT: IPR ISSUES FACING OPEN ACCESS .....	7
2.6	OPEN ARCHIVES INITIATIVE: WHERE WE ARE, WHERE WE ARE GOING .....	8
2.7	THESES ALIVE! - AN ETD MANAGEMENT SYSTEM FOR THE UK .....	8
2.8	AQUITAINE PATRIMOINES & CYBERDOCS .....	9
2.9	DARE: DIGITAL ACADEMIC REPOSITORIES - A NEW AGE IN ACADEMIC INFORMATION PROVISION IN THE NETHERLANDS .....	10
2.10	SCIENCE AND CULTURE - DEVELOPING A KNOWLEDGE SITE IN DISTRIBUTED INFORMATION ENVIRONMENTS .....	10
2.11	HARVESTING THE FITZWILLIAM (JISC-FAIR PROJECT) .....	11
2.12	OA-FORUM ORGANISATIONAL ISSUES WORKING GROUP REPORT .....	11
2.13	OPEN ARCHIVES, OPEN ACCESS AND THE SCHOLARLY COMMUNICATION PROCESS .....	11
<b>3</b>	<b>THE BREAKOUT SESSIONS.....</b>	<b>12</b>
3.1	QUALITY ISSUES .....	12
3.2	E-THESES.....	12
3.3	SUSTAINABILITY ISSUES .....	13
<b>4</b>	<b>THE PANEL SESSION .....</b>	<b>13</b>
	<i>Panel Session and discussion on the future of Open Archives.....</i>	<i>13</i>
<b>5</b>	<b>OUTCOMES.....</b>	<b>14</b>
<b>6</b>	<b>LESSON LEARNED.....</b>	<b>14</b>
	ABOUT THE DIFFUSION AND USE OF OAI-PMH: .....	14
	ABOUT THE WORKSHOP: .....	15

## 1 GENERAL OVERVIEW

This workshop focused on good practice in the implementation of open archives, with an eye to future development of the technology. A particular theme of the workshop was the use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) in the area of Cultural Heritage, where multimedia is an important issue. The workshop looked also at the use of the OAI-PMH protocol as a way of publishing information about university theses, and how that might contribute to developing useful content for institutional (as opposed to subject-based) e-print archives.

The workshop was build on issues discussed during the whole project. In particular,

- the European experience of open archives regarding technical issues, organisational issues and Intellectual Property Rights, were reviewed;
- a report on organisational issues written by an OA-Forum working group was illustrated;
- an online tutorial, delivered as part of the OA-Forum activities, which gives guidance to those wishing to implement a project using the OAI-PMH was presented;
- breakout sessions offered the opportunity to discuss issues of practice with others working at the sharp end of implementation;
- a poster session provided the opportunity to disseminate information about existing projects in the field.
- a panel session closing the second day of the workshop offered the opportunity to exchange views about the future direction of open archives, and about our experience of the open archives approach so far.

This last workshop had, as all the previous workshops, a very good attendance, with more than seventy registered participants attending which included ten invited speakers,. There were representatives from many EU and Nationally funded projects. Eleven countries were represented: Austria, Germany, Italy, The Netherlands, Norway, United Kingdom, USA Portugal, Canada, Belgium, and France. The audience was quite heterogeneous, there were project managers, librarians, researchers, system developers, etc.

There was a significant number of presentations by projects working within the cultural heritage sector (the number was commented on favourably at the time by one of the participants), illustrated that the scope of the OAI PMH is much wider than it might have been imagined at the time it was conceived at the Santa Fe Convention in 1999. Much of the discussion during the workshop focussed on organisational issues and the sustainability of initiatives and programmes making use of the OAI Protocol for Metadata Harvesting. It was also interesting to hear Carl Lagoze speaking at some length on where he thinks the Protocol is going, and where metadata creation is going also.

## 2 THE INVITED PRESENTATIONS

This section contains slight modified versions of the abstracts of the presentations provided us by the invited speakers and any brief comments/notes 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/bath\\_programme.php](http://www.oaforum.org/workshops/bath_programme.php)).

