OPEN ARCHIVES FORUM: PROJECT DELIVERABLE

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

Deliverable Number:	D4.5
Contractual Date of Delivery:	September 2002
Actual Date of Delivery:	May 2003
Title of Deliverable:	Workshop Report 3
Workpackage contributing to the Deliverable:	WP4
Nature of the Deliverable:	Report
URL:	http://www.oaforum.org/otherfiles/oaf_d45_workshop3.pdf
Author:	Donatella Castelli, Leona Carpenter, Susanne Dobratz, Philip Hunter, Birgit Matthaei
Contact Details:	Donatella Castelli, +39 050 3152902

Abstract	This report documents the Third Open Archives Workshop held in Berlin (Germany) on 27th -29th of March 2003.
Keywords	Open archives, Open Archives Forum, Open Archives Initiative, e-print archives, multimedia archives, intellectual property rights.

Distribution List:	Public
Issue:	Version 1.0
Reference:	32015.DEL4.5-20030530
Total Number of Pages:	20

1 GENERAL OVERVIEW

There are now many different types of media in use, such as video, animation, audio, still images, etc. - both digitally-originated materials, and existing media resources which have been digitised - which digital libraries need to store and manage. The aim of the Berlin workshop was to explore which specific requirements and demands ought to be carefully weighed and considered before a digital media archive is made available via the Internet.

While traditional museums strive to become also digital museums, hosting virtual collections, plenty of new questions covering organisational as well as technical issues arise, such as copyright, collection policies, and open access demands.

In setting up subject gateways it makes sense to connect different media and information resources via standardised and interoperable network gateways that hide their technical specification from the users. Within this workshop we discussed different approaches to network media repositories, libraries, archives and other information resources using both the Open Archives Initiative technical framework for metadata harvesting and other approaches.

The workshop was organised, as the other workshops, in 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. Some of these issues were chosen taking into account the suggestions indicated in the evaluation questionnaires of the previous workshops. In addition, a meeting of the Organisational Issues Working group, that was spontaneously created at the First OA-Forum workshop, took place.

The workshop had, as the previous workshops, a very good attendance, with more than fifty registered participants attending, along with five invited speakers, four tutorialists and six OA-Forum project workers. Most of the registered participants (84%) attended one of the two tutorials. There were representatives from many EU and Nationally funded projects. Fourteen countries were represented: Austria, Germany, Italy, The Netherlands, Norway, United Kingdom, Switzerland, Latvia, Romania, Moldavia, USA, Australia, Ukraine, Albania. Eight of the attendants were funded by Open Society Institute Zug Foundation, a part of the Soros Foundation network. Most of the attendants were project leaders, and technicians, but there were also librarians, archivists, researchers, etc.

As in the other workshops, a number of contacts and ideas for future collaborations were established. In particular, there was a lot attention dedicated to the experiences of institutional archives and to the merging of the OAI approach with automatic source description techniques.

2. Tutorials

Two tutorials on the OAI-PMH protocol, one in German and the other one in English, were held the day before the workshop. These we organised in response to the requests raised at the previous workshops.

The tutorial in English was held by Pete Cliff (UKOLN) and Uwe Muller (Humboldt University). It was a revised version of the tutorial presented at the second Workshop in Lisbon. Changes with respect to the previous editions were stimulated by suggestions and comments returned by the Lisbon participants. In particular, we extended its duration in order to have more time for introducing the different aspects of the OAI-PMH more gradually.

Some of the issues raised during the workshop were:

- need for base-level subject scheme to complement DC base-level metadata format
- need to make clear to decision makers that OAI provides
  - support for interoperability via a metadata sharing solution
  - metadata harvesting (gather from many servers to an aggregated database)
  - building services based on the harvested metadata is not part of OAI-PMH

This tutorial was attended by 60 people, and 14 countries were represented altogether.

The tutorial which was delivered in German was the responsibility of Heinrich Stamerjohanns (University of Oldenburg) and Bruno Klotz-Berendes (University of Dortmund).

The tutorials were attended by 84% of the regular workshop attendees (i.e., not OA-F participants or invited speakers).

As shown by the huge attendance, there is a strong interest in knowing more about the technical and organizational implications of the OAI-PHM protocol. To respond to this request we decided to create in the next months an on-line tutorial available on the Web, which can accessed by everyone, at any time, even after the end of the project. In the meantime the powerpoint slides used by Pete Cliff and Uwe Muller for the tutorial in English is available at:

http://www.oaforum.org/otherfiles/berl_oai-tutorial_e.ppt

2 THE INVITED PRESENTATIONS

This section contains 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/berl_programme.php)**.

The Workshop started with two presentations given by members of the OA-Forum projects. These presentations provided an overview of the current situation in Europe on open archives. The technical overview was collected by interviewing nearly fifty European data and service providers.

Four presentations about European and Nationally funded projects followed over the two days of the Workshop. A fifth presentation was planned but it was cancelled at short notice in cause of illness. These focussed on different issues related to the openness of (multimedia) archives and on the implementation of services that provide advanced functionality on top of them. A presentation on the current status of the Open Archives Initiative (OAI) and on its future plans was also given by a member of the OAI steering committee.

