OPEN ARCHIVES FORUM: PROJECT DELIVERABLE D3.2

Project Number:	IST-2001-32015
Project Title:	Open Archives Forum
Deliverable Type:	Public
Deliverable Number:	D3.2
Contractual Date of Delivery:	Month 24 (September 2003)
Actual Date of Delivery:	September 2003
Title of Deliverable:	Final review of organisational issues
Workpackage contributing to the Deliverable:	WP3
Nature of the Deliverable:	Report
URL:	http://www.oaforum.org/resources/deliverables/
Author:	Leona Carpenter, with Donatella Castelli, Michael Day, Rachel Heery, Philip Hunter; Dennis Nicholson, William Nixon, Paul Child, Elizabeth Gadd, and a contribution on IPR by Mark Bide (Rightscom Limited)
Contact Details:	Address: UKOLN, University of Bath, Bath BA2 7AY, UK email: r.heery@ukoln.ac.uk, phone: +44 (0)1225 386724, fax: +44 (0)1225 386724, URL
Abstract	This review explores the validity of the open archive approach through a close examination of organisational issues. It describes the approach and the Open Archives Initiative (OAI), and supports European organisations in benefiting from the added value of open archive technology by exploring the incentives for participation in an exchange-based model for providing information, archiving and metadata sharing services. A summary of Intellectual Property Right (IPR) issues related to open archives is included. Other topics covered are governance of the OAI, business models, content management, metadata issues and quality assurance.
Keywords	Open Archives Initiative, OAI, open archives approach, business models, Intellectual Property Rights, IPR, quality assurance, participation, incentives
Distribution List:	Public (via Web site)
Issue:	0.1 (Draft of final version)
Reference:	32015-DEL3.1-20030930
Total Number of Pages:	29

TABLE OF CONTENTS

PART I Title Page

OPEN ARCHIVES FORUM: PROJECT DELIVERABLE D3.2 1

PART II - MANAGEMENT OVERVIEW 4

Document Control 4

Executive Summary 4

Scope Statement 5

PART III - DELIVERABLE CONTENT 6

Introduction 6

Glossary 6

1 The open archives approach 7
- 1.1 What is the open archives approach? 7
- 1.2 Communities that do or could benefit from the open archives approach 10
  - 1.2.1 Potential benefits and incentives for participation 10
  - 1.2.2 The scholarly community and research libraries 10
  - 1.2.3 The astrophysical community 11
  - 1.2.4 The archival organisation community 12

2 OAI and the open archive approach in Europe 12
- 2.1 Governance of the Open Archives Initiative (OAI) 12
- 2.2 Recommendations [to be added in final version] 13

3 Business models for open archives 13
- 3.1 Growth scenarios 13
- 3.2 Possible/emerging business models 14
  - 3.2.1 What is a business model, especially in the context of the Internet? 14
  - 3.2.2 Which might be applicable to open archives? 15
- 3.3 Risk analysis [to be added in final version] 16
- 3.4 Recommendations for sustainable business models. 16
  - 3.4.1 Why are business models important? 16
  - 3.4.2 Business models and Public Organisations 17
  - 3.4.3 Business models and Publicly Funded Projects. 17
  - 3.4.4 Business models and the OA movement. 18
  - 3.4.5 Conclusions 18

4 Intellectual Property Right (IPR) issues for open archives 18
- 4.1 Overview 18
- 4.2 Defining IPRs 19
- 4.3 An infrastructure to support IPR on the network 19
- 4.4 Metadata and other types of content 19
- 4.5 Stakeholder attitudes 19
- 4.6 Implications for Open Archive Services 20

5 Quality assurance for open archives [draft for final version] 21
- 5.1 Quality assurance for open archives 21
- 5.2 Dimensions of quality 21
- 5.3 Tools. 22
- 5.4 Who is responsible 22

6 Metadata [Draft outline for final version] 23
- 6.1 Metadata Standards 23

7 Subject Interoperability 25

7.1 Subject Interoperability: Guidelines for Best Practice: An Illustrative Draft [Draft of Section for final version of deliverable] ... 25
7.2 Overview ... 25
7.3 Draft Guidelines ... 25
7.4 Illustrations ... 25
7.5 RENARDUS ... 25
7.6 MACS (Multilingual Access to Subjects) ... 26
7.7 HILT (High Level Thesaurus Project) Phase II ... 26
7.8 SPEIR (Scottish Portals for Education, Information, and Research) ... 26

8 Guidelines for Best Practice Managing Institutionally-created Electronic Resources using a Content Management System [Draft for final version] ... 26

9 The Importance of Organisational Issues in the take-up of the OAI PMH ... 28
9.1 The downside of the simplicity of the Protocol for Metadata Harvesting. ... 28
9.2 The Concept of the 'Deposited' resource. ... 28
9.3 The reluctance to see the Protocol as a part of a publishing process. ... 28
9.4 IPR, Knowledge Management, and Content Management beyond the world of Open Archives. ... 28
9.5 The use of the OAI PMH in the context of Cultural Heritage Institutions. ... 28
9.6 Metadata and Quality control. ... 28
9.7 In search of the appropriate business models ... 28
9.8 Organisational Issues and the Future of the PMH ... 28

PART IV ... 28

10 References ... 28

PART II - MANAGEMENT OVERVIEW

DOCUMENT CONTROL

Issue	*Date of Issue*	*Comments*
0.1	30 August 2003	Draft for review by partners and at Peer Review meeting (this version)
0.2	23 September 2003	Final draft for approval by partners
1.0	30 September 2003	Final version

EXECUTIVE SUMMARY

The essence of the open archives approach is to enable access to Web-accessible material through interoperable repositories for metadata sharing, publishing and archiving. It arose out of the e-print community, where a growing need for a low-barrier interoperability solution to access across fairly heterogeneous repositories lead to the establishment of the Open Archives Initiative (OAI)[1]. The OAI develops and promotes a low-barrier interoperability framework and associated standards, originally to enhance access to e-print archives, but now taking into account access to other digital materials. Many communities are beginning to or potentially could benefit from the open archives approach. The Internet and the growing mass of material in digital format have broadened the potential clientele of many repositories of information. Material can be accessed more widely and also exploited for purposes different from those that originally motivated the creation of the repositories. Moreover, the possibility of accessing multiple repositories enables the construction of new kinds of services that can better serve the needs of the users. An additional incentive is the potential for cost-saving inherent in new models of the scholarly communication process that could be supported with this approach. The situation in the scholarly community and research libraries, the astrophysical community and the archival organisation community provide examples of the differences in incentives, interest and uptake across communities.

As an organisation, the OAI is made up of an Executive for management, and Steering and Technical Committees for policy direction and evaluation. The OAI is funded by the Digital Library Federation (DLF)[2], the Coalition for Networked Information (CNI)[3], and the National Science Foundation (NSF)[4] While the Executive and the funders are USA-based, the success of the OAI is firmly grounded in the participation of a community of people from around the world, particularly Europe as well as North America. Now that there is a well-developed and stable second version of the protocol the need to keep control in the hands of a very small number of people who can take independent and speedy decisions may be less important when weighed against the perception of stability and authority conferred by control through a standards body such as ISO. Hence, it could be in the interest of European implementers of OAI-PMH to advocate the hand-off of the protocol to an international standards body. Recommendations for action by European organisations are at section 2.2 of this document [for final version]

The OAI-PMH protocol could become part of the infrastructure of the Web, as taken-for-granted as the HTTP protocol now is, if a combination of it's relative simplicity and proven success by early implementers in a service context leads to wide-spread uptake by research organisations, publishers, and "memory organisations". In line with recent discussions of the nature of business models in the context of the Internet, a business model

[1] Open Archives Initiative (OAI) <www.oai.org>

[2] Digital Library Federation (DLF) http://www.diglib.org

[3] Coalition for Networked Information (CNI) http://www.cni.org

[4] National Science Foundation (NSF) http://www.nsf.gov/

can be seen as the method of sustaining an activity or organisation through providing value and attracting the resources that make it possible to continue to provide value to some constituency. Perhaps the most applicable business model for open archives is the "virtual community" model (services that gain support from members contributing effort, content or money).