The Workshop started with two presentations given by members of the OA-Forum that described two outcomes of the project: the analysis of the results of the technical validation questionnaire on the open archives activity in Europe, collected during the two years of the project, and an on-line OAI-PHM tutorial, based on the experience of the successful pre-workshop tutorials held in Lisbon (2002) and Berlin (2003). These two presentations were followed by three breakout sessions that touched key issues identified during the previous workshops. A series of invited presentations, which mostly described the results of past and on-going projects, followed until the end of the workshop. A presentation on the current status of the Open Archives Initiative (OAI) and on its future plans was also given by Carl Lagoze, a member of the OAI steering committee.

## 2.1 Open Archives Forum - Technical Validation

*by Birgit Matthaei (Humboldt University)*

A long-term Technical Validation Questionnaire that was started in spring 2002 will close after the Bath workshop. The objective is to provide an overview on status, experiences and future plans regarding the European OAI implementations. The focus of interest is on fundamental questions like: Is there a large common ground and therefore good conditions for cooperating and learning from each other, or are requirements so individual that necessarily many further isolated solutions will be developed? Do the existing instruments for implementation fulfil all requirements or should tools and protocols correspond more than before to the needs of different communities?

Furthermore the project provides information from those projects that have already dealt with the integration of existing technologies. An information space was created in form of a searchable web-based database on projects, software, implementations, and services, that allows interested parties to search for potential project partners, for metadata standards and for information about interoperability issues and to share project developments.

Apart from this offer for self-registration to encourage exchange, systematic inventories of repositories, services and tools were gathered, own experiences with the metadata harvesting pilot were reported or with the implementation of OAI Services focusing on different objectives as well as experiences made with available software tools.

The presentation offers a preview on some final results of the questionnaire about used software, implementation costs, offered spectrum and interoperability, experiences and expectations in different communities and in different countries. Compared with loops to aspects of the other resources a short summarising of the OA-Forum project Technical Validation workpackage will be presented to provide an overview of European activities on OAI in relation to worldwide activities.

### Some remarks

The work done as part of the OA-Forum technical validation activity has been really appreciated since it has provided a lot of feedback to the OAI protocol developers and adopters. The OAI-PMH designers have had the opportunity to be informed about the difficulties encountered by the providers when implementing the protocol and about spreading of the OAI data and service providers in Europe. The adopters have had access to

information about the motivations and the experiences collected by other organisations that have opened their archives.

## **2.2 Introduction to the new OA-Forum online OAI-PMH Tutorial**

*by Leona Carpenter (Freelance consultant)*

In the Open Archives Forum mid-project review, the introductory OAI-PMH tutorials provided in conjunction with the Forum's workshops were identified as an important output of the project, and the project was advised that “the tutorial should become a strategic product of the project” and that “the project should package the tutorial as a clear product” that would remain available beyond the term of EU project funding.

The only way to achieve this was to create an online tutorial. The tutorial would cover a range of topics at an introductory level, focussing on technical aspects of the implementation of the OAI-PMH, and based largely on the material presented at the Lisbon and Berlin pre-workshop tutorials.

The challenges addressed in developing an online version of the tutorial included turning presentation-style outlines into a continuous narrative, providing some level of interactivity to compensate for lack of access to tutors for questions and discussion, and ensuring that the tutorial can easily be maintained in order to extend the period of its usefulness and allow for the possibility of providing versions in languages other than English.

The tutorial, which is available at <http://www.oaforum.org/tutorial/>, breaks down into two main areas: OAI for beginners and OAI-PMH (OAI Protocol for Metadata Harvesting). The beginners section provides some background on what the protocol is and what it does; technical detail is at an introductory level for those considering implementing the protocol. The second part relates to various topics covered in workshop tutorials, largely basic ideas, history and development and the technical basis for implementing the protocol. It also covers XML schemas and metadata formats.

