



*Citation for published version:*

Heery, R & Anderson, S 2005, Digital repositories review. Joint Information Systems Committee.

*Publication date:*

2005

[Link to publication](#)

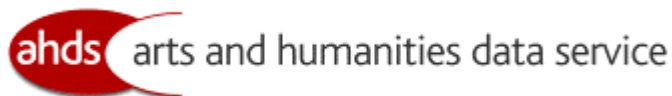
## University of Bath

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



---

# Digital Repositories Review

---

Author:	Rachel Heery, UKOLN, University of Bath Sheila Anderson, Arts and Humanities Data Service
Date:	19-02-2005
Version:	Final version
Document Name:	
Notes:	For circulation

UKOLN is funded by the MLA: The Museums, Libraries and Archives Council, the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils, as well as by project funding from the JISC and the European Union. UKOLN also receives support from the University of Bath where it is based.

The Arts and Humanities Data Service (AHDS) is a national service funded by the Joint Information Systems Committee (JISC) of the Higher and Further Education Funding Councils and the Arts and Humanities Research Council to collect, preserve and promote the electronic resources which result from research and teaching in the arts and humanities.

## **Acknowledgements**

Many thanks to those who contributed to this review by participation in the focus group, questionnaire and interviews. The authors are grateful for contributions from colleagues including Amanda Closier, Leona Carpenter, Liz Lyon, Hamish James, Lorna Campbell, and Phil Barker.

## Contents

<b>1. Introduction.....</b>	<b>1</b>
1.1. Purpose of this review.....	1
1.2. What is a repository?.....	1
1.3. Context.....	3
1.4. Background.....	5
<b>2. Review of current repository activity.....</b>	<b>6</b>
2.1. JISC related activity.....	6
2.2. e-Learning repositories.....	9
2.3. Existing UK national repositories.....	10
2.4. International repository initiatives.....	10
2.5. e-Learning services .....	12
2.6. Data Grids.....	12
<b>3. Ecology of repositories .....</b>	<b>13</b>
3.1. Typology of repositories .....	13
3.2. Ecology of repositories.....	14
<b>4. Technologies.....</b>	<b>15</b>
4.1. Developing user requirements.....	15
4.2. Frameworks.....	15
4.3. Technical architecture challenges.....	16
4.4. Metadata creation.....	18
4.5. Standards .....	18
4.6. Repository software products.....	18
4.7. Other technology issues.....	19
<b>5. Gap analysis.....</b>	<b>19</b>
5.2. Key Actors and Organisations.....	21
5.3. Cultural Issues and Gaps .....	27
5.4. Organisational Issues and Gaps.....	27
5.5. Technical Issues and Gaps.....	28
5.6. Sustainability Issues and Gaps.....	29
<b>6. Concluding Recommendations.....</b>	<b>29</b>

**7. References ..... 31**

Annex 1: JISC Digital Repositories Review Focus Group Report

Annex 2: JISC Digital Repositories Review Software Developers Survey

Annex 3: Repository Issues....from a Teaching and Learning Perspective

# 1. Introduction

## 1.1. Purpose of this review

The Joint Information Systems Committee (JISC) Digital Repositories Programme is initiating a programme of work to assist deployment of digital repositories within the learning and research communities. This review is intended to provide useful background information for participants in this call. The review is not intended to be comprehensive, the intention is to identify useful areas of activity for the programme rather than to prescribe activity in detail.

The methodology of the review has been:

- to undertake a selective review of current activity
- to interview stakeholders by phone and in person
- to hold a focus group of key stakeholders
- to survey by an e-mail questionnaire selected repository software developers
- to undertake a gap analysis

The recommendations made by this report are based both on a review of current activity and on contacts made with a number of interested parties. A gap analysis has been reported within a separate section focusing on feedback from the focus group, survey and interviews.

The potential coverage of this repositories review is vast, coming at a time when there is such a wide range of interest in the subject matter. The authors are aware that much more work is required to thoroughly review the role of repositories and the many issues for deployment. Constrained by lack of time and aware of the context of the review, we have tried to take a broad view of the topic rather than to analyse any particular issue in depth. The real work is left for the forthcoming programme.

## 1.2. What is a repository?

An increasing range of activity areas within the information environment refer to their deposited content collections as 'repositories'. In order to encourage communication across activity areas, and promote interoperability, we need to be able to define the characteristics of 'repositories' and seek the coherence of a common approach.

Increasingly widespread use of a term goes hand in hand with increasing diversity of meanings. Repositories are 'collections of digital objects' but what makes repositories distinctive from other collections of digital objects such as directories, catalogues, databases? What are the defining characteristics of a 'repository'? As with other terms that have been popularised in the digital world (portal, architecture...) some qualification is required: is the repository managed as an *institutional repository*? or a *subject repository*? What is the content of the repository? an *e-prints repository*? a *data repository*? a *learning object repository*? Is the underlying purpose of the repository for *preservation, access, or data management*?

We propose that a digital repository is differentiated from other digital collections by the following characteristics:

- content is deposited in a repository, whether by the content creator, owner or third party
- the repository architecture manages content as well as metadata
- the repository offers a minimum set of basic services e.g. put, get, search, access control
- the repository must be sustainable and trusted, well-supported and well-managed

Enhancing access to scholarly communications has been a main driver for establishing repositories, both institutional repositories (in particular e-print archives) and subject based archives. Many, though by no means all, repositories support 'open access' at least in part. Open access repositories can be distinguished by the following characteristics:

- the repository must provide open access to its content (unless there are legal constraints)
- the repository must provide open access to its metadata for harvesting

The underlying motivations for establishing repositories also differentiate them from other collections. Repositories form an intersection of interest for different communities of practice: digital libraries, research, learning, e-science, publishing, records management, preservation. Within these communities the motivation for focusing on repositories differs somewhat, and the key services that repositories might provide range over several functional areas:

- Enhanced access to resources
- New modes of publication and peer review
- Corporate information management (records management and content management systems)
- Data sharing (re-use of research data, re-use of learning objects)
- Preservation of digital resources

If we consider Lynch's definition of repositories we see an emphasis on the significance of these services rather than on a particular software product or type of content: 'a university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.' (Lynch, 2003)

Various combinations of the areas of functionality listed above have been particularly significant for the growth of institutional repositories, these can be characterised as

- Improved scholarly communication
- Open access
- Content management

The intersection of interest across domains offers possibilities for various crossovers of technologies. There is also potential for sharing experience, sharing tools, and undertaking collaborative development work. It is important that there is co-ordination of this activity and that an appropriate level of interoperability is achieved, without placing barriers on innovative work.

JISC's remit is particularly relevant to such areas of common interest, as stated by JISC in its response to the Select Committee on Science and Technology's Report on Scientific publications: Free for All (House of Commons, Select Committee on Science and Technology, 2004)

"The JISC has a remit to ensure joined-up thinking across the boundaries of research, learning and teaching, and the administration functions within institutions to avoid multiple solutions being

adopted. JISC's vision is to enable the seamless linking of e-research, e-learning, digital library and management information resources, through the co-ordination of technical architectures and standards.”

Within the range of repository activity the development base varies significantly, and it must be said, in many areas development and deployment is patchy and immature. Many existing repositories do support open access and preservation, whether based on the e-print archives model (institutional or discipline-based) or on a digital preservation archive model, typically as a components of the Open Archival Information System Reference Model (OAIS). For a recent overview of OAIS see Lavoie's introductory guide (Lavoie, 2004). There is a somewhat different focus within other communities: for re-purposing learning objects and course materials, for curation of images, research data, and computer software. The learning community has a particular interest in repositories with the development of the IMS Digital Repositories Interoperability Specification (IMS, 2003). Elsewhere managers of institutional information systems, approaching this area from a different perspective, look to content management systems, 'enterprise management systems', or digital asset management systems to fulfil many of the roles of an institutional repository. With the broadening of interest in repositories there is some loss of focus on 'open access' and more concern about how repositories might support institutional strategies.

Still, many continue to have an understanding of 'repository' particular to their own community, for example as an e-print repository, and will have little awareness of activity in other service domains.

**Raising awareness of parallel activity across service domains will be a major contribution of the JISC repositories programme. The call needs to emphasise an inclusive definition of repository: as a managed storage system with content deposited on a personal, departmental, institutional, national, regional, or consortial basis, providing services to designated communities, with content drawn from the range of digital resources that support learning, teaching and research.**

And yet, throughout our discussion of repositories it is worth remembering that the majority of people involved within the education sector have no interest in the existence of repositories, their priority is to gain effective access to the information they need. Should these users have to even know about repositories? The challenge for the programme is to build repository content and deliver the benefits of repositories without burdening content creators and end-users with any additional process

### **1.3. Context**

Repositories must be considered within the context of the wider integrated information environment (the common information environment). Repositories, whatever their flavour, exist within this wider landscape. Fulfilling stakeholders' expectations of return on investment will rely on sometimes complex interactions between repositories and other components of the information environment.

**Future services will rely on well structured work-flow between repositories, and on interfaces between repositories and other components of the information environment.**

The main focus of this review will be on the learning, research and digital library environments. However as an institutional repository can be viewed as a component of an institution's content



management systems (CMS), repositories need to be related to the wider context of administrative systems. Institutions procuring CMS will want to consider digital resource management alongside management of other assets such as web pages, staff and student records, course administration etc. Within institutions one approach is to manage all digital content in one all encompassing system incorporating 'repositories' within the wider system. This is the approach taken by Ohio State University Knowledge Bank (Rogers, 2003), and by the federated repositories at the Los Alamos National Laboratories (LANL) (Jerez, 2004). Repositories also need to interface with other services within institutions such as portals and library catalogues. Some library management systems are now working towards repository functionality as an add-on component (e.g. Innovative, VTLS)

**There is a requirement to explore the role of repositories within wider corporate management systems.**

Just as institutional systems interface with repositories so do personal systems and collections. How do student's portfolios of work fit within an institution's repository strategy? How does a researcher's output stored on their personal hard drives or web pages fit with the institutional repository?