This review raises the Intellectual Property Right (IPR) issues for open archives which are examined in detail in an expert report commissioned by the OA-Forum. The expert report defines and explores these issues, as they relate to both economic and moral rights of exploitation of the works created by authors. Not the least of these issues is the difficulty of managing these rights in a global environment, in the face of substantially different attitudes to IPR in different jurisdictions and cultures. It explores what constitutes "publication" in the online environment, and how IPR applies to content that is freely available on the Internet. The relationship between Open Archives and Intellectual Property is complex, not least because of the complexity of definition of what "Open Archives" are.

The importance of quality assurance within the context of emerging web based information services is becoming more widely recognised. Services often emerge from projects, and so there needs to be consideration of quality assurance from the start of projects. Projects need to consider quality criteria, to establish an evaluation process, and to ensure usability of system solutions. The dimensions of quality for open archives are integrity of data, quality of service provision to end users, quality of data, and sustainability. The issue of who is responsible for quality assurance within the open archives approach is complex. It may be obvious that service providers are responsible for the quality of services they provide, but how this is to be achieved is less obvious, especially in an environment where metadata creation and maintenance is so distributed. Quality policies and best-practice guidelines yet to be developed have an important place in assuring quality.

SCOPE STATEMENT

This is the second of the two reports that are the contracted deliverables of Open Archives Forum Workpackage 3 'Organisational Validation'. These reports review the organisational issues for open archives, in order to increase the understanding of incentives for participation in open archives, the awareness of Intellectual Property Right (IPR) and quality issues for open archives, and make recommendations for sustainable business models for open archives in the European context. This report benefits from the discussion and presentations of the OA-Forum workshops, post-workshop discussions on the OA-Forum "info" mailing list, and the Workpackage 4 expert reports. This report builds on the first interim report, by adding more detail on the topics, plus new sections, including a section on risk management, and another on quality issues, and further recommendations in certain areas.

PART III - DELIVERABLE CONTENT

INTRODUCTION

This review is designed to explore the validity of the open archive approach through a close examination of organisational issues. It describes the open archives approach and the Open Archives Initiative (OAI) from an organisational, rather than technical point of view. It supports European organisations in benefiting from the added value of open archive technology by exploring the incentives for participation in an exchange-based model for providing information, archiving and metadata sharing services. A summary of Intellectual Property Right (IPR) issues related to open archives included here is based on an externally commissioned, more detailed expert review of the subject. Other topics covered are based on the expertise and investigations of the project partners and on input from the Open Archives Forum workshops and the European Organisational Issues working group that was formed at the first workshop in Pisa (2002).

GLOSSARY

Archive
"Archive" in the Open Archive Initiative sense is synonymous with a repository of scholarly papers (reflecting the origins of the OAI); and, more broadly, a repository of stored information. Outside the OAI, archivists have stricter definitions of what an archive is, these definitions include notions of long-term preservation, institutional policy, and document appraisal. See also "Open Archive" in this glossary.

Data Provider
A Data Provider's web server supports the OAI Protocol as a means of exposing metadata about its content (see also Service Provider)

DC
Dublin Core.

Dublin Core
A standard metadata format for Web documents (with 15 core fields).

Interoperability
Interoperability issues are those revolving around the implementation of common standards and the machine readability of data (resulting in data interoperability)

Metadata
In essence, 'data about data'. Descriptive information about resources.

Metadata harvesting
The collection of metadata into a single repository (see repository) from distributed data providers.

OAI
Open Archives Initiative.

OAI-PMH
Open Archives Initiative Protocol for Metadata Harvesting.

Open archive

An 'open archive' is one for which metadata is publicly available (i.e., 'exposed') to researchers via a service provider. The 'open' part of this technical term refers to the architectural aspect of the open archives idea - the definition and promotion of machine interfaces that facilitate the availability of content from a variety of providers. 'Open' does not mean 'free' or available for unlimited use. 'Archive' reflects the origins of the OAI - in the E-prints community the term 'archive generally means 'a repository of scholarly papers' - but the OAI uses the term in a broader sense as a repository for stored information.

Open Archives Initiative (OAI)

A particular initiative to make information about electronic resources widely available, using the OAI Protocol for Metadata Harvesting (OAI-PMH) to provide a low-barrier interoperability solution. The OAI website is at http://www.openarchives.org/

Open Archives Protocol for Metadata Harvesting

The protocol defines a mechanism for harvesting records containing metadata from repositories. The link between this metadata and the related content is not defined by the OAI protocol.

Service Provider

A service provider makes harvested metadata available. The service provider issues OAI protocol requests to data providers, and the resulting metadata repository is the basis for building value-added services (see also Data Provider).

Z39.50

A network protocol that allows searching of remote, heterogeneous databases and retrieval of data.

1 THE OPEN ARCHIVES APPROACH

1.1 What is the open archives approach?

The essence of the open archives approach is to enable access to Web-accessible material through interoperable repositories for metadata sharing, publishing and archiving. There are now many repositories of digital materials of many kinds that either are, or potentially could be, assessable over the Web. The open archives approach arose out of the e-print community, where a growing need for a low-barrier interoperability solution to access across fairly heterogeneous repositories lead to the establishment of the Open Archives Initiative (OAI)[5]. The OAI develops and promotes a low-barrier interoperability framework and associated standards, originally to enhance access to e-print archives, but now taking into account access to other digital materials. These efforts were initially aimed at enabling the author to make resources (and associated metadata) available by means of the archive direct to the user. Within the e-print community, establishing inter-working, interoperable e-print archives showed how services might be layered on top of individual archives.

The potential value to other communities of this approach was soon perceived, and now there is a great deal of experimentation with this approach across different types of organisations including, for example, libraries and publishers. At the heart of the approach is the attempt to disseminate the contents of repositories efficiently through making repository metadata (descriptions of resources in the repositories) available for harvesting and use in (perhaps 3rd-party) services. To be practical, the open archives approach must be supported by low-barrier interoperability solutions that are widely taken up. The key solution arrived at through the work of the OAI has been the development of a harvesting protocol, now known as the Open Archives Initiative Protocol for Metadata Harvesting, or OAI-PMH. It is important to note that metadata, and not ordinarily the content that it describes, is harvested. It is also important to remember that within the OAI 'open archives' has a particular meaning. 'Archives' refers to data repositories; 'open' refers to the availability for harvesting of collections of metadata relating to data repositories. 'Open' may also refer to the standards development process.