### **Some remarks**

The language of the tutorial was recognised as a very important issue for the European OAI potential adopters. Some of the participants to the Workshop spontaneously offered their support for translating the current version of the tutorial into their own native language.

## **2.3 IMLS NLG Collection Registry and Item-Level Metadata Repository Project at the University of Illinois**

*by Timothy Cole, Mathematics Librarian & Professor of Library Administration, University of Illinois at Urbana-Champaign, USA*

The presentation will provide background information on the Institute Museum and Library Services National Leadership Grant Program (IMLS NLG) in the US which funds research and demonstration, digitization, preservation, model programs and new technology in the Library and Museums arena.

An overview of the "IMLS Framework of Guidance for Building Good Digital Collections", published in November 2001 will be given and four general recommendations from the IMLS Forum will be presented. One of these recommendations indicates that the IMLS should encourage the integration of an archiving component into every project plan by requiring a description of how data will be preserved.

An overview will also be provided of an interesting OAI project in which the University of Illinois had been involved under a Mellon grant. The primary objective of this project was to create and demonstrate OAI tools; build a portal using aggregated metadata describing cultural heritage resources; investigate use of EAD (Encoded Archival Description) metadata in an OAI context and research the utility of aggregated metadata. The portal currently has 25 OAI data providers and aggregates some 479,000 metadata items.

A preliminary version of OAI guidelines for static repositories, as a lower barrier option for exposing relatively static and small collections of metadata will be introduced.

The presentation will also discuss also about the OAI services. There are now OAI data provider services being built into many popular digital library applications such as ContentDM, Encompass, DLXS, Dspace, and Eprints.org. However, some of the implementations are limited in that they may support the oai\_dc metadata schema only, or have limited feature sets and metadata mappings which may not be configurable.

### **Some remarks**

During the presentation some concerns were expressed over IP rights issues and uncertainties as to whether licences limit metadata sharing; Timothy Cole's view is that machine readable IP rights attributes are needed to facilitate reuse.

In closing Timothy Cole noticed that considering OAI in context; descriptive item-level metadata alone appears not to be sufficient. It needs to be combined with collection descriptions, user annotations, machine-generated clustering etc.; it is important to note that OAI-PMH is not limited to item-level descriptive metadata.

## **2.4 Open Archives Initiatives Protocol for Metadata Harvesting - Practices for the cultural heritage sector**

*by Muriel Foulonneau (Relais Culture Europe)*

The Open Archives Initiative Protocol for metadata harvesting underlies an organisational model with data providers, service providers and aggregators. In the context of cultural heritage actors, those aggregators can represent an opportunity to match the missing competences in small institutions which need to expose their metadata.

Indeed, the protocol can organise value-added services based on the material created in memory organisations for ages to describe their asset. It can be used to make their resources accessible on the Web, to let other institutions set up cross-institutions and possibly cross-domain services, to allow their data to be re-used in various contexts. It is also becoming a standard way to exchange XML formatted data and to synchronise repositories for downstream services such as name authorities.

Still, the protocol does not remove the interoperability barriers and issues on metadata quality, on heterogeneous original descriptions, organisational issues such as data update, integrated access to aggregated and heterogeneous resources must be dealt with, within a proper partnership between data providers and the service provider.

The conditions of use, guarantees of quality on metadata, content, service and data provider's Websites shall be included in an agreement, whether a charter or a proper contract to ensure responsibilities of each partner, quality of its performances and legal issues related to cultural heritage content.

The development of OAI-based services in the cultural heritage sector seems to be led by several key services funded to set up a proper OAI framework. It appears easier for the service provider to directly set up an aggregator for those institutions for smaller institutions, with poor competences in ICTs.

Still, many areas and subjects are still to be tested, such as the use of OAI with large schemas, services based on data re-use, mixing information retrieval through both full text and metadata especially, the management of aggregated resources is a challenge to service provider since they do not control the evolution of their collections.

