JISC Programme Synthesis Study: Supporting Digital Preservation & Asset Management in Institutions

A review of the 4-04 programme on digital preservation and asset management in institutions for the JISC Information Environment

Part II: Programme synthesis

Document details

Authors:	Maureen Pennock
Date:	January 2008 (links updated since submission to JISC in August 2008)
Version:	Final
Document Name:	404publicreport
Notes:	Circulation: Public

Acknowledgements

MAUREEN PENNOCK, DCC

Contents

MAUREEN PENNOCK, DCC

JISC Synthesis Study on the 4-04 programme: Supporting Digital Preservation & Asset Management in Institutions (DPAM)

Part II: A synthesis of the 4-04 programme on digital preservation and asset management in institutions for the JISC Information Environment

1 Introduction

In mid-2006, JISC requested that the Digital Curation Centre (DCC), in its capacity as a centre of excellence on digital preservation and digital curation, undertake a small-scale study to synthesise and help disseminate the results of projects funded under the Supporting Digital Preservation and Asset Management in Institutions (DPAM) programme. This report is the final outcome of that exercise.

The study ran from November 2006 to June 2007. Study data was collected during interviews and through desk based research, supplemented by e-mail or telephone contact. The study has resulted in production of a two-part report: part I contains an overall programme and project-level evaluation with subsequent recommendations for JISC; part II (this part) contains the synthesis exercise itself. Project documents and deliverables were valuable sources for development of this synthesis and have on occasion been re-used or adapted without specific attribution. Hyperlinks to the relevant project documents have been embedded into the report wherever they were found to be publicly-available.

This synthesis will be made available at the JISC website http://www.jisc.ac.uk, the Digital Curation Centre website http://www.dcc.ac.uk, and the UKOLN website http://www.ukoln.ac.uk. It will also be deposited in the ERPAePrints repository http://eprints.erpanet.org/

1.1 Structure of this report

This synthesis brings together the outputs of the programme into one location for ease of access. It provides details of each project funded by the programme. It also identifies the main themes explored during the projects and presents outputs according to these themes. As such, the outputs can be browsed in three main ways:

- By project
- By funding strand

- By subject-based themes

Several appendices are also provided:

- Appendix A: Related initiatives
- Appendix B: Project questionnaire
- Appendix C: Interview schedule.

2 Background

2.1 Programme aim

The DPAM programme ran from 2004 to 2006 and aimed to establish a basis for further development of institutional strategies and policies for long term preservation and asset management.

2.2 Asset types

The programme focussed on preservation or asset management of a limited range of digital object types commonly found in institutions or national services, particularly those addressed in previous JISC-funded feasibility studies.[1] This included e-journals, web resources, e-prints, e-learning objects and materials, and scientific data.
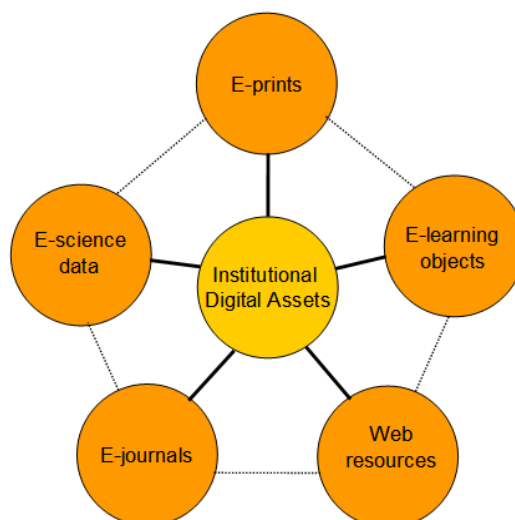
Figure 1: Primary asset types covered by DPAM

Corporate assets and information that had been covered by the previous community call for records management did not fall under the scope of the call unless they formed parts of broader activities with other asset types, or built on and extended previous work.

---

[1]   See http://www.jisc.ac.uk/index.cfm?name=programme_preservation for reports of the respective studies. These studies focused on the archiving of e-journals, web resources, e-prints, e-learning objects and materials, file format representation and rendering, and e-science data

3 Project overviews

- [Assessment of UK Data Archive and The National Archives compliance with OAIS/METS](#)

- [DAAT: Digital Asset Assessment Tool](#)

- [DPTP: Digital Preservation Training Programme](#)

- [ESPIDA: An effective Strategic model for the Preservation & disposal of Institutional Digital Assets](#)

- [LIFE (life cycle Information for E-literature)](#)

- [MANDATE: Managing Digital Assets in Tertiary Education](#)

- [Managing Risk: a Model Business Preservation Strategy for Corporate Digital Assets](#)

- [MAT: METS Awareness Training](#)

- [PARADIGM: Personal Archives Accessible in Digital Media](#)

- [PRESERV: PReservation Eprint SERVices](#)

- [SHERPA DP](#)

3.1.1 Assessment of UK Data Archive and The National Archives compliance with OAIS/METS

The UK Data Archive (UKDA) and The National Archives (TNA) have long-established responsibilities for digital preservation of materials created on electronic media. Each organisation has developed similar but different systems for record keeping and the generation and storage of metadata relating to the files stored. However, these systems were developed and put in place prior to the definition of the Open Archival Information System (OAIS) Reference model and the Metadata and Encoding Transmission Standard (METS) and therefore whilst it is assumed that both organisations broadly comply with the standards, neither has fully tested its infrastructure or metadata systems for compatibility with them.[2] This project permitted a thorough testing of the assumption.

The main elements of the project included:

- mapping of the systems and metadata currently in use by the two organisations against those in the OAIS reference model and the METS standard

- practical testing of the theoretical argument that the two partners comply with OAIS (though it is widely accepted that compliance can take a number of forms)

- exploration of the suitability of METS within an OAIS environment.