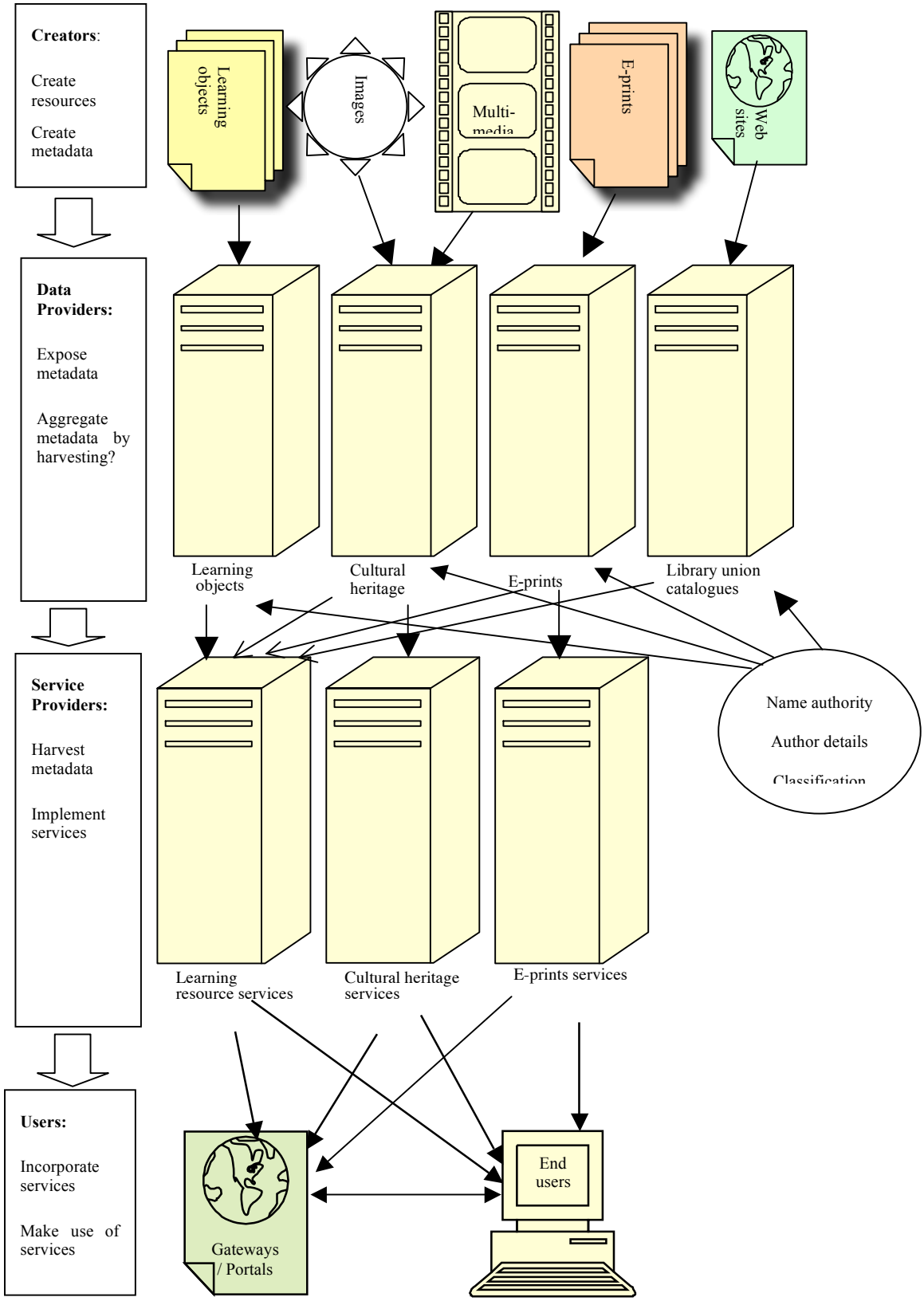
[5] Open Archives Initiative (OAI) <www.oai.org>

The OAI has promised much on the basis of this model, a new pattern for scholarly communication being the most publicised. Perhaps more achievable are the goals of surfacing 'hidden resources' and low cost interoperability. Although the OAI-PMH is technically very simple, building coherent services that meet user requirements remains complex. The OAI-PMH gives a simple technical option for data providers to make their metadata available to services. Although this gives data providers a low-cost entry to interoperability, in the OAI model there are a number of actors involved in making the components of the overall framework. Diagram 1 illustrates some of the relationships between these actors.

Creators are those who create either resources or metadata describing resources, or both. The resources are ordinarily (but not necessarily) such as may be held in a digital repository, such as e-prints, images, learning objects, multimedia, Web sites. Resources may be the products of original authors (for example, a research paper), or of intermediary organisations (for example, the results of a digitisation project in a museum). Creators may also act as data providers. Data providers expose metadata for harvesting; the metadata exposed may itself have been aggregated by harvesting from other data providers. Examples of types of data providers include repositories of e-prints, learning objects, cultural heritage resources, and even union catalogues. All of these depend for their usefulness on adequate processes of subject classification and name authority control relating to author details. Service providers harvest metadata from the data providers and implement services based on this metadata. Examples of service providers include learning resource services, cultural heritage services, and e-print services. Service providers may harvest metadata from different types of data providers; for example, a provider of a learning resource service may harvest metadata relating to learning objects, cultural heritage items, e-prints, and from library union catalogues. Users may be end users of the service providers' services, or may be organisations providing, for example, subject-based gateways or institutional portals.

From this rather simplified illustration, it can be seen how important it is for there to be good communication among the various parties involved in building services based on the open archives approach – and also, perhaps, how potentially difficult this might be.

Diagram 1. The open archives approach to interoperability and metadata sharing



1.2 Communities that do or could benefit from the open archives approach

1.2.1 Potential benefits and incentives for participation

The interest in the open archives approach has been growing since 1999 when the OAI started discussing this approach within the e-print community. There are numerous potential benefits to end-users of the open archives approach and incentives to organisations (and perhaps individuals) of participation in open archives, so this interest can be expected to continue to grow.

In order to support interoperability among open archives OAI has proposed a technical solution based on the use of a harvesting protocol. At the present this is the most widely known and discussed solution for supporting interoperability among repositories of electronic documentation. However, the e-print community is not the only community that is interested in open archives, nor is the OAI solution the only one proposed for sharing information.

With the advent of the Internet, many organizations and communities decided to produce digital material or to digitise existing material. The digital format and the possibility of remotely accessing it have broadened the potential clientele of many of these repositories of information. The data can be accessed more widely and they can be exploited for purposes that are different from those that originally motivated the creation of the repositories. Moreover, the possibility of accessing multiple repositories enables the construction of new kinds of services that can better serve the needs of the users.

Further consideration should be given to:

- the potential impact of OAI from the viewpoint of the end user and the benefits to the learner, researcher and citizen arising from European organisations participating in the OAI approach, for example optimised cross-domain searching, access to 'hidden' documents, changes to the publishing paradigm;
- the benefits to the end user of surfacing European documents by means of the OAI approach, and which content types would be of particular value to the end-user;
- the added value of different user interfaces to the services arising e.g. OAI portal, mix of OAI and other services, integration of document delivery; and
- the incentive to organisations provided by a low-cost, low-barrier interoperability solution.

1.2.2 The scholarly community and research libraries

The scholarly communities are currently among the most active on the open archives approach. Many of them see the OAI-PMH as a valid and low cost of adoption solution for building a new scholarly communication model. Many institutional and discipline-oriented archives are now OAI-compliant and their number is estimated to grow very fast in the near future. A big contribution to this expansion comes from the University libraries. Many of these libraries see in the OAI approach the possibility of making rapidly available worldwide the University publications. In many cases they also see it as an instrument for supporting an alternative form of scientific information distribution that is much cheaper than the current one. Many University libraries are experiencing budget problems due to the ever growing number of publications available, and so the cost is a key issue for them.

Many of the University Libraries, especially those of the Mediterranean and Eastern European countries, do not yet have network-accessible digital repositories. They are now in a phase in which they are looking for a software solution that would allow them to submit and disseminate their documents and make these accessible worldwide. As the OAI protocol promises to satisfy their requirements, there is a lot of attention to the few tools that are now available for building OAI-compliant repositories, especially from EPrints.org[6].

[6] Eprints.org <http://www.eprints.org/>

The OAI approach is now also used for supporting harvesting within the National SMETE[7] Digital Library (NSDL) [Arms2002]. The NSDL is a digital library for education in science, mathematics, engineering and technology, funded by the National Science Foundation (NSF) Division of Undergraduate Education, built integrating existing collections. Interoperability among existing collections is a central issue in the construction of NSDL. The potential collections have a wide variety of data types, metadata standards, protocols, authentication schemes, and business models. The OAI-PMH is one of the mechanisms used to cope with this diversity. It is expected that many organisations will join NSDL through this mechanism due to the low cost of participation vs. the level of functionality achieved.

Many any other groups have very similar problems to those faced by the e-print community, including libraries, museum, commercial journal publishers, and community of scholars who need to share distributed data resources [Lynch2001]. These groups have started discussing the adoption of the OAI approach as a mean for supporting interoperability within their community. The OAI-PMH can be used not only as a protocol for intra-community exchange but also as a mechanism for supporting inter-community interoperability.