**Repositories need to be positioned within the workflow of personal content creation and the 'personal information environment' of users.**

Institutional repositories must be considered within the wider information environment. Creating small scale 'silos' of information within institutional repositories is not, on face value, a compelling information management strategy in the 'Google age'. The benefits of institutional repositories must be clearly stated. How do institutional repositories benefit the institution itself, and how do they benefit users outside institutional boundaries? Of course an institutional repository must play a role in supporting the institution's management of its own assets, however the promise that institutional repositories will underpin open access and enhance scholarly communication relies on *external service providers* exploiting the combined network of institutional repositories on a global scale. To fulfil this promise services must be built on coherent aggregation of content from a network of institutional repositories.

**Modelling the international network of repositories, developing frameworks for interaction, and providing the necessary infrastructure needs to be addressed by the call.**

A number of services might be layered on either institutional or subject ('themed') repositories. Such services would typically be of benefit to the wider educational community outside the single institution, whilst perhaps simultaneously supporting enhanced services to members of the institution e.g. metadata enhancement, name authority, impact analysis. Such services might be provided from within the educational sector or by commercial players. Services might include

- aggregation of metadata exposed by institutional repositories
- indexing the content of repositories
- impact analysis and provision of other metrics with regard to content
- metadata enhancement services
- metadata creation
- annotation services

**The call should consider how services built on repository content will be delivered, the role of aggregators and other players whether from within both the education sector and the commercial sector.**

## 1.4. Background

There is a view, as typified in a recent Gartner report, that digital repositories are now on the downward curve of the hype cycle (Yanosky, 2004). Moving beyond the hype, increasingly repositories are positioned as having a place within the scholarly communication cycle (Lyon, 2003), within wider digital asset management (Dempsey, 2005) and more fundamentally as part of content management in its widest sense (Conway, 2004).

Previous studies and reports have considered repositories from a number of different perspectives, however these studies have tended to emerge from one particular community or have considered a particular type of repository. A significant emphasis in existing work has been the role of institutional repositories in reforming scholarly publication, notably in reports for SPARC (Scholarly Publishing & Academic Resources Coalition) (Crow, 2002) and for PALS (Publisher and Library/Learning Solutions) (Ware, 2004). Within the SPARC report there is also some exploration of the role of institutional repositories to serve as an indicator of a university's quality and to demonstrate the relevance of its research activity thereby promoting the institution's value. The dual purpose of an institutional repository is supported by the study: "we will narrow our definition to focus on a particular type of institutional repository—one capable of supporting two complementary purposes: as a component in a restructured scholarly publishing model, and as a tangible embodiment of institutional quality."

To these roles can be added additional motivations for institutional archives: supporting accountability for the 'evidence base' of data produced in laboratories; recording health and safety information; fulfilling freedom of information and data protection responsibilities. Funding bodies require institutions to follow good practice guidelines such as those outlined in 'Safeguarding good scientific practice', a joint statement by the Director General of the Research Councils and the Chief Executives of the UK Research Councils (UK Research Councils (1998).

Moving on to interaction with publishers, there is potential for connecting digital repository infrastructure with the production processes of existing commercial indexing services, both Internet search engines and traditional Abstract and Indexing (A&I) services. Already Google Scholar is demonstrating the possibilities of building services on 'crawling' web based repositories in combination with privately agreed access to publishers full text databases (<http://scholar.google.com/>). Traditional A&I publishers such as Thomson ISI are piloting services working with institutional repositories (<http://www.infotoday.com/newsbreaks/nb040301-1.shtml>). OCLC has been collaborating with other initiatives such as DSpace and the JISC ePrints UK project to explore added value services such as automated subject classification and name authority services. For a wider consideration of OCLC's activity in this area see a forthcoming review (Dempsey, 2004). There is scope for commercial services not only to offer added value as service providers, but to feedback and embed such added value into the institutional repository (e.g. providing name authorities, openURLs, enhanced metadata etc).

**There are opportunities to explore added value services built on repository content, and the potential for feeding back outputs of the services to institutional and themed repositories.**

There has been less work done on modelling the interworking of institutional and 'themed' repository models, such interaction would involve establishing a workflow between repositories, and more work to explore the metadata life-cycle, and an overall framework for such interaction.

Some investigation of possibilities of federated architectures are emerging from LANL (Jerez, 2004).

**There is a need to explore interaction of institutional and subject repositories from an architectural perspective.**

Similarly although there has been some comparison of the users perspective on subject repositories in comparison with institutional repositories (Day, 2003), there remains an underlying question as to whether users might identify more closely with an institutional repository or whether the 'discipline based repository' will always be a preferred source of information. (Peters, 2002).

**There is a need to explore role of institutional and subject repositories from the user perspective.**

## **2. Review of current repository activity**

JISC's interest in repositories is part of its strategy to support institutions in long-term digital asset management and preservation. This area forms a central theme of JISC's Continuing Access and Digital Preservation Strategy. Already JISC is funding activity related to repositories in a number of areas, a selection of which follows.

### **2.1. JISC related activity**

#### **2.1.1. Preservation**

Several JISC funded preservation initiatives are in start-up mode.

- Supporting Digital Preservation and Asset Management in Institutions Programme

This JISC programme considers development and management of institutional repositories as a means to ensure long-term access to digital assets. The aim is to provide institutions with practical support in effective digital preservation and asset management. The programme includes projects with the following aims:

- setting up exemplar digital archives of political papers
- developing of training materials
- costing archiving of e-journals
- preservation of assets within FE
- digital asset strategy for academic, learning, and corporate information
- digital preservation assessment tool
- exploring use of the METS content packaging standard
- use of eprints.org software within the OAIS framework
- use of PRONOM file format registry by eprints.org
- implementing a common preservation environment for SHERPA and AHDS framed around the OAIS
- assessment of UK Data Archive and The National Archives compliance with OAIS/METS
- developing METS Training materials
- implementing an ingest service based on the OAIS reference model for institutional archives

- Digital Curation Centre (DCC)

The aims of the DCC are to build a Centre and Associate Network in such a way that research, development, services and outreach interact positively. The aims are to

- Establish a research programme: by addressing wider issues of data curation
- Nurture strong community relationships: by forming and extending the Associates Network, engaging with scientific digital curators
- Development activity leading into services: by testing and evaluating tools, methods, standards and policies in realistic settings and offering a repository of tools and technical information, a focal point for digital curators
- Achieving the 'virtuous circle': by feeding expertise, experience and need into its research programme on data curation and transforming research-led innovation into services that enhance productivity of practice

### 2.1.2. Virtual Research Environment (VRE) programme

Whilst its main focus may not be repositories, this JISC programme includes projects that will investigate building services on repositories to support research activity, and the role of institutional repositories to support institutional portals. Related activity includes:

- CORE: Collaborative Orthopaedic Research Environment

This project will explore issues and requirements involved with providing Web-services that relate to the storage, access, use and re-use, of research data in repositories and information from digital libraries.

- ELVI : Evaluation of a Large-scale VRE Implementation

This project will produce and demonstrate a framework for the deployment of a generic VRE in an HE environment across all the disciplines represented at the University of Nottingham.

- EVIE

This project will test the integration and deployment of key existing software components within a portal framework to support the White Rose Grid research community at the University of Leeds.

### 2.1.3. Core Middleware programme

Although focusing on infrastructure across the information environment there is some activity within this programme relevant to repositories.

- KC-ROLO - Kidderminster College Repository of Learning Objects

<http://www.kidderminster.ac.uk/kc-rollo>

The primary aim of the project will be to set up Shibboleth architecture between Kidderminster College, RSC West Midlands and University College Worcester to provide a long term method of sharing of institutional learning resource objects. The aim of the project is to identify how Shibboleth and PERMIS Open Source architectures can provide an interoperability framework between institutions that will support staff and students in the sharing of institutional repository learning objects.

### 2.1.4. Focus on Access to Institutional resources (FAIR)

The FAIR programme is coming to an end and its outputs will be considered in detail in a separate report. Some follow-on activities are mentioned here:

- UK HE e-theses Test-bed Project

This project will undertake a test-bed implementation project to deliver an infrastructure to support the deposit, access and use of research theses for the UK Higher Education (HE) sector. Building on OAI-compliant institutional repositories for e-theses, the project will provide a range of services that individual institutions would not be able to offer by themselves. Amongst other activities, the project will explore workflow between nationally provided services and institutional repositories, and the provision of repository facilities to allow institutions without a local repository to make their content and metadata more widely available.

- Delivery models for e-prints and open access journals

A study published in late 2004 by the Electronic Publishing Innovation Centre (EPIC) and Key Perspectives Limited, on behalf of JISC, looks at ways in which the delivery, management and access of e-prints and open access journals might be provided (Swan et al., 2004). The report offers a view of different models that might support the newly emerging open access environment, looking at three possible models of open access provision in the UK - centralised, distributed and harvesting.

- Copyright management

Within the FAIR programme the Romeo project delivered useful outputs regarding copyright. SHERPA are taking forward the SHERPA/ROMEO Publishers' Copyright Listings (<http://www.sherpa.ac.uk/romeo.php>). JISC are now co-operating with SURFnet in joint activity in this area (<http://www.surf.nl/copyright/>). A committee has been established, the Zwolle Group, comprising representatives of principal stakeholders in the effort: authors, publishers, librarians, and universities.

- Directory of Open Access Repositories (DOAR)

This project will implement a Directory of Open Access Repositories and will be undertaken by SHERPA in collaboration with the University of Lund's Directory of Open Access Journals.

### 2.1.5. Institutional Records Management perspective

The aim of the JISC Supporting Institutional Records Management Programme is to help both FE and HE institutions implement institutional records management programmes that will establish good practice for the management of records and digital assets throughout their lifecycle. This is of relevance to institutional repositories. Outputs from this programme include:

- Managing primary research data & records

[http://online.unn.ac.uk/faculties/art/information\\_studies/imri/rarea/rm/rm\\_primary\\_res\\_data.htm](http://online.unn.ac.uk/faculties/art/information_studies/imri/rarea/rm/rm_primary_res_data.htm)

This project provided a case study of the practical implementation of the retention guidelines for research data, records and digital assets which will be applicable to many other HEIs. The project utilised and evaluated the application of the records continuum theory to the practical management of digital records and their associated systems. The outputs are presented in the form of a report and a series of 'maps' documenting the issues and recommended solutions.