2.1 Technical validation Questionnaire – interim results

by Birgit Matthaei (Humboldt University)

The Open Archives Forum started a first Technical Validation Questionnaire in preparation for the first OA-Forum workshop in Pisa. The objective was to provide an overview on status, experiences and future plans regarding the workshop participants' OAI implementations. At this time exclusively participants of this workshop asked to respond. In Pisa a high interest was raised on the results of this small survey and the OA-Forum project

received feedback indicating that it would be a good idea to collect experiences from a broader spectrum of OAI implementers as well as to learn more about the starting conditions of those planning to implement.
The focus of interest was 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?
Thus in the second questionnaire we added or changed some questions and extended the duration. Beside this, we expanded the target audience for the questionnaire and subdivided the form to account for those projects that have not yet integrated OAI-PMH in addition to those who are experienced implementers.
This second, long-term survey will continue through autumn 2003. The presentation offers interim results of the information the participants gave till now about used software, implementation costs, offered spectrum and interoperability, experiences and expectations in different communities and in different countries.

http://www.oaforum.org/otherfiles/berl_tvq.ppt

Some remarks

The presentation was followed by a discussion on the following issues:

- Although there are many eprint/science implementers, it is remarkable how high a percentage of implementers are libraries.
- Issue of data quality and conversion is likely to be much more time consuming than implementing the OAI-PMH as such.
- The questionnaire surprisingly shows that many people has implemented their own OAI compliant interfaces and they have not used existing tools. Most likely this is because it was filled in by many early implementers (pioneers) that started work before the release of general tools. Now that the OAI-PMH is well-known, the number of implementers who are interested in simple solutions without large development costs is increasing. This will inevitably promote the use of the new tools.

2.2 Overview – European activities on open archives multimedia projects

by Philip Hunter (UKOLN)

This presentation looked at a small selection of European multimedia projects which are using (or are intending to use) the OAI Protocol for Metadata Harvesting.
As a prologue to this brief survey, the presentation examines the problems which arise while looking for multimedia resources on the Web using standard search tools such as Google. The example used is a search for a rights-managed digitisation of William Blake's illustrated edition of his poem 'The Tyger'.
In contrast to the limited search options which are available using a search service which does not use metadata (such as Google), the options available to the researcher with services using interoperable harvested metadata are explored.

The ways in which various multimedia projects are implementing the OAI Protocol are discussed in the context of examples - these range from linguistic archives, television archives, collections of digitized paintings, collections of historical photographs, and projects involved in 3-D modelling.

http://www.oaforum.org/otherfiles/berl_overview.ppt

Some remarks

The presentation raised a discussion on some key issues for multimedia open archives like: IPR, problem of unqualified DC for describing multimedia resources; commercial multimedia controlled access to paid subscribers, etc.

2.3 Prometheus – the distributed digital image archive for research and tuition

by Georg Hohmann (University of Cologne)

As part of its "New Media in education"- program the German Federal Ministry of Education and Research is financing the cooperative university project "prometheus - the distributed digital image archive for research and tuition". The three-year project set to work in April 2001. The partners are the University of Cologne, the Humboldt University of Berlin, the Justus-Liebig-University of Giessen and the University of Applied Sciences of Anhalt at Dessau/ Köthen.
The aim of prometheus is to provide a unified interface to a conceptually very large number of different image data bases that focuses history of art and archaeology. The basic philosophy of the project is, that the individual image databases can have arbitrary different formats, which are unified by a server acting as a technical - and potentially conceptual - "broker". Based on this joined image archive und its media specific potential, prometheus will provide a variety of didactic units to support academic teaching and (e-)learning in the disciplines of Art History, Classical Archaeology, and Design History.
The prometheus central server uses a data model developed over the years for historical research, which build upon the idea of semantic network data bases. In recent years it turned out, that the data structures which can be administered by the system - kleio - are a superset of the data structures which can be expressed by XML. The stage one solution - the contributing data bases send XML dumps to the central server, which maps their structures and semantics into a common system - is currently being replaced by stage two, where instead of dumps being transferred the contributing data bases are mapped dynamically.
If we see the OAI as an attempt to provide integrated access to heterogeneous data sources by a specific protocol discipline required of the contributors, prometheus might be seen as the opposite end, as all the effort in the integration is taken care of by the central server, making no specific requirements of the contributing systems. Providing OAI access to all the contributing databases, simply by supporting the protocol by the server in this way, would be easy. It is not planned for, however, at the moment: Among other reasons, as that would make the handling of existing copyright restrictions all the more difficult.

http://www.prometheus-bildarchiv.de
http://www.oaforum.org/otherfiles/berl_prometheus.ppt

Some remarks

The presentation highlighted that the institutions involved in the project still needed a small and simple data model to achieve interoperability among heterogeneous distributed databases for the simple, lowest common denominator search, although keeping richer data for advanced searching across those records that have more elements.
Structure mapping as well as semantic mapping are also required to provide a central database with a virtually heterogeneous structure.
IPR is a difficult issue for this project.. A workshop attendant from Oldenbourg University let the audience know that his team just lost a copyright case when they were sued for taking the same attitude that it is fine for academic uses.