As already outlined above, however, the OAI-MPH is not the only possible approach to open archives. Many other approaches, more or less general and successful, are being used. Z39.50 is probably the best known. It has been proposed initially within the library community but then its application has been extended to other areas like museums, geographic systems, etc. Z39.50 is a well-defined and sophisticated protocol that allows a client to search a remote information server across a network. It can be used as a tool to build federated search systems. This protocol has been successful only in restricted areas, because building Z39.50 clients has turned out to be very difficult and expensive. The debate between the fans of Z39.50 and those that promote OAI is very lively. It is important to notice, however, that the two protocols have been meant for different purposes, although they can both been used as building blocks in the construction of search services.

1.2.3 The astrophysical community

Other special interest communities have developed their protocols and standards for supporting the interoperability among their archives. A notable case in point is the astrophysical community. This is a small but strong community. It is a good example of a well-organised community. In order to enhance exploitation of ground and space observations astrophysical have launched the idea of a Virtual Observatory built upon heterogeneous sources of data, and services. Several RTD projects have been accepted recently for the construction of this Virtual Observatory. The problem of interoperability is one of the main issues for the construction of this observatory. Different initiatives have been dedicated to this issue. A Working group, called OPTICON[8] (), has been set up for studying cost effective tools and standards for improving access to and data exchange from data archives and information services. An important pre-requisite established by this working group for the solution to be found was, as for the OAI proposal, to keep to a minimum the additional workload on data providers.

The National Virtual Observatory (NVO)[9] is a project specifically dedicated to interoperability issues. It has established a simple protocol to be implemented by a curator of an astronomical source catalogue for publishing it in such a way that a simple cone search can be done. The data remains in the control of the curator, served through a Web server to the worlds, but the query and response formats follow a defined protocol. This protocol, even if specifically tailored for the astrophysical requests, shares many similarities with the OAI protocol. In particular, the requests are formatted as http GET requests and the responses are XML documents in a particular format.

The Astrophysical Virtual Observatory (AVO) project[10] is a more complex project dedicated to the construction of a Virtual Observatory prototype to be made available to the community for scientific usage, in order to obtain

---

[7] SMETE is an acronym for Science, Mathematics, Engineering and Technology Education.

[8] OPTICON interoperability working group <http://www.astro-opticon.org/interoperability.html>

[9] National Virtual Observatory (NVO) <http://www.us-vo.org/>

[10] Astrophysical Virtual Observatory (AVO) <http://www.eso.org/avo>

science results and user feedback at an early stage of the project. One of its objectives is to establish a set of usable recommendations for helping archive managers to implement remote interoperability.

These recently established projects represent the natural follow-up of a number of other initiatives set up in the last years to define standards for the astronomical archives. These initiatives have produced a set of de-facto interoperability standards (e.g. the description of tabular data or bibcode). Recently, the astrophysical community has started to discuss the possibility of mapping some of these de-facto standards into more widely accepted standards in order to enlarge the access and the exploitation of their documents. The discussions on the choices to be made for such mappings are still in progress, and no consolidated result yet exists.

1.2.4 The archival organisation community

Another community much interested and presently largely engaged in open access to archives is the Public Administration sector (together with National Archives Services).

Archives services have become increasingly prominent in the Commission's action and support programmes [DLM1999, IS2002]. The context in which such activities are developed has, however, specific characteristics that differ from those found in other realities, like the e-prints one. In particular:

- "[digital] archive" has, in that context, the meaning of "set of [electronic] records produced or received by a person or organisation – whether a Public Administration or a private corporation - in the course of its life or activity, and retained by that person or organisation. A record may incorporate one or several *documents* (e.g. when one document has attachments), and may be on any medium in any format. In addition to the content of the document(s), it should include contextual information and, if applicable, structural information (i.e. information which describes the components of the record). A key feature of a record is that it cannot be changed. The management of such records, and their life cycle is subject to stringent national rules.

- National legislation encourages more openness and interoperability as a means of building information systems capable of creating trust between government, regional communities and citizens. In the archival context, however, managing electronic records is a complex issue: in particular, accessing archives other than the historical ones incorporates important issues related to authenticity and privacy. Access to public archives is only permissible under the legislation and technical procedures defined to deliver information to the citizen and to protect the citizen from the inappropriate exploitation of personal and private information. In this context, the task of "service provider" can be taken only by certified organizations that must assume a lot of legal responsibilities.

2 OAI AND THE OPEN ARCHIVE APPROACH IN EUROPE

2.1 Governance of the Open Archives Initiative (OAI)

As an organisation, the OAI is made up of an Executive for management, a Steering Committee for policy direction, and a Technical Committee for evaluation. Carl Lagoze and Herbert Van de Sompel, the Executive, are at time of writing based in the USA. There are a few Europeans on the Steering Committee and the Technical Committee. For example, in June 2002, at the time of the publication of version 2 of the OAI Protocol for Metadata Harvesting, three of the sixteen members of the Technical Committee were based in Europe, and about half of the additional alpha testers were based outside the USA, mostly in Europe. The OAI is supported by funding from the Digital Library Federation (DLF)[11], the Coalition for Networked Information (CNI)[12], and the National Science Foundation (NSF)[13], and by community participation. [14]

[11] Digital Library Federation (DLF) http://www.diglib.org

[12] Coalition for Networked Information (CNI) http://www.cni.org

[13] National Science Foundation (NSF) http://www.nsf.gov/

It is notable that while the Executive and the funders of OAI are USA-based, the success of the OAI is firmly grounded in the participation of a community of people from around the world, particularly Europe as well as North America. Within the limits of its well-defined remit, the development and promotion of a low-barrier interoperability framework, the OAI is quick to respond to community concerns. Nevertheless, the OAI-PMH is tightly controlled by the OAI Executive. Although this is probably a major factor in its rapid development and deployment, it can also be a source of frustration for some in the community. The possibility of turning over control, support and any further development of the OAI-PMH to an existing standards body remains a possibility by the OAI. It may well be in the interest of European implementers of OAI-PMH to advocate the hand-off of the protocol to an international standards body. Now that there is a well-developed and stable second version of the protocol the need to keep control of the development in the hands of a very small number of people who can take independent and speedy decisions may be less important when weighed against the perception of stability and authority conferred (especially in the eyes of higher management in organisations) by, for example, the ISO "brand".

The preponderance of U.S. members, and the dependence on U.S. sources of funding for the OAI may have other implications for open archives in Europe. OAI members, and especially the OAI Executive, have actively participated in European events related to open archives, such as CERN and Open Archives Forum workshops[15]. In addition, the OAI held a European workshop to introduce version 1 of the metadata harvesting protocol. It is essential that this level of communication should be maintained, and that communication should just that: be a two-way flow. Within Europe, it is important to publicise the extent to which the protocol is the result of international collaboration and validation, and in particular the degree of European participation. Early adopters should publicise the benefits being realised through the open archives approach. In doing so, they should take a properly critical stance that analyses any difficulties or deficiencies their experience may uncover.

2.2 Recommendations [to be added in final version]

[To be added]

3 BUSINESS MODELS FOR OPEN ARCHIVES

3.1 Growth scenarios

Active participants in the Open Archives Initiative expect that in a few years time the OAI-PMH protocol will be embedded in the infrastructure of the Web, as taken-for-granted as the HTTP protocol now is. If this is to be so, it will be not only because of the relative simplicity of the OAI framework for interoperability and metadata sharing, but also because of uptake by:

- research organisations, including universities
  (as part of a change to the pattern of scholarly communication)
- publishers, especially learned society publishers
  (adding value to the process of dissemination)
- "memory organisations", i.e., libraries, archives, museums
  (extending access to the citizen)

[14] The OAI is sometimes confused with the OAIS (Reference Model for an Open Archival Information System), which defines an organisational and technical model for the implementation of preservation systems and is supported by an unrelated organisational structure.

