

✦ Background

The Paradigm Workbook is an evolving resource based on an exemplar project at the academic research libraries of the Universities of Oxford and Manchester. Between January 2005 and February 2007, the Paradigm project explored the issues involved in the long term preservation of personal digital archives using today's politicians and their personal archives as a testbed.

Why personal archives?

Despite the exponential growth in digital content being created by individuals, there is a marked lack of research and development dedicated to the preservation of this kind of content. To date, digital preservation projects have tended to be sponsored by corporate bodies or state archives, where legal, financial and organisational drivers have provided an impetus to develop policies, procedures and systems designed to preserve digital objects deemed to be of value to the organisation. These developments are welcome, but working with individuals to preserve their personal digital archives presents different challenges which require different solutions.

Project methodology

The project team investigated various tools, models, software, standards, metadata, strategies, policies and procedures associated with preserving digital objects. During the course of the project these have been mapped to traditional archival workflows and the particular circumstances of institutions collecting personal digital archives.

Audience

It is hoped that the Workbook will be helpful to all that are interested in the preservation of digital objects with long-term value. Its principal audience is organisations, of any flavour, which care for the personal archives of politicians, scientists, writers, journalists, academics or of other individuals. Personal archives illuminate the history inscribed in official records; developing the capacity to collect, preserve and provide long-term access to such materials is critical if institutions, such as academic research libraries, are to continue fulfilling their roles as humanity's remembrancers.

The authors

The principal authors of the Workbook are the project team: Susan Thomas (Project Manager and Digital Archivist, Oxford), Renhart Gittens (Software Engineer, Oxford), Janette Martin (part-time Digital Archivist, Manchester) and Fran Baker (part-time Digital Archivist, Manchester).

The web version of the Workbook

The Paradigm Workbook was originally published cumulatively as a web document and was conceived to record the lessons learned by the project team. The Online Workbook contains some additional materials of a more practical nature, which are referenced from this print edition. References to the Online Workbook take the following form: Online Paradigm Workbook. URL: <<http://www.paradigm.ac.uk/workbook/url.html>>. The Online Workbook contains a Glossary of relevant terms;¹ for each underlined word in the text of this print edition there is a corresponding entry in the Online Workbook Glossary. Readers should also be aware that the Workbook makes references to several online resources; links to these resources were all valid on 7 September 2007, though it is expected that access to some of these resources may disappear over time.

Funding

The funding which has made this project possible came from the JISC's Institutional Digital Preservation and Asset Management Programme;² and from the University of Oxford's Research Development Fund (now the John Fell OUP Research Fund).

1 Online Paradigm Workbook. URL: <<http://www.paradigm.ac.uk/workbook/glossary/index.html>>

2 JISC, 'Digital Preservation and Asset Management Strand' (1 September 2006), *Joint Information Systems Committee website*. URL: <http://www.jisc.ac.uk/whatwedo/programmes/programme_preservation/programme_404.aspx>

Acknowledgements

The project team would like to thank the politicians and their staff, who made this project possible by contributing their time and archival materials. We are also grateful to those that provided input to the project and the Workbook, including the following: Paul Bevan (National Library of Wales), Stephen Bird (Labour history Archive and Study Centre), Stella Butler (John Rylands University Library), Chris Fletcher (Bodleian Library), Richard Gartner (Oxford Digital Library), Stella Halkyard (John Rylands University Library), Sarah Higgins (Digital Curation Centre), Amanda Hill (Archives Hub), Helen Hockx-Yu (JISC), John Hodgson (John Rylands University Library), Jeremy John (British Library), Helen Langley (Bodleian Library), Charlotte McKillop-Mash (Bodleian Library), Matthew Neely (Bodleian Library), Richard Ovenden (Bodleian Library), Tim Padfield (The National Archives), James Peters (John Rylands University Library), Michael Popham (Oxford Digital Library), Glen Robson (National Library of Wales), Jane Stevenson (Archives Hub), Emily Tarrant (Bodleian Library) and Dave Thompson (Wellcome Library). Thanks go too to the members of Paradigm’s academic advisory Board: Lawrence Goldman (University of Oxford), Simon Bailey (Oxford University Archives), Martin Conway (University of Oxford), John Davis (University of Oxford), Steven Fielding (University of Salford), Alex May (University of Oxford), Kevin Morgan (University of Manchester) and to the following organisations: Digital Curation Centre, Digital Preservation Coalition, the Fedora community and JISC Legal.