The Community report on practices the Open Archives Initiative of Metadata Harvesting by cultural heritage actors is available on the OAForum Website <http://www.oaforum.org/documents/>. It studies the conditions for memory organisations to use the protocol, what they can do with it, the issues to take into account and the solutions already implemented or being considered.

### **Some Remarks**

This presentation was really appreciated by the audience since it provided a picture of the situation in a large and important application sector that can very profitably exploit the open archives approach. There were many follow-ups to this presentations and new collaboration links were established by groups working in the culture heritage sector.

## **2.5 The RoMEO Project: IPR issues facing open access**

*by Elisabeth Gadd (Loughborough University)*

Inspired by the Open Archives Initiative, the United Kingdom (UK) Joint Information Systems Committee (JISC) established the FAIR (Focus on Access to Institutional Repositories) programme in 2002. One of the programme's objectives was to "explore the challenges associated with disclosure and sharing [of content], including IPR and the role of institutional repositories". To this end, the JISC funded a one-year project called RoMEO (Rights Metadata for Open archiving). RoMEO, which took place between 2002–2003, specifically looked at the self-archiving of academic research papers, and the subsequent disclosure and harvesting of metadata about those papers using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) by OAI Data and Service Providers.

The RoMEO project aimed to develop simple rights metadata by which academics could protect their research papers in an open-access environment and also to develop a means by which OAI Data and Service Providers could protect their open-access metadata. RoMEO proposed to show how such rights solutions might be disclosed and harvested under OAI-PMH.

The RoMEO project was divided into two phases: a data-gathering phase and a development phase.

More information about these phases can be found at <http://www.dlib.org/dlib/september03/gadd/09gadd.html>

### **Some Remarks**

The project team produced a series of six studies based on their work [Gadd, Oppenheim, and Proberts,

<http://www.lboro.ac.uk/departments/ls/disresearch/romeo/Romeo%20Deliverables.htm>]. In particular, an article that provides an overview of all the activities of the RoMEO project and reports on its key findings and recommendations can be found on D-LibMagazine (<http://www.dlib.org/dlib/september03/gadd/09gadd.html>)

## **2.6 Open Archives Initiative: Where we are, where we are going**

*by Carl Lagoze (Cornell Information Science)*

This presentation will review the current status of the OAI-PHM protocol. This protocol is currently a de facto standard for Internet information exchange which has been deployed extensively and internationally in the framework of (digital) libraries, museums, e-print repositories, research projects. In particular, the presentation will focus on its application in the context of the NSDL and will outline the lesson learned and the directions that have been identified for the next future.

### **Some Remarks**

The presentation given by Carl Lagoze was provocative and stimulated discussion among the participants. In particular, he pointed out that the growing quality that automatic methods for metadata extraction have reached recently is pushing him to reconsider the superiority of the manual description, especially given the high amount of digital information that must be described today and the cost that this description implies. This statement was highly criticised by many librarians that still believe in the human intervention as the best and unique mean for achieving an high quality content description and classification.

## **2.7 Theses Alive! - an ETD management system for the UK**

*by Theo Andrew and Richard Jones (Edinburgh University Library)*

The JISC-funded *Theses Alive!* project based at the University of Edinburgh, with pilot partners at Cambridge, Cranfield, Leeds, Manchester Metropolitan, started in November 2002. With the general aim of promoting ETDs, the *Theses Alive!* project is developing an OAI-compliant thesis archive and submission system for use in all participating universities. With this infrastructure in place our target is to enable e-theses to be published on the web to the extent that a minimum of 500 e-theses exist within the UK segment of the NDLTD after 2 years.

The project will also look at developing and implementing a metadata export system (crosswalk) capable of delivering our metadata to relevant metadata repositories for UK thesis information (e.g. the British Library, the Index to Theses service). In addition, we would like to help other universities by producing a 'checklist approach' to use as they develop e-theses capability.



The project started by evaluating the main open source software packages developed for the management of digital objects, specifically regarding e-theses management. This comparative evaluation is freely available from our website (<http://www.thesesalive.ac.uk/archive/ComparativeEvaluation.pdf>).