[15] 2nd workshop on the Open Archives Initiative (OAI): Gaining independence with e-prints, archives and OAI, held at CERN, Geneva, Switzerland, 17-19 October 2002, http://library.cern.ch/Announcement.htm; Open Archives Forum workshops, URL http://www.oaforum.org/workshops/

3.2 Possible/emerging business models

3.2.1 What is a business model, especially in the context of the Internet?

The simplest definition of a business model is that it is the "method of doing business by which a company can sustain itself – that is, generate revenue" (Rappa, 2001). This does not mean that a business model is *only* concerned with revenue; it should also relate to the value of services and goods provided and the organisation's position in the product supply chain. Thus Mahadevan (2000, p. 59):

> *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.*

There is a wide range of business models in use. Rappa (2001) notes that some models are quite simple: a company "produces a good or service and sells it to customers. If all goes well, the revenues from sales exceed the cost of operation and the company realizes a profit." Others are more complicated and are based on organisations as intermediaries or facilitators. The recent growth in electronic commerce (e-commerce) means that at the moment there is quite a lot of interest in Internet business models, both new and traditional (e.g., Jutla, *et al*., 1999; Werbach, 2000; Feeny, 2001).

Table 1: Taxonomy of business models identified by Rappa (2001)

Business model:	***Brief description:***
Brokerage model	Those that bring buyers and sellers together and facilitate transactions (often fee based)
Advertising model	Supported by advertising revenue, a Web site will provide content and services together with advertising (e.g., banner ads)
Infomediary model	Collecting data about consumers and their purchasing habits and selling this information to other businesses
Merchant model	Selling of goods and services on the traditional retail model
Manufacturer model	Direct selling by the creator of a product or service to consumers, cutting out intermediaries
Affiliate model	Offering financial incentives to affiliated partner sites
Community model	Where users themselves invest in a site, e.g. by the contribution of content, money or time. This can be combined with other models, e.g. advertising or subscription
Subscription model	Where consumers (users) pay for access to the site, usually for high added-value content, e.g. financial information, newspapers, journals
Utility model	A model based on metered usage or pay-as-you-go; depends on micropayments

Source: Rappa (2001)

Mahadevan (2000, p. 59) has commented that there have not been very many attempts to formally define and classify business models in the Internet context. However, there have been some recent attempts to organise and classify them. In one attempt, Rappa (2001) has arranged Internet business models into nine generic categories (Table 1). These include some traditional models that have been adapted for use on the Internet; e.g. those based

on advertising, retailing or subscriptions, as well as models that have been developed specifically to support e-commerce.

An older taxonomy by Timmers (1998) classified eleven business models that were in use or being experimented with to support Internet e-commerce (Table 2). Timmers's classification of commercial business models in use on the Internet mentioned several potential revenue streams. He noted that some models would be able to raise revenue through membership fees (e.g. for 3rd party marketplaces or virtual communities), while others might be based on charging by service or transaction provided.

Table 2: Internet business models identified by Timmers (1999)

Business model:	***Brief description:***
E-shop	Marketing of a company or shop
E-procurement	Electronic tendering and procurement of goods and services
E-auction	Based on electronic bidding, on the traditional auction model but which may integrate contracts, payment and delivery
E-mall	A virtual collection of e-shops
Third party marketplace	Common marketing front-end and transaction support for multiple businesses
Virtual communities	Virtual communities based on communication and information exchange between members, e.g. customers or partners
Value chain service provider	Specialists in specific functions of the value chain
Value chain integrator	Integrator of multiple steps in the value chain
Collaboration platforms	Providers of tools and an information environment for collaboration
Information brokerage, trust and other services	Adding value to data available on the open networks, e.g. searching, customer profiling, etc.

Source: Timmers (1999), Pereira & Fife (2000)

Many of these models are broadly similar to (or are based on) those business models used in traditional (i.e., non-electronic) contexts, e.g. shops, auctions or advertising. The key difference is that the more innovative Internet business models are based on the existence of cheap communication costs. There is, therefore, much interest in services that link different businesses or add some kind of value.

3.2.2 Which might be applicable to open archives?

Taking Rappa and Timmers's taxonomies together, many of these commercial (or quasi-commercial) business models will be familiar to those who work in the research community or in academic libraries and other cultural heritage organisations. For example, publishers have used subscription models for many years to provide journals or monographic series. Libraries have also used intermediaries (brokers) like subscription agents and, more recently, content aggregators like Stanford University's HighWire Press or CatchWord (e.g., Inger, 2001). It is possible also, that some of these commercial business models would be of interest to those cultural heritage organisations that are themselves creating digital content (e.g., Harvard Consultancy Services, 2000). However, the most interesting business models from an open archives perspective might be Rappa's 'community model' or Timmers's related idea of 'virtual communities.' These, as currently defined, are services that gain support from members contributing effort, content or money. Thus Timmers (1998, p. 6) writes that the ultimate value of virtual communities comes from "the members (customers or partners), who add their information onto a basic

environment provided by the virtual community company." If we ignore the specifically commercial aspect, this is broadly similar to Rappa's more generic community model, one based on user investment. As an example of a community model, Rappa (2001) cites knowledge networks:

> *Sites are typically run like a forum where persons seeking information can pose questions and receive answers from (presumably) someone knowledgeable about the subject. The experts may be employed staff, a regular cadre of volunteers, or in some cases, simply anyone on the web who wishes to respond.*

This is broadly the type of model employed by the open-source software movement; described by Ljungberg (2000, p. 208) as "a loosely coupled community kept together by strong common values such that software should be free." The Open Archives Initiative itself could be seen as a similar type of virtual community. Other business models that may have relevance to open archives are Timmers's 'collaboration platforms' (a type of virtual community based on the existence of common tools) and value-added 'information brokers'. Certainly, the movement toward self-archiving and paradigm change in scholarly communication (which is well supported by, and the origin of the open archive approach) is a good fit with the virtual community model.

3.3 Risk analysis [to be added in final version]

[To be added]

3.4 Recommendations for sustainable business models.

3.4.1 Why are business models important?

Business models are a method for reflecting real world processes and could be thought of as an intellectual exercise. It seems unlikely that an entrepreneur or strategic decision maker would think about what they do in terms of business models, more likely they would consider the future and functioning of their organisations in terms of the practicalities of supply and demand, of extending markets or perhaps their strategic goals or mission statement. They will clearly be aware that in order to convert their logistics stream into a revenue stream they will need to put forward a value proposition that will attract customers, funding or sponsorship.

An awareness of the model being used can be valuable in a number of ways - maintaining the relevance of business activities and focusing on aims and objectives. If the model works in the first place then "maintenance" of the model will help to sustain the business. However the world isn't static and broader knowledge of other possible business models will allow a company to adjust or even completely change it's model to better fit circumstances

Table 3: A comparison of Taxonomies

Rappa		Timmers	OA
Community	Open source Public Broadcasting Knowledge networks	Virtual community Collaboration platform Value chain service providers	High applicability to data creator/providers and users forming networks.
Brokerage	Marketplace exchange Buy/sell fulfillment Demand collection system Auction broker Transaction broker	e-auction e-mall 3rd Party market place Information brokerage	High applicability to service providers

	Distributor Search agent Virtual mall		
Subscriber	Content services Person to person network services Trust services ISP's	Value chain integrator	Applicable to data providers

In table3 we see three of Rappa's categories replicated with the individual models included and the corresponding models from Timmers. The final column indicates applicability to OA type organisations such as Data producers, data providers and service providers. Thinking about this the other way around i.e. can the OA approach be used within the other models? the answer is a qualified yes. The case can be made that, even in the potentially least open circumstance of the Merchant and Manufacturer models, open access to metadata about, for instance, stocking level could be useful in business to business transactions.

3.4.2 Business models and Public Organisations

What is not clear so far is how business models might apply to public organisations such as universities, archives, libraries etc. Many if not most public organisations do nevertheless have extant business models, usually in the form of a mission statement or corporate strategy. In most cases these recognise the three streams (revenue, value and logistic) that underpin our definition of business models. A brief poll of mission statements on the internet of public organisations (primarily universities) both in the UK and Europe revealed all had a mission statement or strategic policy. Approximately 25% clearly indicated recognition of all three streams from our definition. Almost all recognised both value and logistics streams. The omission of the revenue stream is not entirely surprising given the considerably lower risk of loosing revenue stream in public when compared to private organisations.