### 2.1.6. Semantic Web and Autonomic Computing programme

Within this programme the eBank UK project is concerned with enhancing scholarly communication (<http://www.ukoln.ac.uk/projects/ebank-uk/>). The project is investigating ways to link e-prints and peer-reviewed articles to the primary research data upon which they are based. Working in the domain of combinatorial chemistry, the project has developed a proof-of-concept demonstrator service which has populated an institutional repository with data relating to crystal structures, linked this information to related e-prints derived from the primary data and made the information available through a Resource Discovery Network science portal (PSIgate) used in

learning and teaching chemistry. The project also demonstrates the benefits of depositing research data in open access institutional repositories ('publication at source') making the data available for sharing and re-use in a timely fashion without the delay inherent in linking research data dissemination to the traditional journal publishing process (Heery et al., 2004).

The eBank demonstrator is based on OAI-PMH and is underpinned by a data model and metadata schema for crystallography datasets. Phase 2 of the project is planned for 2005-6 and will consider how the eBank schema relates to other emerging science data models such as the CCLRC Scientific Data Model (Sufi and Matthews, 2004) as well as to research data models in other disciplines. More work is needed in this area, both within the chemistry domain and in other domains. In addition, there are semantic issues associated with the interoperability of descriptive terms for research data both within and between related domains.

#### 2.1.7. SunCat

Edina is developing SunCat a national union catalogue of serials covering holdings in twenty-two research libraries throughout the UK (<http://edina.ac.uk/suncat/>). The pilot has been launched in February 2005 in partnership with Ex Libris and with the National Library of Scotland and the libraries of Cambridge, Edinburgh, Glasgow and Oxford, as well as a close working relationship with the British Library and the ISSN Network. SunCat Phase 2 commenced in January 2005 to take forward service delivery and to explore the extension of scope to include research libraries beyond those of the largest twenty-two. There may be potential for 'appropriate copy' interfaces to be developed between the aggregated content of institutional repositories and SunCat using OpenURL technology (Burnhill et al., 2004).

### 2.2. e-Learning repositories

Lists of learning object repositories are available from the University of Texas at San Antonio (<http://elearning.utsa.edu/guides/LO-repositories.htm>) and from the the Academic Advanced Distributed Learning (ADL) Co-Lab ([http://projects.aadicolab.org/repository-directory/repository\\_listing.asp](http://projects.aadicolab.org/repository-directory/repository_listing.asp)).

The JISC Exchange for learning (X4L) programme has explored re-purposing learning content, both JISC funded content and content created by other bodies and agencies where intellectual property rights allow for educational use, or can be negotiated. The programme has had as one of its main aims establishing and populating a national repository for the deposit of re-purposable learning materials, case studies and exemplars.

- **JORUM**

As part of the JISC Exchange for Learning (X4L) programme, the JORUM+ project investigated provision of a learning object repository service through development of practical test-bed services. JORUM is now transitioning from a development project to a JISC production service, whilst continuing with a research strand. The JORUM repository service has now procured a suitable software platform, IntraLibrary from Intrallect (<http://www.intrallect.com>). JORUM will establish and launch a learning repository service for all Further and Higher Education Institutions in the UK that wish to take the service from August 2005.

Other UK repositories have been established with funding from institutions and other central funding agencies:

- **Stòr Cùram**

A digital repository of learning objects for Scottish social work education funded by the Scottish Institute for Excellence in Social Work Education. The project is led by the University of Strathclyde (<http://www.storcuram.ac.uk/>).

- Learning Object Repository for Edinburgh University (LORE )  
LORE provides a learning object repository for the University of Edinburgh's e-learning projects and will investigate the provision of a university wide repository. Initially the repository is only available to University of Edinburgh staff (<http://www.lore.ed.ac.uk/>).
- High Level Skills for Industry Project (HLSI)  
Funded by Yorkshire Forward (the Yorkshire and Humber Regional Development Agency), the initial remit of the HLSI Project was to establish an on-line repository of learning materials to support the delivery of learning programmes in the subject areas of engineering and manufacturing from Key Stage 4 to Higher Education levels. This repository has now been developed and the project has expanded to take in the whole of the Yorkshire and Humber region with a wider subject coverage - from psychology to animal care (<http://www.hlsi.org.uk/>).

### **2.3. Existing UK national repositories**

There are a number of well known national services serving the HE and wider research communities providing archives for research data. These include AHDS, CCLRC ATLAS, the NERC Data Centres, the Data Archive with the associated ESRC service, as well as the National Archives, The British Library, The National Libraries of Scotland and Wales, Northern Irish Archives and Libraries National Archive, the National Digital Archive of Datasets. New initiatives will need to integrate with these and other national services.

### **2.4. International repository initiatives**

Activities of interest include:

- ARROW (Australian Research Repositories Online to the World)  
ARROW is funded by the Australian Department of Education Skills and Training (DEST) to explore and test solutions to establish institutional repositories at Monash University (the lead institution for the project), Swinburne University of Technology, The University of New South Wales and the National Library of Australia. ARROW has chosen the Fedora open source software as the storage layer software for repositories to be established at the four ARROW partner sites. Repository content management workflows and searching are supported by the VITAL software from VTLS. ARROW's resource discovery service is built on metadata harvested from the project repositories and is currently being tested. Initially ARROW is focusing on textual research outputs (e-prints, electronic publishing and digital theses) with the intention of encompassing learning objects and research datasets at a later stage of the project. A recent report on the project was given to the Information Online Conference in Sydney, February 2005 (Payne, 2005).

- CORDRA (Content Object Repository Discovery and Registration/Resolution Architecture)

The CORDRA reference model is intended to provide for the federation of learning object repositories, building on technologies developed for other content repositories and digital libraries. The main focus is on finding and re-using learning content. The overall approach is to federate learning object metadata from several repositories via a registration process and then to enable search on the combined metadata. The CORDRA activity is being co-ordinated by the Advanced Distributed Learning Initiative (ADL) with support from the Corporation for National Research Initiatives, (CNRI) and the Carnegie Mellon Learning Systems Architecture Lab (LSAL). CORDRA has emerged from the ADL's SCORM initiative, but is not intended to be limited to SCORM content. Creation of CORDRA infrastructure demonstrators are being

developed in Australia and the US, plans include pilot testing of the ADL Registry and putting the ADL Registry into limited production. For recent discussion see a report on "Interoperability state of play" at the February 2005 IMS Melbourne meeting (Kraan, 2005).

- Johns Hopkins University Digital Knowledge Centre: A Technology Analysis of Repositories and Services

The Digital Knowledge Center (DKC) at Johns Hopkins University, working with the University of Virginia (UVA), the Massachusetts Institute of Technology (MIT) and the Sheridan Libraries network of international partners, has been funded by Mellon Foundation to carry out an evaluative analysis of repository software and services supporting e-publishing, e-learning, and digital preservation (<http://ldp.library.jhu.edu/repository.html>). A number of systems will be evaluated against a series of use cases. The outputs of the study will inform the development of Fedora and DSpace, and will result in a typology of repositories and repository users.

- Digital Library Federation (DLF)

The Digital Library Federation in partnership with Emory University, the University of Illinois at Urbana-Champaign, and the University of Michigan, received in September 2004 an Institute of Museum and Library Services (IMLS) National Leadership Grant to research, design, and prototype a "second generation" OAI finding system, building on the lessons learned from the first wave of OAI harvesting (<http://www.diglib.org/architectures.htm>). The project will use as its test-bed collections drawn from across the DLF membership. The aim is to foster better teaching and scholarship through easier, more relevant discovery of digital resources, and enhance libraries' ability to build local services on top of distributed metadata collections.

- National Digital Information Infrastructure and Preservation Program (NDIIPP)

NDIIP, a national digital strategy effort in the US, is led by the Library of Congress in co-operation with other federal, research and private libraries and institutions including the National Library of Medicine, the National Agricultural Library, the Research Libraries Group, OCLC and the Council on Library and Information Resources. As part of the program, in September 2004 the Library of Congress awarded funding of \$14.9 million to eight high profile initiatives to identify, collect and preserve digital materials within a nationwide digital preservation infrastructure. NDIIP has also carried out a series of studies to support planning for the program considering policies, protocols, strategies and technological infrastructure for the long-term preservation of digital materials, including a proposed architecture (NDIIP, 2003). The architecture considers requirements for federation between institutions and repositories, and is therefore of particular interest to this call.

- Digital Academic Repositories (DARE)

DARE (<http://www.darenet.nl/en/>) is a joint initiative of the Dutch universities to make all their research results digitally accessible. The KB (National Library of the Netherlands), the KNAW (Royal Netherlands Academy of Arts and Sciences) and the NWO (Netherlands Organisation for Scientific Research) are also cooperating partners. The SURF Foundation is coordinating the programme is being taken care of by the. The programme runs from 2003 until the end of 2006. A pilot search service is now available, and a number of services are in development: **CoMa: Copyright Management** supporting interaction with publishers' regarding copyright terms for deposit in institutional repositories. **Connecting-Africa** providing access to African research information and materials produced in the Netherlands. **Essays Online** is the start of a national database of graduate papers and extended essays harvested from repositories; **P-Web: a tool**



**for publishing proceedings online** is a Web-based tool for publishing conference proceedings on-line by using institutional repositories for the entry and storage of documents.

## 2.5. e-Learning services

There are a number of catalogues (also known as gateways) linking to educational materials held elsewhere e.g. RDN, GEM, ARIADNE, Edna. As these services do not include deposited content they are not 'repositories' by the definition used within this review, rather these are 'service providers' (catalogues or directories) that create or aggregate metadata. As there is potential for such services to inter-work with repositories a selection of such services are included here. MERLOT is mentioned as it is distinctive in its inclusion of peer reviews

- Curriculum Online

Curriculum Online is intended to improve access within schools to multimedia resources. The service provides search of a catalogue of metadata describing thousands of digital learning resources in the UK. Users can locate relevant content and are linked to suppliers of that content with whom the user negotiates access (using the government e-learning credit system). The service is managed by Becta, funded by the DfES, to give teachers easy online access to a wide range of digital learning materials, which they can use to support their teaching across the curriculum.