2.4 The OAI and OAI-PMH , How did we here, and where do go from here?

by Herbert Van de Sompel (Los Alamos National Laboratory)
The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) has its roots in the Santa Fe Convention of the Open Archives Initiative (OAI). The motivation to launch the OAI was to facilitate transformations in the scholarly communication system through specifying technical interoperability between nodes of such a system. That initial quest led the OAI into the realm of defining a generic protocol for Metadata Harvesting that can be used well beyond the initial application domain.
Now that the stable version 2 of the OAI-PMH is in place, the OAI is reflecting on its mission for the years to come, and refocusing on the original scholarly communication domain is high on the list of priorities. The keynote will address the original motivation to launch the OAI, and it will describe the evolution of the OAI work since its launch in 1999. It will also explain the areas of e-print interoperability that the OAI is interested in focusing on in future work, and it will discuss novel uses of the OAI-PMH in areas that go well beyond the typical realm of resource discovery.

http://www.openarchives.org/

Some remarks

The presentation given by Herbert Van de Sompel was highly articulate and touched different aspects.
He started his presentation by explain to the audience the reasons that led him to initiate the work that has brought to the definition of the OAI-PMH. His initial objective was to reposition libraries in the information chain -- getting closer to the creator of scholarly information. He decided that the first step to do in order to achieve this objective was to find a solution to the technical issues, the idea being that providing a solution for these issues would have provided a basis for moving on to solve issues in other areas, such as law, sociology (1999). The lynch-pin was SFX linking as a validation of e-print worth, and insights regarding lack of interoperability. This lead to the launch of the Open Archives Initiative by Ginsparg, Luce, Van de Sompel. Then Carl Lagoze got involved, and they were funded to move beyond eprints to generic metadata sharing for interoperability. Lots has happened since the launch of OAI, but, according to Herbert Van de Sompel, the big change has not happened. His feeling is that libraries (in the States, at least) have lost focus by getting caught up in "learning systems". Consortium buying has been seen by some libraries as a solution to the original problems of journal costs and IPR drain, but he argued this has

inherent dangers. One alterative solution is offered by the Creative Commons framework. [http://www.creativecommons.org/]
The OAI plans for the future are to go back to its e-print roots moving beyond interoperability at the level of discovery. OAI has no current funding, will apply for a grant, the research will be broader than e-prints, but it will be applied to e-prints.

After this introductory talk, Herbert Van de Sompel began to talk about the new planned work around the OAI-PMH.

1. Although this protocol is low-barrier, the barrier can be non-trivial for small collections and where ISP forbids access. A simpler solution has been proposed. This is described in a paper that will be presented at JCDL 2003. [http://lib-www.lanl.gov/~herbertv/papers/jcdl2003-submitted-draft.pdf]
2. OAI-PMH can be used to support harvesting of different information. In particular, it can be employed for access to a DL usage log. This use of the protocol can support the creation of new recommender systems.
3. The OAI-PMH can also be used as a comformant OpenURL Registry

See slides at: **http://www.oaforum.org/otherfiles/berl_desompel.ppt**

A question was raised at the end of the workshop regarding agent and document: ”Using the techniques described it is only possible to know that a user has accessed a document, not what he have done with it or whether the document were successful in his own terms?” Herbert Van de Sompel explained that they have an intermediate step with a button asking the user what they want to do, including download the journal or find more stuff by the same author, which does at least partly answer this. There was a lot of discussion around this.

2.5 Building Digital Multimedia Libraries using MILESS and MyCoRe

by Frank Lützenkirchen (University of Essen)

MyCoRe is an Open Source project for the development of Digital Library and archive solutions (or, put more generally, "Content Repositories" >> CoRe). In the MyCoRe project a group of universities is working on the development of a shared software core for such applications. This core will be adjustable to local requirements and easy to modify. This is expressed by the "My" in MyCoRe, which represents the local adaptability. On the basis of this core which will be available under the open source GNU General Public License, specific local applications will emerge at the participating institutes. The technical base of the system is formed of Java class libraries, XML technology and, besides Open Source database backends, IBM Content Manager and IBM DB2 for large applications.
The Core Functionalities of MyCoRe include the following: Document and Person Metadata, Internal logical Filesystem, Hierarchical Classification System, User and Rights Management, User and Author Editor Interfaces, Distributed Search Function and Interfaces for OAI and Web Services.
The project roots in the MILESS Project of the University of Essen, where a Digital Library application consisting of Java servlets and applets was developed on the basis of the IBM database solution Content Manager. MILESS contains a collection of multimedia teaching and learning materials like animations, audio, video, images, and full text files. It is mainly local material produced in Essen or being used there which is managed with the MILESS