Most public organisations, in common with large corporate bodies, tend to employ multiple concurrent business models. For example University Libraries will employ a merchant model in its dealings with publishers and a community model in it's dealings with students. Although some of the models, outlined in the taxonomies above, can be applied in part to public organisations they do to some extent fall short. The overriding *modus operendi* for UK public organisations, however, is a model that is essentially fixed by the community through the government. For Universities this model works something like this: The government collects taxes and gives part of that money to educators. The educators increase the potential value of the workforce by producing well educated students. The students join the workforce, earn money and pay tax to the Government. This type of model could be called a social subscription model.

3.4.3 *Business models and Publicly Funded Projects.*

A particularly difficult issue is the relationship between the OA movement, OAI-PMH and projects. Projects in this context are digitisation and research projects funded by governments or their surrogates, such as HEFC in the UK. Projects have a split personality from the point of view of business models in that there is a dilemma between the model of the hosting organisation and the model of the funding body. It can be argued that a project is a separate entity that is bound by contractual obligation agreed between the parent bodies. This assumes that clear requirements are given in the contract. In the UK, funding bodies increasingly require digitisation projects to make their digital materials available in "perpetuity". However, there is a lack of linkage between this and clear and consistent advice as to formats, metadata standards and data mark-up schema's. Furthermore, there is little or no strategic co-ordination between funding bodies to ensure cross disciplinary consistency between and within data repositories.

3.4.4 Business models and the OA movement.

Any organisation considering whether the open archives approach can be of use to them needs to consider how their existing business model is effected. Will this new approach form a primary function of the organisation or will it only form one small part? In an organisation where the primary revenue stream is through selling information, it is unlikely that the OA approach would be embraced for data sharing. On the other hand the organisation could benefit from making metadata harvestable.

Public organisations should embrace the OA approach for the following reasons. Sharing knowledge is a primary function and as we have seen is already written into their business models. It provides one means of conforming to the freedom of information acts - enacted Europe wide in 2000 and 2001

3.4.5 Conclusions

Any organisation using a public subscription model is very likely to be both amenable to and gain benefit from integrating the OA approach. Many private organisations could benefit from the OA approach at some level. There needs to be more strategic level coherency between organisations.

4 INTELLECTUAL PROPERTY RIGHT (IPR) ISSUES FOR OPEN ARCHIVES

This section provides a combined abstract and expanded outline of the expert report on IPR commissioned from Mark Bide by UKOLN on behalf of the Open Archives Forum, which was presented at the 2nd OA-Forum workshop, Lisbon, 6-7 December 2002, with the final version of the report published shortly thereafter. Although the term "Intellectual Property" includes a number of different types of intangible property, including patents, trademarks and rights in design, the focus of this report will be copyright (and related authorial rights) and the "sui generis" database right.

4.1 Overview

The relationship between Open Archives and Intellectual Property is complex, not least because of the complexity of definition of what "Open Archives" are. Any reader of the "Open Archives" literature will quickly discover that there are at least three (possibly more) views of what the "Open Archives" initiative is about:

- At its most straightforward, what is proposed is a set of technical standards for the "harvesting" and aggregation of simple descriptions of resources[16] (metadata). The supposition is that the controller of those resources wishes to make information about them more widely available (whether for commercial or other reasons) and is therefore willing to make the metadata available for harvesting in standardised form on a Web site. Through the aggregation of this metadata, new services (particularly but not exclusively resource discovery services) can be developed for users (perhaps targeted at a particular academic discipline, for example). This is entirely a technical protocol. The question of "open access" to the resources themselves is entirely separable from the metadata – as indeed is the question of access to the metadata that has been harvested.

- At another level, the Open Archives movement is seen as way of simplifying the process whereby academic institutions[17] can become publishers of the intellectual output of their own academic staff through the development of online repositories. This may, for example, involve the online publication of "e-prints", perhaps before or after more formal publication in the traditional literature (or perhaps without publication

[16] We use the term "resources" here deliberately – so far as we can tell, there is no reason why those resources described by metadata in an OAI metadata repository should themselves necessarily be digital resources – or (if they are) accessible on the Web.

[17] It is, of course, equally possible for authors to become their own publishers.

elsewhere). The Open Archives approach allows an efficient way of "co-operative marketing" of the content of those archives, encouraging the widest possible dissemination and exploitation.

- At what may be regarded as the most contentious level, some supporters of the Open Archives movement see it as underpinning a strategy to develop what is being called "Free Online Scholarship" – a reversal of the typical scholarly journal publishing model, involving supply-side payment rather than demand-side. Authors (or those who stand proxy for them) pay for publication rather than readers (or those who stand proxy for them).

Although these three aspects of Open Archives are closely inter-related, the questions they raise about Intellectual Property Rights (IPRs) – and perhaps equally importantly the relationship between Open Archives and those businesses which depend on the exploitation of IPRs to support their business model – are significantly different.

4.2 Defining IPRs

The expert report defines and explore these issues, as they relate to both economic and moral rights of exploitation of the works created by authors. Not the least of these issues is the difficulty of managing these rights in a global environment, in the face of substantially different attitudes to IPRs in different jurisdictions and cultures.[18] It explores what constitutes "publication" in the online environment, and how IPRs apply to content that is freely available on the Internet.

4.3 An infrastructure to support IPR on the network

In the global network environment, recognition of and compliance with IPRs, if it is to be achieved at all, will require infrastructural technology support.[19] The expert report explores current approaches to technological support of IPRs being developed within the "content industries" and their relevance to Open Archives and how these relate to protection by law and by licence.

4.4 Metadata and other types of content

What is the difference between metadata and any other type of content? The question of whether and when metadata records themselves are subject to IPRs is an important topic in the context of Open Archives. There are clearly significant differences between items of "intrinsic" metadata which (in the case of textual resources, at least) can be extracted from the resource itself and "extrinsic" metadata, such as qualitative or subject categorisation, which may have a much higher human creative input and be potentially of much higher value.

4.5 Stakeholder attitudes

The interests and attitudes of the various stakeholders – authors, academic institutions, publishers, and users of archives – are also considered in the expert report.[20] What are the significant IPR issues for each of these groups as they relate to Open Archives, and what might be the motivations for each group of stakeholders to co-operate in the development of Open Archives initiatives?

---

[18] This includes, for example, consideration of the growing demand from certain countries for the recognition of collective rights in "indigenous culture".

[19] It is important here to distinguish between such infrastructural support and the technological enforcement of permissions granted by the rights owners. The latter, which is sometimes referred to generically as "Digital Rights Management", is just one of the applications that depends on the existence of the infrastructure.

[20] This will be based on published sources and a limited number of interviews with primary and secondary publishers.

4.6 Implications for Open Archive Services

Implications for open archives services explored in the expert report include the following.

- What are the implications of IPRs for Open Archive Services?
- How might their operation be constrained by third party IPRs?
- What are the substantive risks that they may run related to IPRs and how might these be minimised?
- What are consequences for their own business models?

5 QUALITY ASSURANCE FOR OPEN ARCHIVES [DRAFT FOR FINAL VERSION]

The importance of quality assurance within the context of emerging web based information services is becoming more widely recognised. Within this section of the report we will investigate particular quality issues facing implementations within the open archives context. It is important for those projects considering an open archives approach to consider quality assurance. Experience has shown that often 'production services' emerge from projects, and indeed it is regarded as a sign of success if a project can translate to an on-going service. In order to ensure high quality services there needs to be consideration of quality assurance from the starting date of projects: projects need to consider quality criteria, to establish an evaluation process, and to ensure usability of system solutions.

In order to assist quality assurance the OA-Forum will identify dimensions of quality particularly relevant to the open archives approach, explore how implementations might address these quality issues, and consider who within the open archive framework is responsible for quality.