Arrangement of the Workbook

The Workbook is loosely arranged around the lifecycle of a personal digital archive which was assembled by the Paradigm project as a means of fusing the ideas and standards relating to digital curation and preservation with established archival practice.

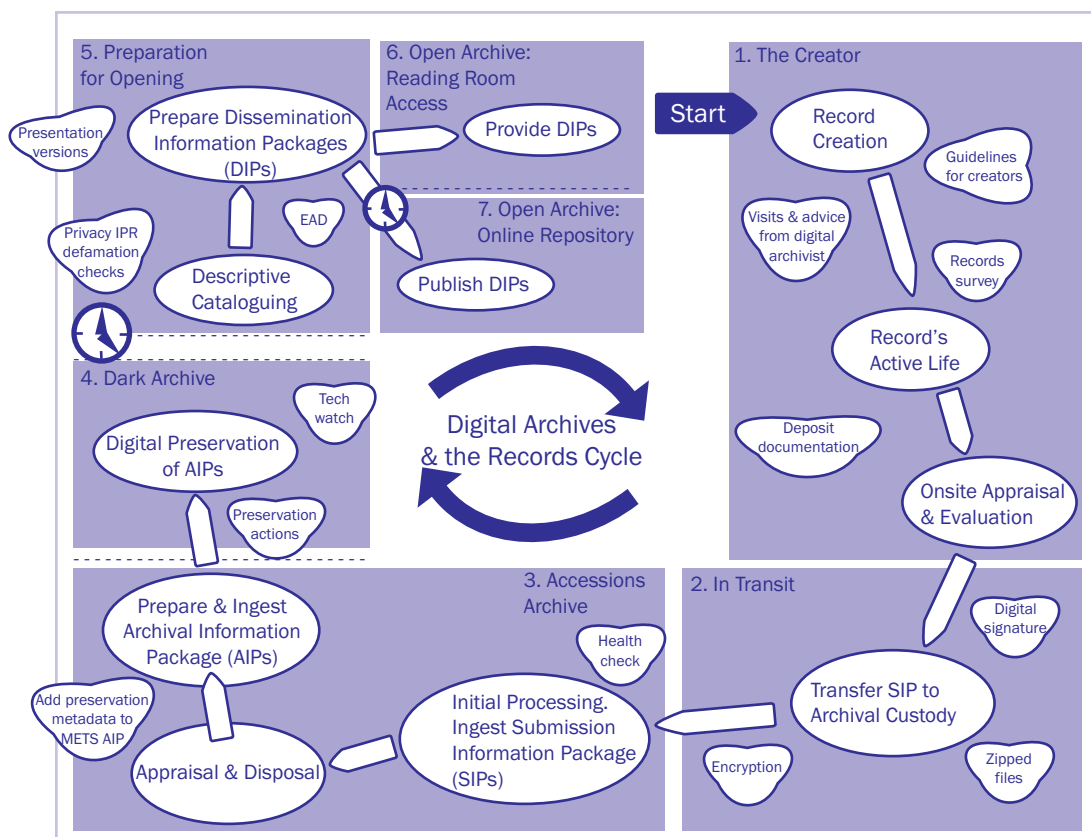


Figure 1: Digital Archives & the Records Cycle

The Paradigm lifecycle moves through seven key stages from interactions with those creating archival materials; to effecting authentic and secure transfer of materials to archival custody; to processing new accessions and preparing them for archival storage; to preservation monitoring and actions while materials are in archival storage; to preparing materials for researcher access in a reading room; and finally to publishing materials to online digital repositories once rights have

expired. Record management concepts have played an important role in the development of this lifecycle as have the concepts of the *Reference Model for an Open Archival Information System* (ISO 14721:2003). In recognition of the importance of the OAIS model to all areas of Paradigm's work and this Workbook, a short introduction to the model is provided here.