- Multimedia Educational Resource for Learning and Online Teaching (MERLOT)

MERLOT stores metadata and peer reviews for learning material stored elsewhere. MERLOT is a free and open resource designed primarily for staff and students in higher education. Anyone can contribute descriptions of learning materials to the catalogue, or use MERLOT material subject to licensing and rights agreements. MERLOT conducts structured peer reviews of online learning materials carried out by editorial boards.

## 2.6. Data Grids

Data grids have been envisioned by the computer science community as distributed computer networks which provide access to data for a range of scientific and engineering applications. A number of exemplars are being developed, these include:

- National / international data grids such as the EU DataGrid which is now part of the EGEE project <http://egee-intranet.web.cern.ch/egee-intranet/gateway.html>

Subject-based grids such as:

- AstroGrid the data grid for UK astronomy <http://www.astrogrid.org/>
- Particle physics Data Grid <http://www.ppdg.net/>
- Geon - the Geosciences Network <http://www.geongrid.org/>
- EcoGrid which is part of the Science Environment for Ecological Knowledge (SEEK) at the SDSC <http://seek.ecoinformatics.org/Wiki.jsp?page=EcoGrid>

The data repositories involved have different data models, different schema and different underlying software and there are major issues of data integration and management. In addition, the publishing of datasets as part of the scholarly communications process is becoming more evident and there are issues of provenance related to the dynamic nature of the datasets,

### 3. Ecology of repositories

#### 3.1. Typology of repositories

Although there has been a lot of discussion and debate around repositories the implemented base is small, and with only a few exceptions, tends to consist of single-role, small-scale repositories. The variety in the types of e-prints repository can be seen by browsing the registry at the University of Southampton e-prints repository (<http://archives.eprints.org/>) or the registry of data providers and service providers at the Open Archives Initiative (<http://www.openarchives.org/community/index.html>) or the registry of archives at University of Illinois at Urbana-Champaign (<http://gita.grainger.uiuc.edu/registry/>). In addition to such registered OAI-PMH compliant archives, there are a number of repositories concerned with other resource types, such as research data and learning resources, that are not yet OAI compliant.

**In order to aid communication it would be useful to develop a typology of repositories.**

##### 3.1.1. Initial typology of repositories

A suggested simple typology of repositories follows:

*By Content type:*

- Raw research data
- Derived research data
- Full text pre-print scholarly papers
- Full text peer-reviewed final drafts of journal/conference proceedings papers
- e-theses
- Full text original publications (institutional or departmental technical reports)
- Learning objects
- Corporate records (staff and student records, licences etc)

There is some content that currently appears to be largely missing within deployed repositories. For example there is little evidence of awareness within repository deployment of connections with archival management of courses as opposed to learning objects, what Lynch refers to as 'composite structures (such as entire courses – in various sense, including both course "frameworks" and actual populated "instances" of courses within such frameworks – exported from learning management systems)...' (Lynch, 2003).

A significant number of entries in institutional repositories are 'metadata only' with no link to the full text. This appears to be due to caution regarding copyright and IPR. Repository administrators and authors are reluctant to come into conflict with publishers regarding copyright issues so will not include 'full-text' when there is doubt about copyright. In addition some repositories will only include links to full text for those entries published and/or authored whilst the author was employed by the institution. So for example only a percentage of entries within the Southampton ECS e-prints repository link to full text, and the CCLRC repository has a significant percentage of metadata only records.

*By Coverage:*

- Personal (author's personal archive)
- Journal (output of a single journal or group of journals)
- Departmental
- Institutional

- Inter-institutional (regional)
- National
- International

*By primary **Functionality** of repository:*

- Enhanced access to resources (resource discovery and location)
- Subject access to resources (resource discovery and location)
- Preservation of digital resources
- New modes of dissemination (new modes of publication)
- Institutional asset management
- Sharing and re-use of resources

*By **target user group***

- Learners
- Teachers
- Researchers

### **3.2. Ecology of repositories**

How do these different types of repository interact? How far is interoperability achievable between repositories of such diverse types?

At present there is very little interoperability between repositories. For example, e-print institutional repositories are unlikely to be linked to or interact with repositories for teaching and learning. Software does not facilitate sharing services between repositories, or provide the full range of functionality that users might require - users in the broadest sense to mean those submitting content, those managing content, and those using content.

Interesting work is now emerging considering interaction between repositories in the context of the digital library and learning community (McLean and Lynch, 2004). Recent work by Blinco and others seeks to map a repository landscape, and to place some order on the present somewhat confused and fragmented picture (Blinco et al., 2004). McLean has put forward a tentative 'ecology' of repository services that may help to identify common services and to bring about a convergence of service domains (McLean, 2004). It would be useful to take this forward to consider commonality of services. This work usefully spans input from the US, Australia, Canada as well as the UK. It would be timely to take this forward at a technical/operational level.

There is potential to expand such work to also consider research repositories, the e-science community, and corporate repositories, though mindful that scope might be too challenging.

***A framework needs to be established for repositories that would encompass:***

- ***relation between repositories***
- ***data flow between repositories***
- ***workflow issues***

This would begin to address fundamental questions, such as how institutional repositories relate to thematic, subject repositories? Within institutions, how do repositories relate across the

'service domains' of research, learning, administration? A meeting point is required at various levels, both as regards service provision and technical infrastructure.

**JISC's role in establishing a framework for repositories might provide a co-ordinating focus for such UK activity nationally, and feed into international activity.**

Note: this is an area of transition, that needs to mature before mandating practice.

## **4. Technologies**

### **4.1. Developing user requirements**

In order to guarantee success it is vital that repositories meet the short-term needs of users as well as underpinning the longer-term strategic objectives of funders and institutions. There is a need to explore user requirements and prioritise them in the development of repositories, to articulate scenarios investigating how real users would benefit from deployment of repositories.

**The process of repository development needs to engage the user community in a real way by such methods as usability studies, participative development process developing and refining use cases on an on-going basis.**

### **4.2. Frameworks**

There is potential to leverage the current widespread interest in repositories to progress consensus on a common framework across service domains. Reaching a common perspective on repositories will assist development of a common framework.

There is a need to reach consensus on a framework that will support a distributed network of institutional repositories interacting with national and international initiatives. This framework needs to encompass the range of repositories involved in provision of preservation, enhanced access and resource discovery. Given the increasing heterogeneity of repositories and their content, such a framework will be a perspective on the wider integrated information environment framework. This perspective will be a view concentrating on 'user facing services' provided by repositories and the required m2m services, work-flow and data-flow (rather than on the internal workings of repositories).

**JISC can encourage existing activities to articulate a common framework amongst the various service domains of research, e-science, e-learning, digital libraries, administrative computing.**

Given the diversity of repository activity there needs to be some judgement exercised as to balancing the costs of interoperability with benefits. Facing this issue within the realm of digital preservation, the NDIIP has sought to provide an architecture (or framework) appropriate across domains recognising that systems are designed primarily to fulfil the requirements of a particular organisation, concluding that "the trivial interoperability of 'everyone uses the same tools and formats' and the deeper interoperability of 'everyone uses the same conceptual model' are both unattainable, now and for the foreseeable future" (NDIIP, 2003). The NDIIP architecture seeks to support co-operation whilst allowing for different technical solutions.

OAIS provides a conceptual reference model, developed within the context of digital preservation, this model is also very relevant to repositories whose primary purpose is access. Many aspects of OAIS may be applicable to other sorts of archives. An associated strand of work between Research Libraries Group and OCLC in 2002 considered the attributes of digital repositories within the context of research organisations, recommending the essential attributes that would form a framework “for trusted, reliable, sustainable digital repositories” (RLG, 2002). Informed by the OAIS this work is positioned in the context of research institutions, and a checklist of operational responsibilities has been drawn up.

Later collaborative work in 2004 under the aegis of the DLF considered how repositories might better support the use of digital library content in course management systems (Flecker and McLean, 2004). As part of their work this group drew up a checklist of both essential and desirable operational services for digital repositories inter-working with teaching and research applications. The report includes recommended best practices and guidelines, and through exploring use cases, creates a model for interaction of tools, users and repositories.

Already there are several projects within the JISC Digital Preservation and Asset Management in Institutions Programme considering applicability of OAIS within institutions. The DCC remit also includes drawing up and disseminating good practice guidelines not only for preservation but for curation (management and access) of repositories.

**There is potential collaboration between the repositories and digital preservation strands of JISC activity. This work will need to join up with international activity.**

**Need for demonstration projects**

**to explore interoperability between different repository types.**

**to implement common services across repositories**

**Need to encourage communication across domains and stakeholders.**

### **4.3. Technical architecture challenges**

There is a risk that the services supported by such a complex ecology will be fragmented, incoherent and overlapping. Ideally services would take a common approach to standards, protocols and interfaces between components of the framework. To provide effective resource discovery and preservation across distributed repositories there must be agreement on an overall technical architecture: metadata standards, agreed method of linking to digital resources, and common resource discovery protocols.

Experience from repository implementation over the last year has thrown up a number of issues that must be addressed in order to facilitate more effective services being built on repositories. In particular there is need to agree technical approaches for the following:

- Harvesting full text of e-prints

Typically added value services built on repository content involve harvesting full content of the e-print (or other resource) and manipulating that resource in some way e.g. indexing, automated subject classification, applying name authority, citation analysis. However initial experience of harvesting within the ePrints UK project has shown that extracting the full-text of an e-print can be a complex process (Tourte, 2005). It is non-trivial to locate the actual files that contain full text as repositories have not been consistent in the way they link from the metadata to the full text.

**A relatively simple solution might be to mandate, within the collaborating federation of UK repositories, a consistent approach to linking to resources from the metadata.** This is recommended as a short-term solution within the UK, and could be promoted by UKOLN and SHERPA in their existing roles.