5.1 Quality assurance for open archives

The importance of quality assurance within the context of emerging web based information services is becoming more widely recognised. Within this section of the report we will explore particular quality issues facing implementations within the open archives context. It is important for projects taking an open archives approach to consider quality assurance, particularly as this approach involves collaboration between organisations and data exchange. In order to ensure high quality services there needs to be consideration of quality assurance from the starting date of projects: projects need to ensure quality of their systems from a technical perspective, but also need to establish policies and evaluation processes which will encourage archive users to have trust in the the archive's service provision.

In order to support quality assurance the OA-Forum has identified dimensions of quality particularly relevant to the open archives approach, explored how implementations might address these quality issues, and considered where responsible for quality lies within the open archive framework.

5.2 Dimensions of quality

The open archive approach is characterised by separation of data provision and service provision, with data (both metadata and original resources) being passed between the data provider and the service provider. There may be a number of parties involved in the system both as 'users' of the data repository and the service(s) based on that data, as well as the providers themselves. This means that open archive implementations must be concerned with managing quality issues in a distributed system, typically a complex distributed system with a number of different stakeholders who have differing requirements.

In order to illustrate this we can consider the potential high-level aims of some key stakeholders in a higher education environment where there is a collaborating 'mesh' of local university institutional e-print archives, subject archives at a national or international level, and other archive services built on these.

Stakeholder (primary e-print tasks)	High-level aims
Researcher (self-archives a paper in institutional archive; retrieves a paper from external e-print service)	Enhanced scholarly communication Effective dissemination of research findings Digital preservation
Institution (as data provider archives research papers in institutional data repository)	Cost-effective access to scholarly output of institution Promotion of the reputation of institution Digital preservation

Subject-based European e-print service (as service provider harvests data from data providers and adds value through services such as authority control, automatic classification, ontology mapping, links to author bibliographies etc)	Providing high quality resource discovery service Document delivery
Student (retrieves a paper from European e-print service)	One stop shop for locating and accessing resources Access to a coherent collection of material

One also needs to keep in mind that each data provider and service provider may well be interacting with a network of other providers. Within such a complex system a number of dimensions of quality can be suggested as particularly significant:

Quality dimension	Elements / Influences
Integrity of data	Synchronising versioning of resources and metadata Harvesting large collections of metadata Linking metadata to resources
Quality of service provision to end-users	Building coherent collections Providing appropriate metadata Providing high quality added value services that meet requirements
Quality of data	Ensuring high quality metadata at metadata creation stage Adequacy of Dublin Core as mandatory OAI-PMH metadata format Editorial control on resources archived in repositories
Sustainability	Explicit commitments to digital preservation policies Consensus on digital preservation strategy

These themes will form basis for discussion at break-out in Fourth Open Archive Forum Workshop.

5.3 Tools.

There are a number of quality assurance 'tools' that already exist that might be re-purposed within the open archive context. These include the DESIRE quality criteria [DESIRE1999], the EDNER evaluation criteria, and the quality standards for cultural sites accessible on the Web: the Brussels Quality Framework [Brussels2001].

5.4 Who is responsible

It would be beneficial for data providers and service providers to have a 'quality policy' that they make freely available to stakeholders.

All services should take responsible for the quality of the service they deliver. Within the open archive approach there will be a number of parties providing 'services', all of these must be responsible for the quality of the

service they provide. An initial step would be for each of these services to make explicit the level of service they provide, ideally following an agreed framework. Given the close interaction and reliance between data providers and service providers it would be sensible to have agreed policies in common amongst collaborating 'providers' whether as regards technical infrastructure such as metadata formats in use, or in relation to organasational issues such as agreements between authors and archives, persistence policy.

More problematic is how to ensure the service to the end-user meets their specific requirements. It would be useful for groups of data and service providers to jointly explore the requirements of their target audiences in order to reach consensus on appropriate collection policies, appropriate added value services, metadata formats, and subject provision. This is quite a time consuming activity and might usefully be considered as a potential area of investigation within national programmes such as FAIR, DARE.

6 METADATA [DRAFT OUTLINE FOR FINAL VERSION]

Organisational Issues - OA Forum [draft sketch]

6.1 Metadata Standards

One of the key organisational issues in the Open Archives community is the development and implementation of metadata standards [and recommendations] for the content of the data providers

The ePrints UK project, funded under the JISC FAIR Programme has produced a set of "best practice" guidelines for the implementation of simple Dublin Core metadata.

The document, entitled "Using simple Dublin Core to describe eprints" by Andy Powell, Michael Day and Pete Cliff

Provides a range of recommendations drawn from:

- DCMI Elements and Element Refinements - a current list [3]
- Using Dublin Core [4]
- The RDN Cataloguing Guidelines [5]

This document provides a range of Eprint-specific recommendations for the for the use of simple Dublin Core Metadata. These recommendations are critical to the ideal of interoperability which underpins the vision of the OAI movement. "Standardised metadata is crucial to interoperability" [Barton]. They also help to draw a distinction between

- Structure and Content [Barton].
- Building Quality Assurance into Metadata Creation

Basic level of metadata quality assurance: spelling as well as name authorities, use of capitalisation as well as broader issues of subject schema and classification.

- Implications for self-archiving
- OAI community has focussed on machine based solutions [barton]
- OA Programmes and Initiatives
- Netherlands - SURF
- UK - FAIR Programme [in particular HaIRST]

The NSDL is a broad program to build a digital library for education in science, mathematics,

engineering and technology. It is funded by the National Science Foundation (NSF) Division of Undergraduate Education http://www.dlib.org/dlib/january02/arms/01arms.html

References -

ARC - http://www.dlib.org/dlib/april01/liu/04liu.html

7 SUBJECT INTEROPERABILITY

7.1 Subject Interoperability: Guidelines for Best Practice: An Illustrative Draft [Draft of Section for final version of deliverable]

7.2 Overview

This document is divided into two parts:

- Guidelines for best practice – rules to apply when choosing your approach in order of priority
- Illustrative examples of how particular guidelines might be applied in practice

7.3 Draft Guidelines

If you can, choose a well-maintained subject or class scheme that suits the needs of your user group, is able to be accurately and consistently applied by your staff, and is in use by the majority, if not all, of the data providers in the group you need to interoperate with. (Note that your community may not just be other archives you are cooperating with, but other parts of your institution and the group(s) they are interoperating with)

If possible without impairing its usefulness, use the scheme without extending or amending it.

If extension and amendment is needed, put in place a mechanism to ensure that you 'harmonise' extension and amendment activity with the other repositories you are interoperating with

Make sure your users have easy to access and use information on the terminology set you are using, that (if possible) they have training in its use, and that your staff have training in its use as a descriptive tool. Make sure that this includes any extensions or amendments.

If creating extensions or amendments, be consistent in the rules you apply when extending or amending and try to 'harmonise' your approach with the approach taken by the scheme maintainers. This will normally require some kind of mechanism to ensure this – see, for example, illustration x below

If it is impossible to agree on a single subject or class scheme for the group of repositories you wish to interoperate with, choose one in use in the group that suits the needs of your repository and its users and apply this following the rules above as appropriate. Attempt to establish a mechanism to ensure interoperability between the different schemes in the group – attempt to provide users with a mapping between schemes, or attempt to agree a single scheme that all repositories can and will apply in addition to their normal scheme (note that this guideline arguably applies to all OAI repositories, and, in fact, to all services across the planet), At the very least, provide users with good guidance of which schemes apply to which repositories.

7.4 Illustrations

This section will provide some illustrations of approaches to implementing the above rules from projects like Renardus and HILT and known practices in the world at large (e.g. use of LC's NACO and SACO programmes in Scotland). These will be based on something like the following:

7.5 RENARDUS

RENARDUS (see http://renardus.lub.lu.se/) is an EU-funded project. Its overall aim is to establish a collaborative framework for European subject gateways that will benefit both users in terms of enhanced services, and the gateways themselves in terms of shared solutions. An important part of the RENARDUS service is its attempt to provide some kind of subject directory browsing service across the participating

gateways. The classification scheme chosen by the project to create a central browse structure is DDC. Mappings are made from DDC to the subject browse hierarchies used by participating gateways. Further information is available in Heery's article (2001).