✧ Introduction to OAIS

Introduction

What is the OAIS model and why should I use it?

ISO 14721:2003 defines an OAIS as:

“an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community.”

Where

“The information being maintained is deemed to need Long Term Preservation, even if the OAIS itself is not permanent. Long Term is long enough to be concerned with the impacts of changing technologies, including support for new media and data formats, or with a changing user community”.

The Open Archival Information System, usually referred to as the OAIS model, is a reference model that has been widely accepted by the digital preservation community as a key standard for digital repositories. The OAIS model specifies how digital assets should be preserved for a community of users from the moment digital material is ingested into the digital storage area, through subsequent preservation strategies to the creation of a dissemination package for the end user. The OAIS reference model is a high-level reference model, which means it is flexible enough to use in a wide variety of environments. The down side of this flexibility is that in some areas, for example ingest, more detailed steps and workflows will have to be developed by the implementing institution. The OAIS model was adopted as an ISO standard in 2003 (ISO 14721:2003 OAIS).

OAIS: a common language for the information professionals

Although designed by space data curators, the OAIS model aims to be as context-neutral as possible. OAIS deliberately avoids jargon from both the IT and archival professions; this is very useful as it makes both groups speak the same language. Once acquired, the terms and language of the OAIS model enable the digital curator to communicate effectively with other national and international projects. The downside is that the jargon can deter those not yet immersed in ‘OAIS speak’ and act as a barrier to understanding and cooperation. The complexity of the OAIS reference model has led some practitioners to call for an ‘OAIS lite’ which would make the model more accessible to smaller and less well funded institutions.

Besides supplying a common framework and vocabulary, the OAIS model serves as a planning tool for designing new digital repositories and a benchmark for evaluating the capabilities of more established digital repositories.

Overview of the OAIS model: Environment (OAIS 2.1)

The OAIS has relationships with three entities (see diagram on next page):

- Producers, which deliver material to the OAIS.
- Consumers, which obtain material from the OAIS.
- Management, which is responsible for managing the OAIS.

The actors in each entity may be human, machine, or both.

A special class of Consumers is the OAIS' Designated Community. The Designated Community is the OAIS' primary user group; the information supplied by the OAIS should be understandable by this group; this means that the OAIS must have an appreciation of the community's Knowledge Base.

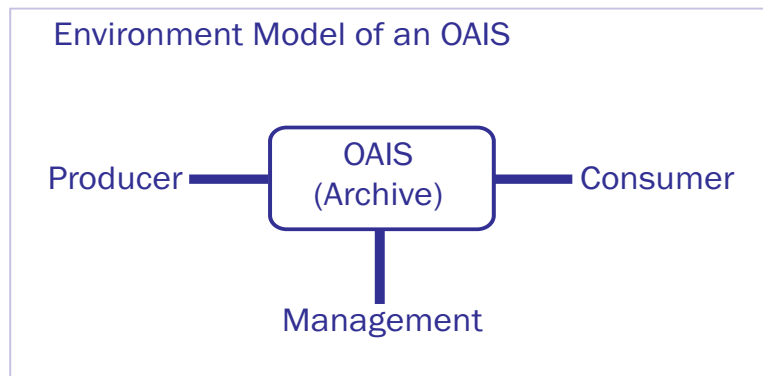


Figure 2: Environment Model of an OAIS

Overview of the OAIS model: Information model (OAIS 2.2)

In addition to defining the parties involved in the long-term preservation of digital materials, OAIS provides an information model for managing the digital materials as they pass through the system. A significant component of this model is the Information Package (IP). Each IP consists of:

1. The digital object(s) to be preserved.
2. The metadata required at that point in the system.
3. The Packaging Information which relates 1 and 2.

OAIS outlines three types of Information Package:

- Submission Information Packages (SIPs).
- Archival Information Packages (AIPs).
- Dissemination Information Packages (DIPs).