system. Since MILESS was developed to fit the local needs in Essen it was never a primary goal to create a product that would flexibly adjust to the requirements of other locations. Out of the first group of later MILESS appliers to which the University of Jena (Urmel) and the University of Leipzig (Quästur, Bach Digital) belong, the "MILESS Community" emerged (>> "M...y CoRe"). Within this community a detailed idea of the general requirements of Digital Library applications, their common structure and possible differences, was formulated. Out of this dialog grew the decision to develop a shared software core for the different local applications based on the experiences with MILESS. This core is MyCoRe system.
MyCoRe is an Open Source product under the GNU General Public License. The System will be realized on the basis of Java. It will be a serverside application built of Java applications and Java servlets. The import and export format for the describing data will be XML. For now the IBM Content Manager and IBM DB2 will be used as database backend. But the system is generally designed to employ also other backends (especially those developed as Open Source products and applying XML technology) in the future. Adjustability, extensibility, and open interfaces are fundamental design premises. To permit as many local applications by "configuration in place of programming" as possible is the main task.

http://miless.uni-essen.de/
http://www.mycore.de/engl/index.html

Some remarks
This presentation was cancelled due to the illness of the speaker.

2.5 Discovering Good Practice: Metadata and the NINCH Guide

by Ian Anderson (HATII)

The National Initiative for a Networked Cultural Heritage (NINCH) "Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials"[2] is unique in being practice based and expert led. The Humanities Advanced Technology and Information Institute (HATII) at the University of Glasgow was contracted to undertake extensive research on current practice in digitisation on both sides of the Atlantic. Thus the Guide was based on empirical research, and offered good practice from some of the world's best-established digitisation projects. The NINCH Working Group, who conceived and brought the Guide to publication, strengthened the Guide with input from some of the leading experts in the field. This ensured that the Guide was not only timely but could highlight emergent trends, technologies and strategies. The Guide looks to the future as well as reflecting present processes.
The Guide highlights a variety of approaches to metadata amongst the projects analysed and interviewed by the HATII team. This diversity was not only a consequence of the variety of collections - text, image, sound and moving images - but a result of the different institutional contexts in which projects developed, the legacy of analogue cataloguing methods and different technological choices. Methods for representing metadata include: MARC, EAD, DC, TEI, TIFF, XML, and SGML. Thesauri and controlled vocabularies include: LCSC, CDWA, AAT, VRA, TGN, TGM, and ULAN. As this range of acronyms indicates, most projects adopted a hybrid approach to metadata creation, adopting and adapting various standards and technologies according to the type of metadata being created and project requirements.

Although projects were creating metadata to recognised standards and protocols that would enable interoperability, few took a pro-active approach to this. Whilst there was awareness of initiatives such as OAI, METS, CIMI and SMIL projects were adopting a 'wait and see approach'. This cautious approach was not only a result of the immaturity of these initiatives but reflected problems with existing metadata creation, particularly in the descriptive field. Even with institution or project based searching many projects struggled to reconcile accurate descriptions of their digital collections with absent or inadequate thesauri, subject classifications and name control files. As initiatives such as OAI come on stream parallel developments such as the UK Archival Thesaurus may help solve these problems. Nevertheless, the greatest challenge facing multimedia repositories may be populating interoperable metadata frameworks rather than implementing the technology.

http://www.oaforum.org/otherfiles/berl_ninch.ppt

Some remarks

The NINCH Guide has a broader scope than open archives. It deals with an issue of great importance to successful implementation of open archives: how to achieve quality metadata. The NINCH Guide has been designed to be continually updated and extended. It is a community based document, coming out of the digitisation community, authored by a working group. This document: emphasises "good practice", which can be discovered and may exist in several varieties within one area of practice, rather than "best practice" which suggests perhaps a prescription for one way of doing things. First of the NINCH Working Group Principles was "optimise interoperability of materials", second was "enable broadest use".

2.6 ArtWorld

by Paul Child (University of East Anglia, UK)

Projects are temporary. They have a defined beginning and defined end. As project workers, we would like our work to live on after the project has finished. The most common way of ensuring this longevity is to make it interoperable with the widest number of other systems that we can. This can be a daunting task for a relatively short lived organisation and complicating factors such as dealing with multiple media types and the need to reconcile project aims with the interoperability goal can only make the situation worse.
ArtWorld began in 2000 as a three year project, led by the University of East Anglia and is funded by the Joint Information Systems Committee. ArtWorld provides access to primary visual resource materials for the enhancement of learning and teaching in world art studies. It is a consortium project comprising art museums, university departments and research institutes in England centered at the University of East Anglia, Norwich, and the University of Durham. Resources are being built by a team including teachers, students, museum curators from the consortium together with external IT consultants.
In this presentation I will outline the current status of the ArtWorld project and how the project team has approached the difficulties in reconciling multimedia types, interoperability and conflicting project aims.

http://artworld.uea.ac.uk/
http://www.oaforum.org/otherfiles/berl_artworld.ppt

Some remarks

This presentation outlined that there are often conflicting issues in metadata requirements, especially in case of teaching packages. For example, academics have particular pedagogic requirements that may not necessarily be matched by the standards that software specialists use. The NINCH Guide offers a valuable contribution to overcome these conflicts. It is very important that sufficient time and person resources be devoted to a sound planning of how to solve these conflicts before starting the project.