The *Theses Alive!* project is scheduled to 'go live' in December 2003 with the launch of our e-theses repository within the 'Edinburgh Research Archives'. In Spring 2004 we aim to test a submission system for the examination of e-theses within the University of Edinburgh. To do so we are not only looking at the physical act of building and populating an e-theses archive, but in doing so we will have to address the requirements of the university administrators, examiners, students and academics. This requires a new role for the Library/Information Services in that the service we are providing is not just simply a replacement for traditional interlibrary loan

## 2.8 Aquitaine Patrimoines & Cyberdocs

by Rasik Pandey, Developer, AJLSM

*Aquitaine Patrimoines* (<http://ajlsm-sdx.hopto.org/sdx-22h/pa-portail/>) is a cultural heritage portal or service provider in OAI terms. The portal attempts to pull together diverse cultural information sourced from libraries, media libraries, archives, museums, cultural heritage education centers, centers of documentation, etc. The data, harvested by means of the OAI protocol, describes various heritage resources concerning the Aquitaine region of France. The contributors and actors from international to local levels are interested in validating methodologies and technologies for sharing resources in a distributed environment as well as investigating the services which can be derived from these sources. The experiences encountered during the development of this portal brought to light the issues surrounding the creation an OAI service provider for cultural heritage purposes which ranged from technical to content concerns.

*Cyberdocs* (<http://sourcesup.cru.fr/cybertheses/>, <http://mirror-fr.cybertheses.org/>) is a free/open-source platform for publishing structured electronic documents. Cyberdocs was realized as a result of experiences from the Cyberthèses project, an information processing platform for scholarly publishing initiated by Presses de l'Université de Montréal in 1997. The platform consists of modules serving three purposes: conversion, management, and publication (including OAI Repository and Harvester implementations). The use of standard technologies and open-source software to create this platform provides various technical and organizational benefits which can be of value to a larger community. The future of the Cyberdocs project will focus on incorporation more resources to support various types electronic structured documents, greater support for multilingualism in interfaces, and encouraging the involvement of more interested parties; therefore, allowing the platform to grow and become more useful to a larger public.

Both of these projects incorporate SDX, an open-source system for searching and publishing XML documents, which is built upon the Apache Cocoon framework and incorporates the Apache Lucene search-engine. (<http://sdx.culture.fr/sdx/> (documentation in English currently NOT available), <http://cocoon.apache.org>, <http://jakarta.apache.org/lucene/>)

### Some Remarks

The presentation highlighted that one of the major difficulty found in carrying out the *Aquitaine Patrimoines* project has been in trying to find common threads in diverse content by which resources can be presented such that value can be added by the service provider. They also found that DC was insufficient and added seven additional terms.

## **2.9 DARE: Digital Academic Repositories - a new age in academic information provision in the Netherlands**

*by Henk Ellermann, Project Leader, Erasmus University, Rotterdam*

DARE is the Dutch equivalent of the open archives movement, or Digital Academic Repositories. It aims to have repositories at all universities catering for the archive of all academic output which would include: theses, articles, data sets, lecture slides, etc. A major aim is to enable reuse of such resources and the provision of services based on the repositories. A tender process is currently in progress for the provision of services.

### **Some Remarks**

When the DARE project was presented at the OA-Forum workshop in Berlin it was moving its first steps. DARE participants started to discuss the use of the OAI-PHM during this workshop stimulated by the Herbert Van de Sompel's suggestions. This presentation highlighted the impressive progresses made by DARE in the short period and demonstrated the power and the simplicity of adoption of the OAI-PHM.