SIP

At the SIP stage, the metadata accompanying the digital object is, ideally, supplied by the Producer who is generally the original creator of the material; in the case of personal archives it is perhaps more likely that a digital archivist working with the creator will provide the metadata. At this stage, the metadata will probably lack structure and may not be comprehensive at all levels of the archive. SIPs may also be supplied to an OAIS from another digital repository. Where another digital repository has supplied SIPs, the use of interoperable metadata standards will minimise the effort required to ingest the material into the new repository.

AIP

At the AIP stage, the SIPs are prepared for preservation. During this process, the digital materials submitted for preservation are known as Content Data Objects and they are combined with the Preservation Description Information (PDI, see OAIS 4.2.1.4.2) needed to administer their preservation. OAIS breaks the PDI down into four sections:

- Reference Information: a unique and persistent identifier.
- Provenance Information: the history of the archived object.
- Context Information: relationship to other objects, e.g. the hierarchical structure of a digital archive.
- Fixity Information: a demonstration of authenticity, such as a hash value.

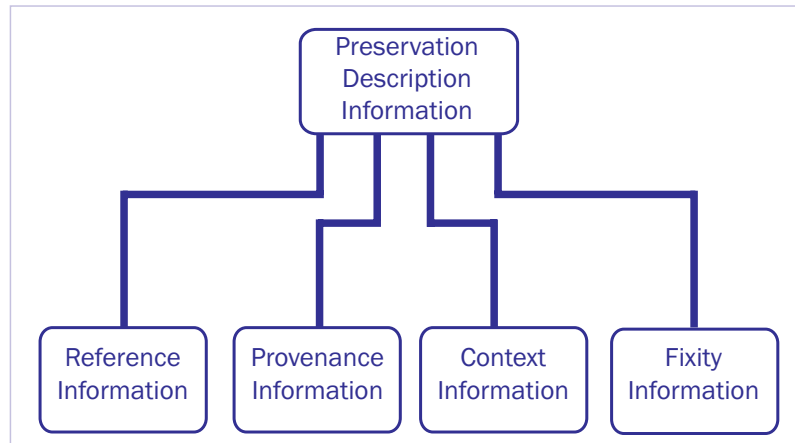


Figure 3: Preservation Description Information

OAIS also requires the archive to maintain the Representation Information required to render the object intelligible to its designated community. This might include information regarding the hardware and software environment needed to view the content data object.

Detailed information on the AIP can be found in OAIS 4.2.2.3.

DIP

The DIP stage happens when a user (or Consumer in OAIS terms!) requests a digital object or group of objects from the OAIS. The OAIS supplies the object(s) packaged as a DIP comprising the object and relevant metadata. It is likely that the metadata accompanying the object at this stage will be more descriptive than technical; the end user is unlikely to want to see complicated metadata relating to fixity or representation. In all instances, the Knowledge Base of the OAIS' Designated Community will guide the type and extent of metadata supplied.

METS and Information Packages

The METS (Metadata Encoding and Transmission Standard, see p. 117) standard was designed to fulfil the OAIS 'Information Package' concept. METS can be implemented in many ways, but the key consideration for preservation repositories is that METS provides a means of associating all the metadata pertaining to an object, including that object's relationships with other objects, with the object.

Representation Information

Representation Information is an important part of the OAIS information model. Representation Information is used to interpret a Data Object so that it can be rendered into a useful Information Object. Representation Information permits the expression of dependencies within a digital archive. Since many repositories contain broadly similar objects, central Representation Information registries that can be used by several repositories are being developed; one example is the Digital Curation Centre's Representation Information Registry¹ (under development). See OAIS 2.2.1 and 4.2.1.3 for further information.

Descriptive Information

OAIS also includes descriptive metadata as part of its information model. Descriptive metadata is not part of the Information Package, rather it describes the package.