A more profound solution to this problem would be to create a structured package to more fully describe the 'complex object' that makes up the e-print, see below.

- Handling complex objects

Implementation experience within institutional repositories in UK, Netherlands and US has shown that even within the relatively narrow scope of an e-print archive there is a requirement to handle complex objects. In this context the complex object will typically include metadata and multiple formats of a journal article (PDF, WORD, HTML). Repository implementations with more innovative repository content such as research data, multimedia objects, courseware, will need to manage complex objects of a different type. Interoperable repositories need to encode, exchange and describe these complex objects in agreed ways.

OAI technical experts have had initial discussion at meetings and on the OAI technical mailing list. These initial discussions have suggested that work needs to be done on an international basis to reach consensus on ways of packaging structured objects and identifiers. Initial areas for investigation might be METS, MPEG DIDL, IMS Content Packaging Specification.

**It is recommended that an international technical group is established to draw up a specification for OAI-PMH handling of complex objects. This group needs to consult with data providers (repository owners), service providers and repository software developers.**

- OAI-based architecture for federated repositories

As repositories become favoured as a good solution for managing digital assets across institutions then naturally there will be increasing repository numbers with increasingly diverse content. Already we see that within institutions which have embraced repositories a number of distributed archives are established, albeit on a relatively small scale e.g. both the University of Southampton and the University of Glasgow have multiple repositories. A larger scale example is found at the Research Library at Los Alamos National Laboratory

Such distributed repositories might be managed as 'federations' in order to support harvesters downstream of the distributed repositories, and end-users within institution and beyond who wish to see a 'combined view' of distributed repositories.

Various solutions might be explored to support federation, such as the LANL Repository architecture, Dspace, Fedora. Within LANL a 'multi-faceted OAI architecture' using OAI-PMH at various levels has been developed to include Repository Index, OAI-PMH Federator, and Identifier Resolver components (Jerez et al., 2004). Jerez discusses the potential for managed inter-institution federations of repositories and, more tentatively, for loosely-structured federations of repositories distributed over the Web. The Fedora architecture is also of interest, initially developed by Cornell and CNRI, this provides an architecture that aims to store a variety of resource types - books, journals, corporate reports, software - as well as complex multimedia entities. While each type of content has unique aspects, the objective is to manage content in a uniform manner. Payette defines a common set of operations to perform basic repository

management functions such as storing, copying, depositing, and archiving disparate forms of data (Payette et al., 1999)

**Further work is required to demonstrate architectures to enable effective services to be built on harvesting complex objects from distributed repositories with heterogeneous content.**

#### 4.4. Metadata creation

A recurring theme is the need for improvement to the repository ingest process or, in common parlance, 'better metadata creation tools with some level of automated metadata creation'. The need for automated metadata creation as part of the normal workflow was mentioned as a requirement by the focus group. The same issue was raised in a response to the developers' questionnaire: "The big issue in institutional repositories is how to get authors to self-archive, or to design systems that don't rely on author self-archiving. The former will probably depend upon whether the repository front end can be integrated with tools that authors already use."

Some implementations report that the simple Dublin Core default as specified in the OAI-PMH is not adequate to describe repository resources effectively. In addition metadata more sophisticated than simple Dublin Core is needed to allow subject aggregators to deliver real value.

**Simple Dublin Core metadata (oai\_dc) is not sufficiently rich to describe repository content, there needs to be agreement on qualified Dublin Core schemas to describe different repository resource types. The immediate requirement is for agreement on a qualified Dublin Core application profile for describing e-prints.**

Issues are beginning to emerge in relation to lack of interoperability between Dublin Core and IEEE LOM metadata models. This results in complexities for data exchange and cross-searching.

**Need to explore interoperability issues arising from deployment of multiple metadata formats within the existing UK network of repositories.**

#### 4.5. Standards

Interoperability between repositories requires consensus on standards. Development of standards is labour intensive but is a necessary investment to support interoperability. There is therefore a need to continue *involvement in development of standards* to support repositories in the UK in partnership with international standards making bodies. In addition there is need to foster agreement on *common data models, common approach to packaging of complex objects, common schemas, common means of linking to full resource.*

#### 4.6. Repository software products

Repository software is still considered inadequate in terms of functionality. Some concern has been expressed about the long-term viability of open source software solutions. There are few examples of stakeholder analysis or detailed functional requirements for repositories.

**There is a need to build on existing product user groups (or to create such groups if necessary) in order to steer development to the requirements of stakeholders.**

## 4.7. Other technology issues

- Provenance

Copies of e-prints within repositories are 'final draft' rather than the copy actually published in the journal. This raises issues of provenance. There are also issues as to how different repositories deal with versioning of preprints. The eBank UK project is producing a supporting study on provenance of e-prints. More investigative work is required in this area.

- Annotation

Further work might be considered regarding annotations in the context of digital repositories. The W3C proposes that annotations are treated as metadata and a list of annotation tools is given at <http://annotation.semanticweb.org/annotation/tools>. These include Annotea (<http://www.w3.org/2001/Annotea/>) and GATE (<http://gate.ac.uk/>). Different aspects of the role and function of annotations are illustrated by the following exemplars: DLESE (<http://www.dlese.org>) and eProtein (<http://grid.ucl.ac.uk/eProtein.html>). eProtein involves the European Bioinformatics Institute, and is clearly working in a specialized area of bioinformatics, but it illustrates the importance of data annotations to the research process in this field.

Whilst annotation is a named research priority of the Digital Curation Centre and work is underway in this area, there are wider aspects which would merit further investigation:

- User requirements and priorities.
- Experience of annotation of a variety of format types including text, audio, moving images, 3D images, maps etc.
- Development of a common data model and schema for descriptions of annotations.
- Development of an annotations server as a shared service for the JISC Information Environment.
- Practical implementation in a range of repository scenarios in both research and learning & teaching.
- Recommendations on shared tools and semantically-aware services.

## 5. Gap analysis

The following gap analysis is based on interviews with stakeholders, a focus group undertaken as part of the review, and a survey of software developers providing repository systems.

In order to start to start to unpick complexity and to identify the issues and gaps to be addressed by the next JISC call, it is useful to identify the drivers behind repository development, the issues they seek to address, and the expected outcomes they hope to achieve. From the background research undertaken for this review four key drivers have emerged as the forces behind repository development in the UK:

- *The Publications Crisis*
- *Scholarly communication and sharing*
- *Teaching and Learning, including re-use and re-purposing*
- *Management of digital assets and Preservation requirements*



### 5.1.1. The Publications Crisis

Libraries claim that the increasing cost of subscriptions is becoming untenable and they are therefore seeking to resolve this by establishing e-print repositories for the deposit and sharing of published works. Once a critical mass of publications is made available then libraries will be able to reduce subscriptions to all but the most essential journals. In the longer term it may well be that new publishing paradigms emerge (e.g. the author pays model) that reduce (or remove altogether) the cost of subscriptions.

However, evidence thus far indicates that populating e-print repositories is proving difficult. Many scholars are concerned about the impact on traditional publishing methods, and have yet to make the transition to automatic submission of their publications to their local e-print repository. Moreover, there are still considerable problems with ensuring deposit of the 'definitive' or master version of the publication. If e-print repositories are to be widely used and respected as providing access to peer-reviewed versions of publications, then quality control of some sort needs to be applied for both metadata and content.

### 5.1.2. Scholarly communication and sharing

Scholars need to share their research results (and the data these are based upon) more quickly and easily than is available to them through the usual publishing route and are therefore arguing that their publications should be deposited in e-print repositories, and the underlying data in either institutional or subject repositories for sharing with others. However it seems that this is sporadic and more advanced in some subject areas than others. In the research context, scholars are more likely to think along subject lines and to share, and indeed be working with colleagues based at different institutions nationally and globally.

Complexity of materials also becomes an issue here – increasingly research data is large-scale and complex, and even where not, brings with it significant issues for its management in terms of metadata creation and quality assurance, IPR issues, and preservation and accessibility.

In addition, many researchers and teachers are using more informal methods for sharing such as informal networks, wikis, peer-to-peer mechanisms; and providing access through owner created and managed websites. Repository development must take into account these less formal and less regulated methods of sharing and managing access to content, consider how they might be supported, and at what point (if at all) content might be accessioned into a more formal managed repository environment.

### 5.1.3. Teaching and Learning, including re-use and re-purposing

There is clear scope for further use and re-purposing of many digital objects and collections. Repositories have been established to manage and disseminate digital objects and to enable them to be packaged and re-used. The recent review of the JISC X4L programme (Baldwin, 2004) stated that the primary purpose of repository development within teaching and learning has been to ensure the availability of content, to improve the quality of the learning experience, and to cater to different learning styles.

Baldwin identified several areas where cultural change may occur or where change is needed to support further use and repurposing. She suggested that teachers must be willing to share their learning materials and that for repurposing to become mainstream, it needs to be embedded in institutions and part of their educational strategies. Embedding will involve understanding how to cascade knowledge from 'pioneers' like those involved in X4L to others, and facilitating it.

Repurposing could stimulate cultural change in teaching and learning, as teachers have to rethink how they deliver their courses and focus on how to improve the quality of the learning experience.

There is also evidence from JORUM to suggest that disaggregating learning materials into discrete learning objects is problematic from an IPR point of view. The proposed alternative is the creation of 'resource stubs' (metadata describing a learning resource) for inclusion in JORUM for the purpose of searching. However, this limits the opportunity for re-purposing, and it could be argued that a solution based around extended OAI-PMH harvesting for learning objects might be a better solution.

It was also suggested by the X4L programme manager that teaching and learning repositories needed to focus on delivering to 'communities of practice' if significant take-up and use of the content was to be achieved.

#### *5.1.4. Management of digital assets and preservation requirements*

The establishment of the Digital Curation Centre (DCC), the success of the Digital Preservation Coalition (DPC), and ongoing support for organisations such as the UK Data Archive (which hosts the Economic and Social Data Service), the Arts and Humanities Data Service, and the NERC Data Centres all point to the increasing importance of service provision for the effective creation and management of digital assets of all kinds.