2.3 Resource Selection and Data Fusion in Distributed Multimedia Digital Libraries: The Mind Approach

by Fabio Crestani (University of Strathclyde, UK)

MIND is an IST project funded under the EC Fifh Framework. It is led by the University of Strathclyde (UK), with the University of Florence (Italy), Duisburg (Germany), Sheffield (UK), and Carnegie Mellon (USA) as partners. The project started in January 2001 and is approaching a conclusion.
MIND addresses some of the issues that arise when people have routine access to thousands of heterogeneous and distributed multimedia Digital Libraries. Today, a person must know where to search, how to query different media, and how to combine information from diverse resources. As Digital Libraries continue to proliferate, in a variety of media, and from a variety of sources, these problems of resource selection and data fusion become major obstacles, as solutions based on a centralised repository of metadata will battle with scalability and substantiality.
In this talk I will give a brief overview of the results achieved in the MIND project. I will also outline the differences and similarities between the MIND and the OAI approach to accessing multimedia information in distributed Digital Libraries. Although, very different and starting from almost opposite assumptions, I hope to be able to show that there is strength in a possible combination of the two approaches.

http://www.mind-project.org/

http://www.oaforum.org/otherfiles/berl_mind.pdf

Some remarks

This project has put a lot of emphasis on supporting and understanding within the software functionality the users'/searchers' tasks. It exemplifies an approach to interoperability which is completely different from the OAI one. This approach assumes that the only cooperation provided by the archives is the facility to search them through whatever, usually web-based, proprietary interface. The searches are real-time, with no harvesting to a central database. Here, the architecture is much more complex than with an OAI implementation, but there is absolutely no effort or cost for the individual archive. Also the burden on the services is limited since the system generates, mostly automatically, the information required for supporting them, such as, for example, the query language of each archives and the query transformation rules relating to schemas of individual archives. During the presentation it was outlined that, even if in some cases there are still significant error rates for some automatic generations with schema mapping, both at document and query levels, the experimentation has shown that this approach is viable.

3 THE BREAK-OUT 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 Local and national services in the OAI model: how to implement them?

This session was facilitated by Henk Ellermann (Erasmus University Rotterdam) and Saskia Franken (University of Utrecht) and reported by Diann Rusch-Feja (International University Bremen).

The Dutch program DARE (Digital Academic REpositories) promotes the development of services, as defined, or constrained, by the OAI model by the institutions participating in DARE. The main objective is to show to the academic community as concretely as possible how electronic publishing can be a rewarding activity for them. The development of such services however is not a trivial task. The OAI protocol only offers a first step and much work is needed in setting up acceptable practices in, say, defining metadata and in coming to terms with a large number of organizational and cultural issues. In this break-out session a few representatives of Dutch universities exposed their plans and ideas as a basis for a wider discussion.

The main issues that were discussed during this session were:

- Metadata issues: Unqualified DC format is not sufficient to described all kinds of objects. In order to describe objects the need to give context through something like dc:relation arises; preservation metadata is also likely to be needed. Standardisation for interoperability remains a key issue.

- Organisational issues: There are many of such issues in realising a project like Dare. Authors/creators must submit metadata and objects; the relationship with and the role of parent organisation (e.g. a university, as in the case of DARE) must be clarified by establishing which are the requirements, the polices, the support provided, etc.

- Actions to support DARE and to broad the work of this project so that other communities can exploit the experience done. Some of the planned actions were:

1. consensus-building on metadata needed

2. organisation of a pan-European meeting on institutional repositories

3. Herbert van de Sompel to support communication on new directions and areas of influence for OAI

4. DARE group to report their work for feedback and re-use within the community

3.2 Meeting of the Organizational Issues Working Group

This session was facilitated by Paul Child (University of East Anglia) and reported by Leona Carpenter (UKOLN).
The meeting considered the approach to developing business model guidelines as set out in an email from: Paul Child (Project Manager: ArtWorld, Sainsbury Centre for Visual Arts) to the OI working group posted on the OA-Forum public mailing list (info@oaforum.org) on: 06 March 2003. Paul took as his starting point is the document "Interim review of organisational issues" http://www.oaforum.org/documents/ pages 12 to 15.