## **2.10 Science and Culture - Developing a Knowledge Site in Distributed Information Environments**

*by Ann Borda, Head of Collections Multimedia, Science Museum Alpay Beler, IS Architect Science Museum Nick Wyatt, Collections Services Librarian, Science Museum Library*

This presentation will describe a project being undertaken by the Science Museum in London with the support of funding from the New Opportunities Fund (NOF). The project involves several museums in the UK, aiming to make a rich quantity of materials and collections accessible. It further aims to contextualise information through intelligent display, searching and relational linking and to develop user-focused activities and personalisation tools. NOF projects are required to use DC and XML, although flexibility is allowed in the implementation. Content is drawn from 5 disparate databases amounting to 40,000 digitised images and associated text; 30,000 library records; 10,000 object records and 50 narrative topics. The project has found authority control to be an essential feature if consistency in the data is required. At the collection level DC and Research Support Libraries Programme Collection Level Description (RSLP CLD) elements are used. Future work is envisaged in the area of creating communities based on interest groups linked to subject hierarchies.

### **2.11 Some remarks:**

This presentation, given by three participants to this project, clearly exemplified how large institutions are now moving toward digitations. It was clear that this large movement will dramatically change the culture heritage sector since it will create a sharing market where the valuable digitised resources will be shared under appropriate control policies.

## **2.12 Harvesting the Fitzwilliam (JISC-FAIR project)**

*by Shaun Osborne, Project Manager, Fitzwilliam Museum, University of Cambridge*

The Fitzwilliam is the Art Museum of the University of Cambridge, providing access to teaching, learning and research as well as access to the general public. The museum has a diverse collection of 500,000 objects managed in 5 curatorial groups: manuscripts and printed books; paintings, drawings and prints; antiquities; applied arts; coins and medals. The goal of the work being undertaken is to develop a unified catalogue of object records which can be used for: collections management; teaching, learning and research; and electronic access. The work is being funded by JISC's FAIR (Focus on Access to Institutional Resources) Programme to support the preparation of records and images to provide access using the OAI-PMH.

## **2.13 OA-Forum Organisational Issues Working Group report**

*by Paul Child, Project Manager: Artworld, University of Cambridge*

This presentation will summarise the review of organisational issues which have emerged through discussions within the OA-Forum project and it will cover the following issues: business models, intellectual property rights, quality assurance, metadata, interoperability, content management systems, and the importance of organisational issues. The review forms a part of the set of deliverables of the project.

### **Some Remarks**

The OA-Forum Working Group on Organisational Issues was spontaneously established at the first OA-Forum workshop in Pisa and its participants regularly met at all the successive workshops. This presentation summarised the results achieved through these discussions and highlighted the wish that the links created by this activity are preserved also beyond the end of the OA-Forum project.

## **2.14 Open Archives, Open Access and the Scholarly Communication Process**

*by David Prosser, Director, SPARC Europe*

The scholarly publishing process comprises four functions: registration (establishing intellectual priority); certification (certifying the quality/validity of the research); awareness (assuring accessibility of research); archiving (preserving research for future use). Looking at each function from an institutional repository perspective, it is clear that registration can be achieved. Certification on the other hand, via the process of peer review is independent of the medium. Awareness can be enhanced by OAI-compliance and interoperability, so that search engines can index the metadata harvested from federated repositories. The advantage in terms of preservation is that an institutional repository helps to put librarians rather than journal publishers in charge of digital archiving.

### **3 THE BREAKOUT SESSIONS**

The topics of the breakout sessions were chosen keeping into account the requests raised through the evaluation questionnaires by the participants to the previous workshops.

#### **3.1 Quality issues**

*facilitated by Rachel Heery (UKOLN, University of Bath)*

This session discussed the identification of good practice and the formulation of generic guidelines for the future. The distributed nature of systems means that quality becomes more and more important the greater the number of contributions. Quality Assurance (QA) needs to be considered right from the start (e.g. all JISC projects are now required to take QA into account, with additional funding for this specific purpose). QA requirements change to meet technical advances, so the process needs to be iterative.

There was a question over who should be responsible for QA, Data Providers, Service Providers or Clusters? At the moment OAI-PMH mandates DC (Dublin Core) as the lowest common denominator for metadata description -this unfortunately makes the default of a very low specificity. The guidelines need to encourage the sharing of richer metadata records such as MARC and IEEE LOM, to enable records to be re-tasked and re-purposed.