However, it is clear that effective curation and preservation for the long term is a complex process and the necessary methods and skilled staff are only now starting to emerge. This is likely to place significant constraints on what can be achieved in this area, particularly in the more informal settings, and in smaller or less well financed institutions. Whilst the DCC is intended to provide advice and support for those undertaking preservation, and the DPC seeks to raise awareness of the issues, it is essential that the requirement for effective curation and preservation of many, if not all, digital assets of all kinds is taken into consideration.

One of the benefits of bringing digital assets into a managed repository framework is the promise of future proofing against technology obsolescence. However, current repository software does not fully support the preservation process, and the staff with the necessary skills to undertake this work are few and far between.

## **5.2. Key Actors and Organisations**

Broadly speaking these might be broken down into three key groups:

- *Institutions – both HE and FE*
- *Individuals or groups of individuals e.g. research groups, teachers, learners*
- *National Services and organisations e.g. British Library, National Library of Scotland, UKDA, AHDS*

In addition, there are other groups and organisations with an interest in repositories, primarily as funders of repositories or as funders of content that might be included in a repository, including government, JISC and the Research Councils. Indeed, both JISC and some Research Councils fund repositories to manage, disseminate and preserve the research data and other outputs arising from their grants or specific data collections funded by them. These tend to be on a subject basis, reflecting the Research Council infrastructure.

RCUK has established a Data Curation and Archiving working group to address the many issues surrounding the curation, preservation and long-term access to the large volumes of research data which they either fund the creation of, or fund access to. JISC should take this work into account when devising its own repositories programme.

RCUK also supports an active scholarly communications group which is currently working on a joint policy regarding the deposit of publications in institutional repositories. It seems likely that that this group will support in some form, the deposit of e-prints in institutional repositories

For the purposes of this analysis commercial publishers and content providers are excluded.

### 5.2.1. Institutions

A limited number of institutions are starting to establish repositories for managing one or more resource types. Approximately 38 UK HE institutions have established working demonstrators of OAI compliant e-print or e-theses institutional archives, most with the help of FAIR funding. A few of the larger universities are including some research data, in particular collections of images and texts (these tend to be easier to create and manage than larger or more complex digital data types).

Name	Number of Records
<b>Institutional e-print repositories</b>	
St Andrews	259
AIM25 - Archives in London	n/a
Durham e-Prints	4
Edinburgh Research Archive	256
e-Prints Soton	3179
CCLRC ePublication Archive	20774
CSC Eprints	43
DSpace at Cranfield University	468
ePrints@Bath	9
Glasgow ePrints Service	647
Nottingham ePrints	70
Oxford Eprints	400
The Open University Library's Eprints Archive	3
UCL Eprints	86
University of Glasgow EPrint Archive	35
White Rose Consortium ePrints Repository	195
<b>Sub Total</b>	<b>64162</b>
<b>Institutional subject archives</b>	
ECS EPrints Service	8735
Modern Languages Publications Archive	57
The Mathematical Institute Eprints Archive	138
PASCAL EPrints	671
Applied Computing Sciences ePrints Service	32
University of Cambridge - Teaching and learning	107

AKT EPrints Archive	165
<b>Sub Total</b>	<b>9905</b>
<b>Journal eprint archives</b>	
BBSPrints Archive	149
<b>Sub Total</b>	<b>149</b>
<b>International subject archives</b>	
BioMed Central	14930
Cogprints	2137
PsycPrints	720
ERP AePRINTS Service	35
IPv6 Eprints Archive	58
<b>Sub Total</b>	<b>17880</b>
<b>e-theses archives</b>	
Nottingham eTheses	44
<b>Sub Total</b>	<b>44</b>
<b>Total</b>	<b>92140</b>

**Table 1: Snapshot of number of records harvested by ePrints UK in January 2005**

Within the UK, repositories for learning objects are being addressed along a more centralised model, in particular within the JORUM repository and the HLSI and Curriculum Online services. Within institutions arguably VLEs could be considered as repositories for the management of learning objects, however it is difficult to track the deployment of VLEs within institutions.

Name	URL	Scope	N <sup>o</sup> of resources	Res. type	Other Notes
<b>Jorum</b>	www.jorum.ac.uk	National UK HE All subjects		Learning objects Metadata Any granularity	Currently setting up Jisc service.
<b>HLSI</b>	www.hlsi.org.uk	Regional Yorkshire Consortial KS4+ Subjects relevant to SMEs.	1 000s	Learning objects Any granularity	Higher level skills for industry
<b>Stòr Cùram</b>	www.storcuram.ac.uk	National UK HE Social Care		Multimedia digital learning resources	
<b>LORE</b>		Institutional strategic UK HE		Learning objects Teaching support materials	Learning Object Repository for Edinburgh Linked to strategic institutional e-learning programme

Name	URL	Scope	N° of resources	Res. type	Other Notes
<b>Curriculum Online</b>	www.curriculumonline.gov.uk	National UK School All subjects Rights expressed	"thousands" (>700 suppliers)	Metadata for resources useful in schools	Possibly the biggest repository of digital learning resources in the UK?

**Table 2: Selected UK learning object repositories and services**

Take-up and use of repositories thus far is limited but this may change should RCUK and/or institutions themselves mandate deposit of e-prints in institutional repositories. However, whilst this seems more likely with regard to e-print publications, it seems far less likely for research data and learning objects. Repositories for research data seem far more likely to be managed on a national level in some form or other, and to provide a subject rather than an institutional focus.

Key findings from interviews and the focus group indicate that where institutions do have repositories of one sort or another:

- They frequently experience difficulty in populating them. Encouraging and facilitating deposit of content can be time-consuming and requires a personal approach and much persuasion. Unless and until the creators of content can see the benefit of deposit in a repository then this problem is likely to continue.
- Where content is being deposited it tends to be from a particular department or subject area. Subject coverage is therefore patchy and incomplete and this may affect willingness to deposit, and usefulness in the longer term. Michael Day provides a useful overview of why this may be the case in his paper "Prospects for institutional e-print repositories in the United Kingdom" (Day, 2003) and Theo Andrew has reported an interesting case study of differences between disciplines in behaviour regarding 'self-posting' of research materials (Andrew, 2003).
- In many cases the full text of e-prints is not included and the content is only a metadata record describing the publication. The benefit of this kind of information for wider sharing in the community is somewhat doubtful.
- More often than not, the driving force behind the establishment of a repository is the library or the learning technology centre. It is unclear in many cases the extent of buy-in from senior management for these initiatives. Whilst this seems not to hinder the initial set-up of a repository, it may have a significant impact on the future sustainability of repositories beyond the period of JISC funding.
- Initiatives within institutions tend not to be joined up except in rare cases. For example, an e-prints repository will not be linked to a VLE to enable users of learning objects to also find relevant publications or the research data from which the object was derived.
- There are some concerns over the sustainability of open source software and the response to community requirements for improvements and new features.
- Existing software does not provide the full range of services and functions that those with repositories would wish to see. This may well act as a barrier to further development and the ability to meet user needs.
- There was common agreement that institutional metadata would be harvested from their repositories and presented via subject or other interfaces that related to 'real life' communities of practice. Some disappointment was expressed at the perceived lack of

progress of the ePrints UK project and it was suggested that this was a much needed service.

Key findings from interviews with institutions that do not currently have a formal repository infrastructure indicate that:

- There is concern about costs and long term sustainability of repository infrastructure
- There is concern about how to populate a repository, and how to ensure the creation of high quality metadata and content
- In some cases there is significant doubt about the value of establishing a repository, and a belief that any activity must relate to the objectives of the institution or meet a need from its staff and/or student body. Indeed one respondent expressed strong views about the lack of research into institutional and community requirements, and felt that the repositories agenda was being driven by a few very vocal advocates of the e-prints repository solution.
- In some cases a repository was seen as a useful way for sharing content within the institution but doubts were expressed about the benefits of sharing with the wider community
- Considerable concern was expressed about IPR and managing copyright

On the assumption that JISC would wish to continue to support the expansion of institutional repositories, several major issues need to be addressed, including:

- The need for support for institutions unable or unwilling to invest to create their own repository
- The need for support for those setting up repositories (perhaps along the lines provided by SHERPA)
- Technical developments and technical sustainability, including greater flexibility in specifying improvements to repository software to provide more services to users and depositors
- The need to embed repository development to meet the strategic aims of the institution – for example, to meet RAE requirements, to showcase research and teaching, or assist in the development of institutional learning strategies etc.
- The need to investigate complementary national/aggregator services for harvesting and re-presenting repository content
- Linking to other services or content held in other repositories

#### 5.2.2. Individuals or Groups of Individuals e.g. research groups

Lorna Campbell in a recent CETIS paper *Repository Issues from a Teaching and Learning Perspective*, 15/12/04 (see Annex 3): argued that

“...digital resources are situated within a more user centred culture where ownership, management and access control may be distributed across institutional roles (e.g. librarians, learning technologists, authors, teachers, researchers, learners, etc.) and communities of practice. Researchers, teachers and students are increasingly developing their own personal information management strategies, assuming control over whom they choose to share their resources with and adopting a wide range of informal tools and applications to support their communities of practice. Advocates of decentralised informal information management systems argue that they are better suited to facilitating a user centred approach to resource management and it is noticeable that discussions surrounding user controlled, decentralised environments

(e.g., peer-to-peer, wikis, ePortfolios) and centralised authority controlled systems (e.g., institutional repositories, digital libraries) are becoming increasingly polarised”.

It is certainly the case that these and other more informal methods of sharing papers, data and other resources have always existed, and will no doubt continue to exist. In addition, many researchers are creating customised websites of their own to either present research data (this is particularly prevalent in the arts and humanities), or with content for teaching purposes for their students to access; or as a personal website containing information about their careers, publications, interests etc. To this we might add virtual research environments (VREs) which are likely to become increasingly important for large-scale cross-institution and global research.

Several issues arise from this more informal landscape:

- These methods are an increasingly popular way of communicating and sharing – should they be supported and encouraged?
- How sustainable is this kind of infrastructure? Should it be tied in to a more managed framework, and if so, how?
- How might one select it for inclusion within a more managed framework?
- How might one better support this kind of activity in way that allows for a seamless transition into a more managed framework, should this prove necessary?