The goals of this section might be as follows: To provide a basis for the assessment of the validity of the open archives approach for any given business model. By:
1. assessing the extent to which the open archives approach can be applied to any given business model.
2. indicating how the open archives approach can be applied to each business model.
These were by no means complete or intended to be absolute - any comments would be appreciated. His feeling was that guidelines would stem primarily from 2. above. The issues and questions discussed in the meeting are as follows:
1. Agree a general definition of business models.
The meeting agreed that the Mahadevan definition of business model can be adopted for our purposes:
"A business model is a unique blend of three streams that are critical to the business. These include the value stream for the business partners and the buyers, the revenue stream and the logistical stream. The value stream identifies the value proposition for the buyers, sellers, and the market makers and portals in an Internet context. The revenue stream is a plan for assuring revenue generation for the business. The logistical stream addresses various issues related to the design of the supply chain for the business."
Mahadevan, B. (2000). "Business models for Internet-based e-commerce." *California Management Review*, 42 (4), 55-69. Qoute from p.59.
2. Agree a taxonomy of relevant business models.
The meeting proposed an amalgamation of taxonomies devised by Rappa and Timmers, with addition of licensing and franchise models. This was based on the experience and knowledge of participants within their own and other organisations, including parent organisations and funding bodies. It was noted in particular that mixed models were often the practical response to the complex situations surrounding projects and services which might be based on open archive implementations. Paul Child agreed to draft such a taxonomy and present it via email for discussion to the OI working group and other members of the OA-Forum public mailing list. After discussion, OA-Forum partners would add this new taxonomy to the project deliverable "Final Review of Organisational Issues."
References:
Rappa, M. (2001) *Business models on the Web.* Part of the open courseware project: "Manageing the Digital Enterprise" http://digitalenterprise.org/models/models.html
Timmers, P. (1998). "Business models for electronic markets." In: Gadient, Y.,Schmid, B.F., Selz, D., eds., Dorian: EM -- Electronic Commerce in Europe. *EM -- Electronic Markets*, 8(2), 3-8.
3. Consider how knowledge of business models can inform decisions at a project level.
The meeting considered that this could be done by :
addressing the problem of relationships between business models of different organisations involved in projects and services
addressing the problem of temporary organisations such as projects by documenting that knowledge in ways that are applicable to projects
4. Discover which business models are represented in those participating in the OA-Forum workshops and discussion list.
As noted above, participants said that this tends to be mixed models -- an example is one of the participants, where a combination of licensing, information services, and brokerage models provide a combination that is proving *commercially* viable because of the value to clients is thus suffiently high to result in an adequate revenue stream. It would be useful to gather experience from more organisations than the few represented at this fairly small meeting.

3.3 Metadata for Multimedia Objects

This session was facilitated by Heike Neuroth (SUB Gottingen) and reported by Susanne Dobratz (Humboldt University.

The session used the following text as a framework for the discussion:

Metadata definition:

- „data about data“, information about information“, etc.
- metadata is associated with an „object“ (information resource), object could be digital or not
- but here: restricted to metadata that describe digital objects

= data about digital objects (?)

Purposes:

- resource description (content metadata) (e.g. E-Learning)
- resource discovery (agreement of core set?)
- identification, location
- usage information (IPRs, DRM, ...)
- management purposes (administrative metadata)
- long-term archiving (preservation metadata)
- technical information (e.g. format, size etc. of image file)

= human and machine-understandable (with encoding formats like XML, RDF, METS, etc.) data

Metadata Schema:

- Dublin Core and its Application Profiles
- MODS

MPEG Standard (http://mpeg.telecomitalialab.com/):

Coding of moving pictures and audio***:***
Moving Picture Experts Group (MPEG) is a working group of ISO/IEC in charge of the development of
standards for coded representation of digital audio and video. Established in 1988, the group has produced:
- MPEG-1: standard on which such products as Video CD and MP3 are based
- MPEG-2: standard on which such products as Digital Television set top boxes and DVD are based
- MPEG-4: standard for multimedia for the fixed and mobile web
- ***MPEG-7: standard for description and search of audio and visual content***

Work on the new standard MPEG-21 "Multimedia Framework" has started in June 2000: describes a multimedia framework and sets out a vision for the future of an environment where delivery and use of all content types by different categories of users in multiple application domains will be possible. MPEG-21 assumes that there are Users (anybody in the value network) and Digital Items (assembly of content) on which Users execute Actions that generate other Digital Items that can become object of Transactions. In order to make this possible a number of technologies are needed that fall under the following categories:

- Digital Item Declaration
- Digital Item Identification and Description
- Intellectual Property Management and Protection
- Terminals and Networks
- Digital Item Management and Usage
- Digital Item Representation
- Event Reporting

Open Issues/Questions:

- General: What do we mean by multimedia metadata? Where do we need multimedia metadata (e.g. e-Learning)? Is OAI Dublin Core sufficient for resource discovery/description for multimedia objects (e.g. movies in general vs. single sceme of a movie)? Integration of different types single (complex) objects in "one object" (e.g. movie, ppt presentation, text, oral talk, etc. as one combined digital object)?
- Interoperability: cross-domain interoperability, concept of Application Profile and Namespace,
Dublin Core as interchange/exchange format?
- Standardization: controlled vocabulary systems, authority files, etc.?
- Multilingual Metadata/Internationalization: other languages than English, language of the primary audience, cross-lingual interoperability, etc.?
- Digital Object: What is a digital object (web page, digital image, multimedia collection, etc.)?
- Granularity: Parts of objects/collection, single item, etc.?
- Registry: Dublin Core, CORES, MetaForm?
- Tools: for metadata creation, are they helpful?