#### **3.2 e-Theses**

*facilitated by Jessica Lindholm (Electronic Information Services Librarian, Lund University Libraries, Sweden)*

The session covered a variety of e-theses-related issues from a range of perspectives. The group identified metadata, workflows, copyright issues, preservation and convincing decision makers as themes of special interest for the session.

In the time available it was not possible to discuss the topics that had been identified as being of common interest in great depth, but the overall impression of the current situation was encouraging. Many developmental projects are underway, some universities are already requiring electronic submission, and there was an indication from one participant that e-theses work, initially funded on a project basis at his university, is now considered a routine part of the work of library and computer staff (and funded accordingly).

With a large number of people undertaking research in this area, problems are being resolved and pockets of expertise are emerging. Generic ready-to-use tools are desired, but local needs are likely to create high levels of expectation and demand on such tools. They are to handle the entire publication life chain, producing a 'useful' document for resource discovery which also accommodates preservation needs. The tools are to support the learning environment as well as the archival environment.

The participants felt that they would benefit from closer collaboration, particularly in terms of sharing best practice in several of the areas that were discussed. For example, a joint approach on standardised copyright agreements between authors and publishers could be interesting to investigate in depth. As more institutions adopt e-theses, it becomes easier to convince decision makers of the merits of this approach.

### 3.3 Sustainability issues

*facilitated by the DARE Project*

In this breakout session participants discussed organisational issues relating to sustainability of open archives. Sustainability entails securing organisational support and providing a service based on an institutional repository to a scientific or scholarly community by building service on top of the archive. A question was raised as to whether a "self-archive" model was appropriate for institutional repositories. Some members of the DARE project reported their experience on this aspect. In particular:

- the Twente representative told that self-archiving has not worked for their institution;
- the Amsterdam representative reported that they are practicing self-archiving in order to check technical aspects. In their experience they found that this archiving procedure is accepted only when all different hierarchical levels are included in the process;
- the Rotterdam representative illustrated the experimentation that is currently underway with DSpace in his institution. In particular, he outlined that the quality control is made manually by a person per faculty, the library only checks the system and takes care of the overall organisation.

The break-out session was closed with a discussion about how the institutions could reduce the self-publication cost by sharing tools either at National or at European level.

## 4 THE PANEL SESSION

### *Panel Session and discussion on the future of Open Archives*

*Tim Cole*

The OAI-PMH technical guidelines have been deemed to be good. But object identity attributes are still not well understood. For example, to what extent does the object's "value-addedness" need to change before it can take on a new identity? Simple DC is neither rich nor structured enough; DC is focused on static objects and operates at an item-level view. CLD may be more important. On a philosophical level, open archives initiatives need to encourage reuse of information objects and provide cost-benefit analyses.

*Carl Lagoze*

We should remember that metadata has uses other than just resource discovery. OAI metadata is largely for dissemination purposes. It may be useful to add work flow processes into the OAI model, so that for example migration of articles can be tracked or information such as annotations or versions can be maintained.

*Andy Powell*

Largely from a technical point of view, as far as the OAI-PMH protocol is concerned, no further development is required. Assignment of rights, such as how to carry Creative Commons licences could be an area to investigate; also how to carry metadata other than DC. There is certainly no need to remove the mandatory status of DC. There is a need for additional mechanisms, (over and above cataloguing guidelines), to cater for QA in metadata provision. Some scepticism was expressed with regard to getting OA eprints adopted in the UK -self-archive may not be the correct model.

### *Muriel Foulonneau*

Largely concerned with how memory organisations take up technology and collaborate (with a French focus). There is a need for preservation metadata to be adopted, as well as data transfer and repository synchronisation. OAI-PMH appears to be under-exploited by memory organisations.