A key issue for JISC to consider is the need to regard repository activity not only as a technical and content problem – regardless of the nature of that content, but also to situate repository development in the practices and processes communities of practice who will be the depositors and users of the services and content. For example, learning objects should be commissioned from and presented to specific communities.

### 5.2.3. National Services and Organisations

A number of national services exist within the UK that form a vital part of the repository infrastructure. These include The National Archives, The British Library; The National Libraries of Scotland and Wales, Northern Irish Archives and Libraries, as well as a number of largely Research Council or government funded archives. Among these are the Economic and Social Data Service, the Arts and Humanities Data Service, CCLRC and the NERC Data Centres.

These Centres, Archives and Libraries all have an identified collecting remit and provide a range of services, including access, curation and preservation. Some, particularly those funded by the Research Councils, have a specific subject remit. All have, or are in the process of creating repositories to curate digital materials, some more complex than others. Within some centres a significant amount of expertise exists in the curation and preservation of digital materials from the simple to the more complex.

In addition there are a number of data stores that provide large capacity storage and preservation services, primarily for scientific data.

Many of the national services covered by this review have obligations, either legal deposit or from their funders, to acquire, curate and preserve, digital resources of all kinds, including data and texts. Issues arising at the focus group (which included staff from the BL and the National Library of Wales) and interviews included the need for certification for trusted digital repositories, some sense of what might be the appropriate relationship between national and local repositories, and how we might ensure no duplication of services; tools to automate or semi-automate repository processes; and the need to provide different search mechanisms and

interfaces tailored to 'communities of practice' to ensure the widest possible exploitation and use of the digital assets.

The focus group suggested that national services might be provided to support institutional repositories and/or to provide services (such as preservation) that institutions might find difficult or expensive. It was also suggested that a national service that could host a repository on behalf of an institution would be an option for those unable to provide their own service.

In addition, most institutional repositories saw national services as complementary, providing services and views they could not or did not want to. 'Real life' communities are interested in content not repositories, and national and international services are required whether from commercial or educational sector. Once again the need for progressing a national approach to services was expressed.

### **5.3. Cultural Issues and Gaps**

Whilst there are obvious and significant technical barriers and gaps, and requirements for further technical development, the primary challenge is effecting the necessary cultural change, and joining up the needs and requirements of the different actors involved in repository development, population and use.

The cultural issues must sit alongside a review of the life-cycle process of different types of digital assets, and a sense of when and how they might be brought into a more managed environment – and where that managed environment might best sit.

A clear vision of what a repository is, what it should do, and what services it might offer is by and large still unclear – policies still require more thought and work, and the overall landscape still requires much further thought and analysis.

There is some lack of knowledge (and understanding) of the culture in which others operate, with subsequent inability to tailor services and requirements accordingly.

There are still strong cultural and other barriers – from the research, teaching and learning communities, and from senior management in institutions. Little conversation seems to take place between those responsible for e-print repositories and e-learning repositories for example. How to embed repository practice within teaching practice, research practice, and institutional strategies requires further work.

Pedagogy and technology still do not easily sit side by side – there appears to be resistance to the use of technology, and some suspicion from the teaching community.

IPR and copyright issues are still a serious impediment to acquiring, preserving and sharing content.

### **5.4. Organisational Issues and Gaps**

The relationship between national, local and regional repositories (if any) is unclear, as is the relationship between subject and institution focused repository activity. Whilst this of itself need not be problematic, the potential for duplication, and repetition of services and development is high. More work on an 'ecology of repositories' would help here, as would a repositories framework. However, any work in this area should not only address the technical issues, but



should also look at functional issues (for example, how does the OAI model fit in), and work flows in an inclusive sense - both within repositories, and also how research, teaching and learning processes and work flows 'fit' with repository infrastructure.

More thought needs to be given to where personal collections of teaching materials/resources and personal research collections sit within the repository framework. Similarly the role and place of peer-to-peer, WIKI etc. solutions would benefit from an analysis of their position within the repository spectrum. If and how they might be supported, and how they might transfer into a more managed environment for further sharing and long-term sustainability of content.

Institutions are likely to have a range of repository like activities – perhaps an e-prints repository, a VLE, an administrative system for managing the RAE and so on. However, there seems to be little coordination between these various efforts. Encouraging 'joining up' of repository developments and interoperability within institutions, and positioning these to meet institutional strategic aims and objectives is essential.

Typically a number of distributed archives are established within institutions that embrace repositories as a solution, reflecting organisational politics and varying interests of disciplines. This is also a consequence of the phased introduction of repositories where different parts of the institution will have different priorities and different strategic requirements. Similar issues are faced by institutions establishing a common CMS across the institution.

In addition, the relationship between institutions and national services would benefit from further work and development. In particular the 'joining up' of content and services is a major area for development, and national services clearly have a vital role to play in focusing on 'communities of practice'.

There are gaps as regards coverage of particular subject areas – both in UK national research data archives, and within institutional archives. Within an institution, typically repository activity is led by individual enthusiastic departments that recognise the benefits of using a repository. Therefore coverage within institutions across disciplines is uneven, with most institutions having input from only a few departments. Should institutions consider providing 'departmental' repositories rather than one monolithic institutional repository (this could of course physically be one repository with departmental interfaces), and focus their activities on those subjects and areas that are most willing and able to provide content?

## **5.5. Technical Issues and Gaps**

Software and tools to support repository activity are still at a relatively early stage of development. Many of these products are open source, which presents issues of coordinated development and sustainability. Concern was expressed at the long-term sustainability of some of these tools. JISC needs to address this issue if institutions and organisations using repository software are to be confident in repository development and embedding that within a service environment.

Lack of functionality in existing products limits the range of services that repositories might offer. Indeed many of the national services have developed their own in-house solutions in order to bypass this lack of functionality. Whilst this may be an option for larger organisations, it is unlikely to be possible in the institutional repository environment as the resource required would be beyond the scope of most libraries.

Priority developments include 'smart tools' for data extraction and automatic classification; format conversion tools improvements in the user interface for submissions, including metadata and submission across multiple repositories; tools to manage interdisciplinary exploitation of repositories; and last but by no means least, a repository module to support the RAE process.

Among the other issues highlighted were the lack of provision for preservation actions; the need to extend OAI compliancy and interoperability between systems to share metadata and digital objects; integration with existing tools used by authors to create and manage their digital outputs; the need for easier installation, configuration and use; automated ingest and processing; provision of a flexible rendering environment and automated testing of whether a digital object will be properly reproduced in such an environment; and a standardised way of collecting and comparing download statistics.

Institutions are often conservative as regards linking to full text of e-prints. This is based on caution as regards copyright conflicts with journals, and inter-institution IPR issues. This means repositories often restrict links to those resources produced whilst staff are employed at the institution. The result is that significant percentage of repository entries are 'metadata only.'

There are real concerns about the quality of metadata and content deposited in repositories – relying on content creators to create the necessary metadata, and even to supply high quality content does not seem a satisfactory solution.

## 5.6. Sustainability Issues and Gaps

There is a growing body of information on cost models for establishing institutional repositories, as illustrated within the Leadirs seminars (LEADIRS, 2004). However, more work needs to be done in this area taking into account on-going maintenance costs and long-term sustainability, and undertaking cost/benefit analysis. For example, we know little about whether the benefits to institutions are cost-effective. There may be a different answer for research led institutions than for learning institutions.

Similar questions arise for national and international services. For example, what are the governance models for national repositories, can such repositories ensure sustainability? What should be the role of JISC in contributing to the sustainability of such repositories? Should these be provided on a value-added, charged for basis? Who takes responsibility for services built on repositories?

Within the distributed environment how might JISC ensure that 'communities of practice', including subject communities, receive the services and interfaces they require in order to do their jobs more effectively? How might national services join-up with international services to provide subject coverage? What are the roles of the Research Councils, JISC, institutions, BL and national libraries? And who should fund what?

## 6. Concluding Recommendations

**More work needs to be done to identify, specify and map the repository landscape. The framework should include not only technical issues but also processes and functions. JISC should consider this as a parallel JISC/DEST activity. This work should include workflows and business processes, and the relationship between national and local, subject and other types of repository.**

**In addition to the above, JISC should consider what might best be provided by a national service (e.g. akin to JORUM for learning objects) and what might be best done at an institutional level. The Digital Rights Management requirements of different organisational models might be a key element of any study in this area.**

**Repository developments should demonstrably be set within the strategic aims of the host institutions or funding bodies and clearly relate to the strategic aims and objectives of the organisation bidding for funds – buy-in from institutional and other senior management must be assured for future sustainability. For example, JISC should consider funding projects seeking to use e-print repositories to support the RAE process.**

**Repository developments should, depending upon their primary focus, relate to the processes and practices of research, teaching or learning – buy-in from the community is unlikely to be extensive unless this happens. JISC should identify current practice of researchers, teachers and learners, and seek to base services on supporting their needs.**

**Support for the research infrastructure should be undertaken in collaboration with the Research Councils and in particular, with the RCUK Digital Curation and Archiving Working Group.**

**Repository interfaces should be directed towards ‘communities of practice’ and more effort should be made to tailor services for specific user communities, rather than producing generic interfaces. This is likely to be particularly important to encourage take-up within FE, although it is still important for HE.**

**Continued support will be required for establishing institutional repositories – JISC may wish to consider funding a co-ordinating ‘focus’ initiative to provide technical support and support with policy and advocacy issues.**

**A different kind of support may be required for those institutions unable to provide their own repository infrastructure. JISC may want to consider providing a national service that smaller and less well funded institutions could use to provide repository services and functions on their behalf. JISC might wish to fund the start-up costs of such a service with a view to such a service becoming self-sustaining in the longer term. We recommend a scoping study be undertaken to assess the need and the costs of such a service.**

**As more and more content becomes available it will become increasingly important to join-up content held in different places. JISC may want to consider funding projects that seek to find and link content held in different types of repository e.g. e-prints with data; learning objects with publications, and to investigate the challenges posed.**

**JISC should consider funding further technical development to support the provision of additional repository services. These could include ‘smart tools’ for automatic data extraction, automatic classification etc., format conversion tools, improved input and export functionality, and facilities to create multiple interfaces for different kinds of users.**

**JISC should also address the issue of sustainability of repository software.**

**JISC should consider further work on IPR and authentication/authorisation mechanisms that would allow some content to be widely shared, and others to be available to more**

**limited groups e.g. students studying on a particular course, or colleagues working on a research project.**

**JISC should consider investigating how more informal networks for sharing content and pre-prints might be supported, and mechanisms for incorporating shared content into a more managed repository framework at some point in the lifecycle.**

## **7. References**

Andrew, T. (2003) "Trends in Self-Posting of Research Material Online by Academic Staff": 30- In Ariadne, Issue 37, October 2003.