There were 20 Participants in this session, the majority of whom were not specialists in this area. Many have projects already, and have to solve pressing tasks, such as the integration of several metadata structures into one in order to provide search, and the need for metadata that allows the description of different types of resources. Some time was spent on the definition of the term "Multimedia Object" Is this best thought of as an "Information object"?

Kinds of metadata were considered, and the user orientation. The need for metadata to facilitate brokerage between different formats of media objects was also considered. Standards - MPEG 21 was discussed broadly, as something which ought to be looked at in detail. The question of whether or not a multimedia object is best thought of as a combination of two or more different media formats (ie, as types of information presentation).

There was also discussion of the way metadata is splitting into different kinds for different purposes. There is no single metadata schema for multimedia, but the possibility of a combination of metadata schemas for different types: audio, picture etc. This leads (perhaps) to application profiles for metadata Metadata, it was agreed, should support retrieval, and should be fit to users needs.

3.4 Technical Copyrights Enforcement and Open Access

This session was facilitated by Volker Grassmuck (Humboldt University) and reported by Diann Rusch-Feja (International University Bremen).

Most of the discussion was focussed on Digital Rights Management (DRM) and its legal protection. This establishes a new global knowledge regime. DRM is about turning the Internet into an e-commerce environment, trying to make information behave like material goods. DRM would lead us to a world filled with "digital barbed wire" if we make give-away literature amenable to copyright management through the technical solutions of DRM. The questions of how DRM affects open access and what can be done to ensure the freedom of scientific communications where raised. Many of the participants believed that organisations/institutions should hold copyright rather than the authors; however, many also thought that some rights should be retained by individuals.

http://www.oaforum.org/otherfiles/berl_bs_pr_copyright.pdf
http://www.oaforum.org/otherfiles/berl_bs_rb_copyright.ppt

3.5 Requirements and lessons from the open archive service providers

This session was facilitated by Donatella Castelli (ISTI-CNR) and reported by Pete Cliff (UKOLN)

The attendance to this session was quite high. This confirmed the strong interest around the development of OAI-PMH compliant services evident at the workshop. Very few of the attendees had implemented a service, and most others were "thinking about it". Most of the planned services are discovery services, i.e. search engines and the like. This means that there is not still much consolidated experience in the development of these services. The discussion was centred around the question “What do services need?”. A number of suggestions were given by the session facilitator to stimulate discussion.

1. A Registry of Data Providers?

There was much discussion and both sides of the argument were presented. Some people believed that a central repository was necessary to facilitate the discovery of appropriate data and service providers .If a central repository was created, the need for a standard schema for describing data-providers (DPs) repository was recognized. It was also suggested that the central repository could be an OAI-compliant repository in itself. Such a repository would certainly need both human and machine interfaces. Alternately, it might be possible to automatically generate summaries of DPs content by analysing the metadata it contains. Ultimately there might be a need for two registries: Service Providers and Data Providers.

Some argued against the creation of a single registry. They claimed that i) there are so few archives at the moment that a service provider is already aware of which data providers it should use. Alternately, the community which the SP serves would have a sufficient knowledge of appropriate DPs that no automatic discovery was required; ii) if a SP builds a "high-profile" service then the DPs will make themselves known to the SP in order to get their metadata into the service; In addition a central repository may be undesirable because it would (if the OAI folk are right) grow to huge proportions very quickly and require some management. It was hard to identify an organization that would take on the responsibility of that maintenance . A DNS-like system was suggested as a solution.

The OAI-PMH provides the "friends" mechanism which allow DPs to inform SPs of similar repositories, but this might not be ideal as small closed groups of friends may appear. It should be noted that the OAI believe a central repository would be a problem as it would soon get very large and OAI have no interest in maintaining such a registry.

The next thing a SP needs is:

2) Appropriate Quality Metadata (or Good/High Quality)

Those who had seen the content of DPs reported that the quality of the metadata in DPs varies:
i) Some DPs do not always provide valid XML for all records
ii) Even seemingly "full" records can cause problems because the "content" is not standardized – e.g. language could be a three letter code, a two letter code, a word (in any language), etc. (To address this, people felt that a suite of tools to do things like normalize metadata i.e. mapping subject terms, language codes, etc., would be useful. Such tools could include name, language, date, subject mapping /normalization; autoclassification; OpCit; etc.)

Some metadata guidelines from the OAI might improve the metadata quality situation, but these would be difficult to enforce. The OAI would be unlikely to mandate "high quality" as part of the PMH because this "raises the barrier" of adoption. OAI-PMH is meant to be a low-barrier solution, and strict metadata guidelines would run counter to the OAI philosophy.

Another solution is for the SPs to mandate a "minimum standard" for inclusion in a service. Again, if the service is "high profile" enough, such as Nature, then the DPs will see the value in improving their metadata and work towards getting into such services.