Overview of the OAIS model: Functional model (OAIS 4)

OAIS provides a functional model, which includes information about the kinds of activities undertaken by each function. The seven functions outlined by OAIS are:

¹ Digital Curation Centre, *Representation Information Registry Repository website*. URL: <<http://registry.dcc.ac.uk/omar/>>

01 Introduction

- Ingest.
- Archival Storage.
- Data Management.
- Administration.
- Access.
- Preservation Planning.
- Common Services.

A useful diagram of the seven OAIS functions and the transition from SIP to AIP to DIP is given below:

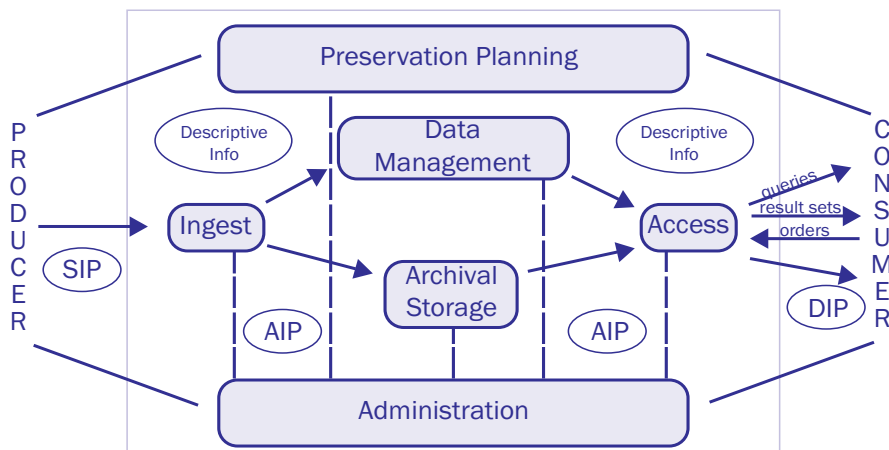


Figure 4: Seven OAIS Functions

Most of these functions will make sense to an archivist or information professional. Ingest is roughly the same as the traditional archival processes undertaken when an archive is newly accessioned and before it is added to the main collection. In the paper world this would involve checking for mould, insects and physical degradation as opposed to the digital equivalent of worms, viruses and bitrot. The meaning of Common Services may also require some explanation: these are the essential services required by any IT system, such as the timely application of security patches.

Overview of the OAIS model: Responsibilities (OAIS 3)

The responsibilities required by the OAIS standard are:

- Negotiate and accept information from Producers.
- Determine which community should become the Designated Community.
- Ensure that Information Packages are independently understandable.
- Ensure Information Packages are preserved.
- Make preserved Information Packages available.

Once the unfamiliar language is grasped, it is clear that these are tasks and responsibilities routinely undertaken by archival institutions in the management of paper records. What's new about the OAIS standard is the detailed intellectual framework, which articulates the policies, functions and metadata crucial for meeting such responsibilities in a digital environment, although the model can be applied to any kind of data, even a moon rock specimen!

OAIS in practice: Some examples

The OAIS concept of ‘Designated Community’ permits a diverse range of groups and institutions collecting digital records for their own purposes to use the model for their own needs. To date various sectors have employed the OAIS model. Below are a few of the more high profile digital initiatives based on OAIS concepts.

LOCKSS

The open source LOCKSS¹ system (Lots of Copies Keep Stuff Safe) developed by Stanford University Libraries is a good example of how the OAIS model can work in practice. Stanford’s system works closely with publishers and enables participating libraries to own copies of the e-journals that they subscribe to as well as content published in open-access e-journals. Each library in the LOCKSS network hosts an inexpensive machine running LOCKSS software, which crawls the websites of e-journal publishers to collect content; by talking to LOCKSS machines at partner libraries, the software engages in a peer-to-peer audit that identifies and repairs any corrupted content using a voting system. LOCKSS only provides libraries with access to content that they have paid for, so it satisfies the needs of publishers.