<http://www.ariadne.ac.uk/issue37/andrew/intro.html>

Baldwin, C. "X4L review: final report". July 2004.

[www.jisc.ac.uk/uploaded\\_documents/X4L%20Review%20-%20Final%20Report.doc](http://www.jisc.ac.uk/uploaded_documents/X4L%20Review%20-%20Final%20Report.doc)

Blinco, K., Mason, J., McLean, N., Wilson, S. (2004). "White paper for alt-i-lab 2004".

Prepared on behalf of DEST (Australia) and JISC-CETIS (UK). Version 2, 19 July 2004.

[http://www.jisc.ac.uk/uploaded\\_documents/Alttilab04-infrastructureV2.pdf](http://www.jisc.ac.uk/uploaded_documents/Alttilab04-infrastructureV2.pdf)

Burnhill, P., Halliday, L., Rozenfeld S., Kidd T., (2004). "SUNCAT: a modern serials union catalogue for the UK." In *Serials: the Journal for the Serials Community*, Volume 17, Number 1, March 2004 pp. 61-67.

Crow, R. (2002). "The Case for Institutional Repositories: A SPARC Position Paper." SPARC, 2002. [http://www.arl.org/sparc/IR/IR\\_Final\\_Release\\_102.pdf](http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf)

Campbell, L., Blinco, K., Mason J. (2004). "Repository management and implementation."

Prepared on behalf of DEST (Australia) and JISC-CETIS (UK), July 2004.

[http://www.jisc.ac.uk/uploaded\\_documents/Alttilab04-repositories.pdf](http://www.jisc.ac.uk/uploaded_documents/Alttilab04-repositories.pdf)

Conway, Paul. (2004). "Institutional repositories: is there anything left to say?" Given as presentation within OCLC Distinguished Seminar Series, October 7 2004. Powerpoint and Audio from <http://www.oclc.org/research/dss/>

Day, M. (2003). "Prospects for institutional e-print repositories in the United Kingdom." ePrints UK supporting study, no. 1, 28 May 2003.

<http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/impact/>

Dempsey, L., Childress, E., Godby, C.J., Hickey, T. B., Vizine-Goetz, D., Young, J. (2004).

"Metadata Switch: thinking about some metadata management and knowledge organization issues in the changing research and learning landscape." Forthcoming in *LITA Guide to E-Scholarship* [working title], ed. Debra Shapiro. August 2004 preprint available at

<http://www.oclc.org/research/publications/archive/2004/dempsey-mslitaguide.pdf>

Dempsey, L. (2005) "Library collections and services - some pictures." Extract from Lorcan Dempsey's Weblog, February 17 2005. <http://orweblog.oclc.org/archives/000555.html>

"Digital Library Repositories and Instructional Support Systems." Repository Interoperability Working Group, DLF, 2004.

[http://www.diglib.org/forums/Spring2004/Johnston0404\\_files/frame.htm](http://www.diglib.org/forums/Spring2004/Johnston0404_files/frame.htm)

Flecker, D., and McLean, N. (2004) "Digital Library Content and Course Management Systems: Issues of Interoperation: Report of a study group." Digital Library Federation, July 2004. <http://www.diglib.org/pubs/cmsdl0407/>

Heery, R., Duke, M., Day, M., Lyon, L., Coles, S., Frey, J., Hursthouse, M., Carr, L., and Gutteridge, C. (2004) "Integrating research data into the publication workflow: eBank experience" In *Proceedings PV-2004: Ensuring the Long-Term Preservation and Adding Value to the Scientific and Technical Data, 5-7 October 2004, ESA/ESRIN, Frascati, Italy*. ESA WPP-232. Noordwijk: European Space Agency, 2004, pp. 135-142. <http://www.ukoln.ac.uk/projects/ebank-uk/dissemination/PV2004-heery.pdf>

House of Commons, Select Committee on Science and Technology. (2004). "Responses to the Committee's tenth report, Session 2003-04, scientific publications, free for all?" HC 1200. London: The Stationery Office. <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/1200/120011.htm#a6>

IMS (2003). "IMS Digital Repositories v1.0 Final specification." January 2003. <http://www.imsglobal.org/digitalrepositories/index.html>

Jerez, H.N., Xiaoming, L., Hochstenbach, P., Van de Sompel, H. (2004) "Repository architectures: the multi-faceted use of the OAI-PMH in the LANL repository." In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries, Tuscon, USA, 2004*. ACM, 11-20. <http://public.lanl.gov/herbertv/papers/jcdl2004-submitted-draft.pdf>

Kraan, W. (2005). "Interoperability state of play at IMS Melbourne meeting" February 13, 2005. Report on CETIS web site. <http://www.cetis.ac.uk/content2/20050213085236>

Lavoie, B. F. (2004). "The Open Archival Information System Reference Model: introductory guide." DPC Technology Watch Series Report 04-01, 2004. [http://www.dpconline.org/docs/lavoie\\_OAIS.pdf](http://www.dpconline.org/docs/lavoie_OAIS.pdf)

LEADIRS (2004). LEarning About Digital Institutional Repositories Seminars (), November 2004. [http://www.lib.cam.ac.uk/leadirs/LEADIRSII\\_presentations.htm](http://www.lib.cam.ac.uk/leadirs/LEADIRSII_presentations.htm)

Lynch, C. (2003). "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL*, 226, February 2003, 1-7. <http://www.arl.org/newsltr/226/ir.html>

Lyon, Liz. (2003) "eBank UK: Building the links between research data, scholarly communication and learning" In *Ariadne* 36, July 2003. <http://www.ariadne.ac.uk/issue36/lyon/intro.html>

McLean, N. "The Ecology of repository services: a cosmic view". A keynote presentation at *Research and Advanced Technology for Digital Libraries: 8th European Conference, ECDL 2004, Bath, UK, September 12-17, 2004*. Powerpoint available from <http://www.ecdl2004.org/presentations/mclean/>

McLean, N., and Lynch, C. (2004). "Interoperability between library information services and learning environments – bridging the gaps." A joint white paper on behalf of the IMS Global learning consortium and the Coalition for Networked Information, May, 2004. [http://www.imsglobal.org/digitalrepositories/CNlandIMS\\_2004.pdf](http://www.imsglobal.org/digitalrepositories/CNlandIMS_2004.pdf)

NDIIP Technical Architecture Version 0.2. (2003?)  
[http://www.digitalpreservation.gov/repor/NDIIPP\\_v02.pdf](http://www.digitalpreservation.gov/repor/NDIIPP_v02.pdf)

Payette, Sandra et al. (1999). "Interoperability for Digital Objects and Repositories: the Cornell/CNRI Experiments." In D-Lib Magazine, Volume 5 Issue 5, May 1999.  
<http://www.dlib.org/dlib/may99/payette/05payette.html>

Payne, G. (2005) "ARROW institutional repositories: a report on the decisions and experiences of the ARROW project." Presentation to the ALIA Information Online Conference, Sydney, 1 February 2005.  
[arrow.edu.au/docs/files/ARROW%20Presentation%20Info%20Online%201%20Feb%202005%20final.pdf](http://arrow.edu.au/docs/files/ARROW%20Presentation%20Info%20Online%201%20Feb%202005%20final.pdf)

Peters, T. A. (2002) "Digital repositories: individual, discipline-based, institutional, consortial, or national?" The Journal of Academic Librarianship, Volume 28, Issue 6, November-December 2002, 414-417.

Rogers, S. A. (2003) "Developing an institutional Knowledge Bank at Ohio State University: from concept to action plan." Libraries and the Academy Series/Report no: vol. 3, no. 1, 2003, 125-136. <http://hdl.handle.net/1811/188>. <https://kb.osu.edu/dspace/handle/1811/188>

Sufi, S., Matthews, B. (2004) "CCLRC scientific metadata model: version 2" CCLRC DL Technical Reports, DL-TR-2004-001, 2004.  
<http://epubs.cclrc.ac.uk/bitstream/485/csmdm.version-2.pdf>

Swan, Alma *et al.* (2004). "Delivery management and access model for e-prints and open access journals within further and higher education." Electronic Publishing Innovation Centre (EPIC) in partnership with Key Perspectives Limited, 2004.  
[http://www.jisc.ac.uk/uploaded\\_documents/ACF1E88.pdf](http://www.jisc.ac.uk/uploaded_documents/ACF1E88.pdf)

Tourte, G. L. (2005). "ePrints UK technical documentation." ePrints UK paper, 2005.  
<http://www.rdn.ac.uk/projects/eprints-uk/docs/technical/eprints-tech-report>

RLG. (2002). "Trusted digital repositories: attributes and responsibilities." An RLG-OCLC Report, 2002. <http://www.rlg.org/longterm/repositories.pdf>

UK Research Councils (1998). "Safeguarding good scientific practice: a joint statement by the Director General of the Research Councils and the Chief Executives of the UK Research Councils." Issued on 18 December 1998. <http://www.ost.gov.uk/research/councils/safe.htm>

Ware, M. (2004) "Publisher and Librarian/Learning Solutions (PALS): Pathfinder research on Web-based repositories: final report." PALS, January 2004.  
<http://www.palsgroup.org.uk/palsweb/palsweb.nsf/>

Yanosky, R., Harris, Marti; Zatrocky, M. "Hype cycle for higher education." Gartner, 2004.