Final discussion suggested that it was too early to assess the OAI-PMH's ability to support complex services as these services had not been build. Current SP implementations are probably quite simple discovery services. There needs to be some consideration by SPs about what services they could and should be providing.

4 OUTCOMES AND ACTIONS

Thanks to the hospitality of Humboldt University, the support of Humbolt staff, EU IST project funding, and most of all the hard work of workshop participants (including invited speakers) the workshop objectives of sharing knowledge and experience and initiating joint work were achieved. The workshop saw a number of emerging themes and identified some future actions. It is interesting to note the extent of common concerns among participants across the range of different kinds of projects and services, organisations, repositories, and subject and data or content types that were represented.
Balancing complexity with simplicity was one recurring theme. Herbert van de Sompel in his keynote presentation expressed the desire to see future OAI work to extend OAI in ways the would ease implementation for small organisations, reducing the requirement for technical skills within an implementing organisation. Many participants commented on the relative complexity of providing services based on harvested metadata, in comparison with the relative simplicity of becoming a data provider. It appeared that the ensuring the quality and standardisation of metadata could make far more work for organisations than any other aspect of exposing that metadata for harvesting.

Many projects and services, themselves often evolving from projects, experience difficulty in terms of sustainability at the end of initial start-up funding periods. Business models for long-term sustainability could prove elusive for some. Case studies showing what has worked in practice would be welcome, and OA-Forum should attempt to identify a number of these for future dissemination.
Over and over again, there was talk of the importance of open standards for interoperability. The usefulness of DC for base-level interoperability of metadata was acknowledged, while it was recognised that richer metadata should also be provided where available, and especially within specialist communities. OAI-PMH was accepted by many as a core interoperability standard. It is notable that respondents to the OA-Forum technical evaluation questionnaire, according to Birgit Mattheiu's presentation of results, judged the importance/advantages of OAI to: range from "to provide access to all of human knowledge" to "nothing other than political expediency".
In the NINCH Guide presentation, Ian Anderson spoke of how, once again, the problem arose of what subject scheme to use, and it was pointed out that the terms used in big standard schemes may provide interoperability but not provide the terms to express local cultures. Many spoke of the need for interoperable subject schemes, and also of the need for a single, simple scheme to provide base-level subject interoperability such as is provided by unqualified Dublin Core for item descriptions.
The need for guidance on good practice, perhaps especially in the area of metadata creation, was another recurring theme. The NINCH Guide has already been helpful for some participants, including ArtWorld. Participants can contribute to the drawing of guidelines in some areas of shared concern through the work of the Organisational Issues working group, and this will continue to be aired on the OA-Forum public mailing list, which all were encouraged to join and use. In addition to using the info@oaforum.org list to discuss organisational issues relating to the open archives approach, participants were encouraged to use the OAI's own implementers' mailing list for discussing technical issues. They were also asked to register information about their repositories, services and projects in the OA-Forum information resources. The 4th OA-Forum workshop was announced, with the title "In Practice, Best Practice: the future of Open Archives", to be held in Bath on 4-5 September.

5 LESSONS LEARNED

One of the most interesting things to emerge from the workshop was the fact that some multimedia projects using the OAI PMH (still relatively few in number) are using the protocol for internal harvesting and administration purposes, without necessarily being interested in either other projects harvesting their metadata, or harvesting anyone else's metadata. These multimedia projects are in effect providing their own custom services (for themselves), using their own data resources. Using the OAI PMH leaves open the possibility of other uses and other services being developed in the future, which might involve the importing or exporting of both metadata and multimedia objects, and collaboration with other archives and services. In which case it appears to be the interoperability and flexibility afforded by adoption of the OAI PMH which is one of the main attractions of the protocol.

About the diffusion and use of OAI-PMH:

There are still very few multimedia archives that have moved towards the open archive model, but the protocol is an attractive proposition for them.

About the Workshop:

This workshop confirmed many of the impressions we have had for the other workshops.

The background of the participants was very different, as was their experience with open archives. They had different expectations about the workshop and the tutorials. Those who were just considering opening their archives and/or were at the beginning of the conversion process were looking for practical guidance, simple software solutions, and experiences from the practice of others with similar goals. Those who were working at technical and organisational solutions (eg. DARE) were looking for tips and discussions with more experienced communities. Finally, experienced implementers were willing to discuss special solutions, finding common grounds, and to offer their experiences to the community. This heterogeneity in the attendees risks creating some problems for the arrangement and design of the presentations and needs to be carefully managed (not everyone necessarily wants the kind of technical detail available in the tutorials, for example). A similar heterogeneity was also noticed among the tutorial participants. This aspect must be taken into account in the preparation of the on-line tutorial.

Also in this workshop there were people that had attended one or both the previous workshops. It was interesting to notice that the long-term influence of the Open Archives Forum has now become concretely recognizable. One of the attendees commented on his participation in the workshop saying: "In Pisa my participation reason was noncommittal curiosity. In the meantime I became an OAI implementer. Therefore, in Berlin I could extract concrete benefit from the Workshop.".