Institutional repositories

Another common type of repository is the institutional repository, designed to store, manage and sometimes preserve digital assets of institutions. An increasing number of universities are establishing these as a means of managing the digital research output and administrative records generated by a university. Repositories for archival material could operate within the framework of an organisation’s institutional repository, but as the area is currently dominated by the needs of academic outputs and the drive for Open Access, archivists need to take a lead in formulating infrastructures, policies and procedures that are appropriate to the requirements of potentially sensitive archival material.

E-theses

Electronic-theses repositories are another good example of an OAIS. A leading project in this area is the UK ETHOS (Electronic Theses Online Service) project,² which aims to develop a prototype online e-theses service, which will be easily scalable and financially sustainable.

Centralised repositories

There are also centralised, often discipline-oriented, research data repositories, such as those managed by the Arts and Humanities Data Service (AHDS).³

OAIS in practice: Compliance & certification

How does your institution rate?

As the OAIS Reference Model gains significance, institutions which have managed digital media for decades may wish to see how they measure up. Such an exercise is useful for ongoing planning, for raising an institution’s profile within the archive community and ensuring that the same degree of professional care is applied to digital archives as to their analogue equivalents. Two institutions in the UK have undertaken this exercise: The National Archive’s ‘Digital Archive’ and the National Digital Archive of Datasets (NDAD). The result of this OAIS compliance investigation has been written up into a useful report.⁴

A model for digital repository software

OAIS also serves as a framework for developers of digital repository software; such repositories include DSpace and Fedora, which Paradigm has tested (see p. 204). The impact of the OAIS model is apparent in the existing versions of DSpace and Fedora, and in the current development priorities of the two communities.

1 LOCKSS, *Lots of Copies Keep Stuff Safe website*. URL: <<http://www.lockss.org/>>

2 ETHOS, *Electronic Theses Online Service website*. URL: <<http://www.ethos.ac.uk/>>

3 AHDS, *Arts and Humanities Data Service website*. URL: <<http://ahds.ac.uk/>>

4 Hilary Beedham, et. al., *Assessment of UKDA and TNA Compliance with OAIS and METS Standards* (Essex: UK Data Archive, 2005). URL: <http://www.jisc.ac.uk/uploaded_documents/oaismets.pdf>

Certification

There is currently no formal certification mechanism for digital repositories, though some assessment of digital archiving competency is available through existing mechanisms, such as the inspection service provided by the Advisory Service at The National Archives¹ based on its *Standard for Record Repositories* document.² There is ongoing debate in the digital preservation community about the merits of a formal certification process; many are concerned about the cost and unconvinced of the value, and most are unsure how certification will work in practice. Despite these uncertainties, certification initiatives are already emerging. Combining the strengths of the OAIS model with the RLG/OCLC's *Trusted Digital Repositories: Attributes and Responsibilities* (2002),³ a RLG and NARA task force published an audit checklist for digital repository certification⁴ in 2005; an updated checklist entitled *Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist* was published in February 2007.⁵ There is also a certification process for German repositories administered by the Deutsche Initiative für Netzwerk Information (DINI).⁶ The DRAMBORA⁷ (Digital Repository Audit Method Based on Risk Assessment) toolkit has also been published by the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE).

✦ Useful resources

OAIS

Consultative Committee for Space Data Systems, *Reference Model for an Open Archival Information System (OAIS). Blue Book (January 2002)*.

URL: <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>

Beedham, Hilary, et. al., *Assessment of UKDA and TNA Compliance with OAIS and METS Standards* (Essex: UK Data Archive, 2005).

URL: <http://www.jisc.ac.uk/uploaded_documents/oaismets.pdf>

Holdsworth, David and Sergeant, Derek M, 'A Blueprint for Representation Information in the OAIS model', *Cedars Project, University of Leeds website*.

URL: <<http://www.personal.leeds.ac.uk/~eclhdh/cedars/ieee00.html>>

Lavoie, Brian F., 'The Open Archival Information System Reference Model: Introductory Guide', *DPC Technology Watch Series Report, 04-01 (January 2004)*.

URL: <http://www.dpconline.org/docs/lavoie_OAIS.pdf>

Note: A good introduction to the OAIS model.