7.6 MACS (Multilingual Access to Subjects)

MACS (see http://infolab.kub.nl/prj/macs/) is a project funded by the Swiss National Library aims to provide multilingual subject access to library catalogues. The multilingual search is catered for by mapping between three indexing languages used in these libraries: SWD (for German), RAMEAU (for French) and LCSH (for English).

7.7 HILT (High Level Thesaurus Project) Phase II

HILT Phase II (see http://hilt.cdlr.strath.ac.uk/) will set up a pilot terminologies service for the JISC Information Environment looking at a range of ways of harmonising subject access across a range of services using a number of schemes. Included are DDC, UNESCO, LCSH, AAT, MeSH, and others.

7.8 SPEIR (Scottish Portals for Education, Information, and Research)

SPEIR (see http://speir.cdlr.strath.ac.uk/) aims to develop a centralised infrastructure to support a co-operative distributed digital library for Scotland. For the most part, there are no significant problems with legacy metadata, although there are some. Participants use LC authority files as a source of standard forms for names and subjects. Names and subjects not found in these global authority files are agreed between partners, created using the LC format, and held in a local authority file. They are then submitted through the NACO and SACO programmes for inclusion in the global authority files. This is done with the active support of the National Library.

8 GUIDELINES FOR BEST PRACTICE MANAGING INSTITUTIONALLY-CREATED ELECTRONIC RESOURCES USING A CONTENT MANAGEMENT SYSTEM [DRAFT FOR FINAL VERSION]

Basic premise: A CMS can help co-ordinate a process that requires the skill and experience of staff based in a number of different areas of an institution. Examples include:

- Academics and researchers who may be involved in peer review/quality control or initial metadata creation or checking the subject content of a particular aspect of an e-learning package
- 'External Relations' staff who may be involved in promoting institutional strengths or protecting the institution's reputation from the effects of poor quality work
- Librarians who enhance initial metadata to improve retrieval and external interoperability to and run the resource public access and management service
- Legal experts who ensure an e-resource is legal, clean and honest
- Institutional unit whose aim is to exploit institutional resources commercially or strategically.
- [And so on]

Questions:

- Will exact requirements vary from institution to institution? If so, OAF guidelines can presumably be illustrative only with Institution-specific guidelines flowing from a preliminary and detailed institutional requirements analysis?

- What form should guidelines for best practice take? A set of questions to consider? (What kinds of material? Who needs to be involved? What institutional policies are relevant? Which of a list of possible steps (see below) are relevant (a) in general (b) for a given type of material?)

- Is the list of steps below complete? Is the order correct`?

- Do we need an initial description of what a CMS is?

- Is there a linking mechanism here that draws together all other organisational issues? Since even institutions who do not have a CMS will need a requirements analysis? If so, then so what?

Draft steps and rough order (each step may have sub-steps – e.g. metadata and subject description):

Question: Do different types of materials require different workflow 'steps' in a different order? For example, institutionally planned and executed course materials may start at * below, but the creation of a research paper may start at **

- *Collection Management and Development Policy (should we draft a model policy? HaIRST?)

- Identification of requirement (e.g. digitisation decision or course creation decision)

- Identification of individual or group to be involved in creation

- Identification of creation schedule and plan

- Identification of constituent digital objects

- Bidding for funds

- **Electronic resource is created

- Metadata is created by resource creator

- Collection managers informed

- Checks: Research quality/Learning standards/Accuracy and currency of information

- Checks: Is metadata standard sufficient?

- Checks: Legal, clean, honest?

- Checks: IPR issues (risk, liability, protection)

- Security of information, including encryption, digital signatures and watermarking

- Checks: In line with institutional policies?

- Current awareness: which classes, students, academics need to know about it?

- Currency update cycle

- Commercial or Strategic Exploitation potential
- Archiving and preservation decisions
- Electronic publication decision(s) (e.g. internal only partners only, external?)
- Permissions?

9 THE IMPORTANCE OF ORGANISATIONAL ISSUES IN THE TAKE-UP OF THE OAI PMH

9.1 The **downside of the simplicity of the Protocol for Metadata Harvesting.**

9.2 The Concept of the 'Deposited' resource.

9.3 The reluctance to see the Protocol as a part of a publishing process.

9.4 IPR, Knowledge Management, and Content Management beyond the world of Open Archives.

9.5 The use of the OAI PMH in the context of Cultural Heritage Institutions.

9.6 Metadata and Quality control.

9.7 In search of the appropriate business models

9.8 Organisational Issues and the Future of the PMH

PART IV

10 REFERENCES

[Arms2002] William Arms et al. A Spectrum of Interoperability: The Site for Science Prototype for the NSDL, D-Lib Magazine, 8 (1), January 2002. http://www.dlib.org/dlib/january02/arms/01arms.html

[Brussels2001] Experts Meeting on coordination of digitisation policies and programmes, 17 July, Brussels. ftp://ftp.cordis.lu/pub/ist/docs/digicult/brussels_concluding_statement_17_july_2001.doc

[Desire1999] Selection Criteria for Quality Controlled Information Gateways. *Work Package 3 of Telematics for Research project DESIRE (RE 1004)* http://www.ukoln.ac.uk/metadata/desire/quality/

[DLM1999] DLM Forum (archived site). http://europa.eu.int/ISPO/dlm/index.html

[Feeny2001] Feeny, D. (2001). "Making business sense of the e-opportunity." *MIT Sloan Management Review*, 42 (2), 41-51.

[Griffiths2002] Jill R.Griffiths and Peter Brophy. Student searching behaviour in the JISC Information Environment. Ariadne Issue 33. http://www.ariadne.ac.uk /issue33/edner/intro.html

[Harvard2000] Harvard Consultancy Services (2000). *Income generation and sustainability*. NOF-Digitise Technical Advisory Service Information Paper. http://www.ukoln.ac.uk/nof/support/help/papers/incomegeneration.htm

[Heery2001] Rachel Heery, Leona Carpenter, Michael Day, Renardus Project Developments and the Wider Digital Library Context, D-Lib Magazine, 7 (4), April 2001 http://www.dlib.org/dlib/april01/heery/04heery.html

[Inger2001] Inger, S. (2001). "The importance of aggregators." *Learned Publishing*, 14 (4), 287-290.

[IS2002] Information Society Web Site. http://europa.eu.int/information_society/index_en.htm

[Jutla1999] Jutla, D., Bodorik, P., Hajnal, C. & Davis, C. (1999). "Making business sense of electronic commerce." *Computer*, 32 (3), 67-75.

[Ljungberg2000] Ljungberg, J. (2000). "Open source movements as a model for organising." *European Journal of Information Systems*, 9 (4), 208-216.

[Lynch2001] Clifford Lynch, Metadata Harvesting and the Open Archives Initiative, ARL Bimonthly Report 217, August 2001. http://www.arl.org/newsltr/217/mhp.html

[Mahadevan2000] Mahadevan, B. (2000). "Business models for Internet-based e-commerce." *California Management Review*, 42 (4), 55-69.

[Pereira&Fife2000] Pereira, F. & Fife, E. (2000). "Meeting consumer needs on the Internet: successful business models." *Journal of the Institution of British Telecommunications Engineers*, 1 (3), 141-147.

[Rappa2001] Rappa, M. (2001). *Business models on the Web*. Part of the open courseware project: "Managing the Digital Enterprise." http://digitalenterprise.org/models/models.html

[Timmers1998] 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. http://www.electronicmarkets.org/netacademy/publications.nsf/all_pk/949

[Werbach2000] Werbach, K. (2000). "Syndication: the emerging model for business in the Internet era." *Harvard Business Review*, 78 (3), May-June, 85-93.