### *Discussion*

To begin with, questions from the floor and discussions centered on the issue of take-up of open archives, in particular institutional repositories. Andy Powell felt that it was difficult to convince academics to self-archive and that there probably would be no significant take-up until there is a requirement to do so by funding bodies. Tim Cole suggested that there is good take-up of archives in some areas such as physics, but studies were required to get a better picture. A member of the audience thought that linking the UK's Research Assessment Exercise to institutional repositories would also encourage take-up in the UK. A significant point made by a member of the audience was that, if institutional repositories are to gain acceptance by academics, there will need to be a change in the scholarly publishing process - a view with which the author agrees.

Further discussions touched the following items: whether publishers would be happy with self-archiving of published material; the cost of running an institutional repository; institutional versus discipline-based repositories (e.g. with joint authorship, where should the article be archived?); and tracking of how harvested metadata is being used.

## **5 OUTCOMES**

As suggested earlier, the number of presentations given at the workshop by projects working within the cultural heritage sector, illustrated that the scope of the OAI PMH is much wider than it might have been imagined at the time it was conceived at the Santa Fe Convention. It is likely that the Protocol will be used significantly beyond the library community, and also beyond the scholarly community (people from television companies have started to turn up to such events - they have large archives of material which they need to manage, and increasingly, to manage as part of a business process which involves making metadata about their holdings available to interested parties.

Organisational issues are also clearly of great significance. In Europe, take up is relatively slow because organisational issues are not yet granted the same importance as they are in the USA for example, where it is understood that the benefits of the OAI approach have to be sold to institutions and users.

Carl Lagoze made it clear that he thinks the battle has been lost for manually created metadata - no-one creates metadata for Google, and Google is, with some notable exceptions, how people find things. Services created on the basis of the OAI Protocol need to compete with that.

## **6 LESSON LEARNED**

### **About the diffusion and use of OAI-PMH:**

The workshop made clear that in the two years lifetime of the project there has been a large adoption of the OAI-PMH. The workshop presented a rich overview of the activities around this protocol, especially those related to the development of services. For the first time, since

the time of the first workshop, the discussion was centred more on the organisational and real technical problems than on the use of Dublin Core. The workshop also highlighted that the diffusion of this protocol will certainly be largely enhanced with the advent of the OAI Static Repository approach that is already been successfully experimented by some organisations.

### **About the Workshop:**

The last workshop, as all the others, had a good attendance. The participants were more experienced than in the other workshops, most of them had implemented or used the OAI-PHM protocol in their organisations. The workshop offered the opportunity to establish synergic links between the different represented organizations, a lot of discussion was held during the breaks. This discussion was also largely stimulated by the poster session that gave very profitable results in terms of dissemination of the OAI related activity. Below, are some of the comments and suggestions received as part of the evaluation questionnaire that was distributed to all the Workshop participants

*“I feel that the aspect that was most valuable was the ability to test my own conceptual framework against that of the international community, well represented in Bath, providing an update on current OAI developments-such as in the excellent report by Muriel Foulonneau on practices for the cultural heritage sector, reviewing current innovative implementations and assessing future directions. In addition, I was surprised and delighted by the range of projects presented and inspired by the well considered presentations made by both Timothy Cole and David Prosser - both of which warrant investigation.”*

*“I liked the fact that the speakers were sharing their experiences with us.”*

*“I found the opportunity to know what are the projects of other institutions and see what problems, solutions and new ideas are developed during implementation time”.*

*“Every presentation was interesting, maybe there were even too many issues together in the same workshop”*

*“I would have liked more time for contacts and discussions, for instance with more time at lunch time and break out sessions, extending the event to 3 days”*

Below are some of the comments on the breakout sessions.

*“The break-out session offered the opportunity to meet people working in similar areas, sharing of perceptions, insight into other peoples projects and their special issues/problems”*

*“Break out sessions could have been more focussed to given subject, but it was interesting to be more wide ranging never the less”*

*“To established in advance the focus areas and circulated on the programme might foster more effective participation in a limited time.”*