Lynch, Clifford, 'Institutional Repositories: Essential infrastructure for Scholarship in the Digital Age', *ARL Bimonthly Report, 226 (February 2003)*.

URL: <<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>

1 The National Archives, 'National Advisory Services', *The National Archives website*. URL: <<http://www.nationalarchives.gov.uk/archives/>>

2 The National Archives, *Standard for Record Repositories* (2004). <<http://www.nationalarchives.gov.uk/documents/standard2005.pdf>>

3 Research Libraries Group and Online Computer Library Center, *Trusted Digital Repositories: Attributes and Responsibilities*. (RLG, May 2002). URL: <<http://www.rlg.org/en/pdfs/repositories.pdf>>

4 Research Libraries Group, 'Audit Checklist for Certifying Digital Repositories', *Research Libraries Group website*. URL: <http://www.rlg.org/en/page.php?Page_ID=20769>

5 RLG-NARA Digital Repository Certification Task Force, *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)*, (OCLC and CRL, February 2007). URL: <<http://bibpurl.oclc.org/web/16712>>

6 DINI, *Deutsche Initiative Für Netzwerkinformation E.V. website*. URL: <<http://www.dini.de/>>

7 Digital Curation Centre and Digital Preservation Europe, *Digital Repository Audit Method Based on Risk Assessment website*. URL: <<http://www.repositoryaudit.eu/>>

RLG-NARA Digital Repository Certification Task Force, *Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)*, (OCLC and CRL, February 2007).

URL: <<http://bibpurl.oclc.org/web/16712>.>

Research Libraries Group and National Archives Records Administration, *An Audit Checklist for the Certification of Trusted Digital Repositories: Draft for Public Comment* (RLG, August 2005).

URL: <<http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>>

Key digital preservation directories

There are many useful sources of information about digital preservation available online. To avoid duplicating the efforts of other organisations, we have limited this list to key directories which are an excellent point of departure for discovering more about digital preservation activities in the UK, Europe and internationally.

DPC - Digital Preservation Coalition (<http://www.dpconline.org/>)

The DPC was established in 2001 to encourage collaborative effort in addressing the challenge of preserving digital resources in the UK. In addition to its advocacy and education work, the DPC conducts surveys, publishes reports and organises forums on topics of interest. The DPC's website includes a directory of projects underway at member institutions and hosts the online edition of *Preservation Management of Digital Materials: A Handbook*.

DCC - Digital Curation Centre (<http://www.dcc.ac.uk/>)

Funded by the JISC and the UK Research Council's e-Science programme, the DCC aims to support creators, curators and re-users of digital data. The Centre provides a national focus for research and development into digital curation issues and promotes expertise and good practice by disseminating information through various channels, including workshops on different aspects of digital curation. The DCC's website boasts a host of useful features including a calendar of relevant events; the DCC Associates Network and Forum; a catalogue of tools; and a helpdesk and advisory service. A *Data Curation Manual* and a peer-reviewed journal, *The International Journal of Digital Curation*, are also under development.

ERPANET - Electronic Resource Preservation and Access Network (<http://www.erpanet.org/>) and DPE – Digital Preservation Europe (<http://www.digitalpreservationeurope.eu/>)

Established with European Commission funding in 2001, ERPANET provided information and advice to those interested in digital preservation techniques and best practice; ERPANET resources continue to be accessible at the project's website. Highlights of the website include erpaDirectory, a listing of digital preservation resources; erpaEvents, a series of expert workshops and seminars on digital preservation issues; and erpaGuidance, a collection of guidance tools. Digital Preservation Europe builds on the work of ERPANET by fostering collaboration between several digital preservation initiatives across Europe.

PADI - Preserving Access to Digital Information (<http://www.nla.gov.au/padi/>)

PADI is a gateway to international digital preservation resources. Maintained by the National Library of Australia, the website allows users to browse resources by type or topic. PADI is also 'safekeeping' resources deemed to be of long-term value by bringing these together within the NLA's PAN-DORA web-archive, where they will remain accessible for the long-term.