---

[2] Open Archival Information System reference model:
http://public.ccsds.org/publications/archive/650x0b1.pdf ;
METS official website: http://www.loc.gov/standards/mets/

The project concluded that, based on its assessment, both organisations exceeded the requirements of the OAIS reference model. Assessment of practical implementations of the model identified potential discrepancies inherent in the current model, particularly an overstated strength in the relationship between the OAIS concepts of producer and Designated Community, and the separation of the stored Archival Information Package from the stored Dissemination Information Package. Such detailed exploration and assessment of the model enabled the project to contribute to the reference model's five year review.

The project's assessment of the potential use of METS within an OAIS environment concluded that although there may be potential for its use in the future, there was little advantage to be gained at this moment in time by using METS over other more formal metadata standards. The flexibility of the schema was on the one hand deemed useful as it could be adapted for preservation purposes in environments with different object and preservation requirements. On the other hand, this very flexibility had potential to limit interoperability in a network of repositories that do not use the same tools and metadata content.

Varying vocabularies and terminology between different stakeholder groups is often cited as a cause for concern when developing or discussing approaches to preservation. The project found the shared terminology of the OAIS model invaluable as a means of communication between the UKDA and TNA, as the working terminologies of each are different despite their shared sectoral and archiving backgrounds.

The project found that the model was not highly scalable. This was felt to be a problem for smaller archives that wished to gain value from the model and achieve compliance with its main elements. As a result of this, TNA has commissioned work on producing an 'OAIS Lite' smaller scale version of the model for smaller scale archival institutions.

This highly specific and localised project produced valuable input for project partners on the role of the OAIS and METS standards within their current technical and organisational infrastructures for long term preservation. This enabled it to make significant contributions to the wider preservation and standards community, most publicly in the form of it contribution to the global OAIS five year review process. From a more practical perspective, the project also produced a valuable text-based tool to assist other organisations in assessing their own storage systems for OAIS compliance. Answers to questions posed in the tool will help when comparing the functions and workflows of an archive with those recommended by the OAIS standard.

Outputs

- Assessment of UKDA and TNA compliance with OAIS and METS standards: final report

    o Set of questions for OAIS compliance self-testing (pp 105 - 107)

3.1.2 DAAT: Digital Asset Assessment Tool

The DAAT project brought together extensive preservation expertise from a broad range of institutions: University of London Computing Centre (ULCC), the

Arts and Humanities Data Service (AHDS), the National Preservation Office (NPO), The National Archives (TNA), the British Library (BL), Kings College London (KCL), the School of Advanced Study of the University of London, and the Digital Preservation Coalition (DPC). Led by ULCC, the project aimed to develop a digital preservation assessment tool for institutions to identify the preservation needs of their digital holdings so that scarce resources could be focussed on those assets where the risk of loss and cost of loss was found to be greatest.

The project expected to base the DAAT tool on an existing tool (PAS Preservation Assessment Survey) produced by the National Preservation Office (NPO) which was already used in various forms to assess traditional collections in libraries, museums and archives and to assess photographic collections.[3] This proved not to be feasible: the project's primary conclusion was that the data collection model common to all varieties of PAS was not suitable for use in the digital environment. Furthermore, it was deemed possible that even the underlying conceptual model of where risk lies and how it is measured does not translate well from traditional materials to digital collections.

The project performed significant work in producing a digital survey tool and reports on what risk factors need to be assessed and the attributes an ideal survey tool would possess. It also assessed a range of automated file format assessment tools such as JHOVE and DROID as possible components of a future, more automated, digital asset assessment tool. However, it fundamentally determined that automated tools are still not capable of assigning a value to digital assets, a vital step in determining where preservation resources are best spent.

Outputs

- [DAAT Project website](#)

- [DAAT project final report](#) (October 2006)

- [Assessment of Risk Factors](#) (September 2006) – an assessment of the risk factors that may affect the survival of digital assets and would therefore need measured by an automatic risk assessment tool

- [Assessment of Attributes](#) (September 2006) – an assessment of the functional attributes needed in a digital asset assessment tool

- [Assessment of file format testing tools](#) (February 2006) - an assessment of selected existing file format testing tools: DROID, JHOVE and AIHT Empirical Walker

- [Database and data entry tool: D-PAS](#) (September 2006) – a pilot system reworking the NPO's PAS tool for digital assets. The database and the questions it contains may be of use to institutions as part of a wider asset management strategy.

  - [D-PAS Guidance Manual](#) (September 2006) – guidance manual for using pilot D-PAS tool

  - [Report on the use of the D-PAS tool](#) (February 2006) – evaluation of the tool by practitioners

---

[3] Further details on PAS are available from the British Library website
http://www.bl.uk/services/npo/paslib.html

Presentations/posters

- [DAAT project introductory poster](#) (July 2005)
- [DAAT tool promotional poster](#) (May 2006)

3.1.3 DPTP: Digital Preservation Training Programme

The University of London Computing Centre, in conjunction with the Digital Preservation Coalition and the British Library, established a modular digital preservation training programme to provide practical support for staff in HE institutions with responsibility for dealing with digital materials. The programme specifically aimed to develop and deliver a modular training programme with content targeted at all levels of staff, from service or project managers to technical and operational staff. DPTP's broader aim was to equip staff with the skills, tools and confidence to be able to embark on a pragmatic and cost-effective digital preservation programme appropriate to their own institutional needs, taking into account the practical reality that any approach to digital preservation is iterative and that no organisation is able to leap instantly into a fully formed digital preservation programme.

The course brought together delegates from technical, managerial and archival backgrounds to learn the same standards and methods for digital preservation, examining the need for policies, planning, strategies, standards and procedures. Content was derived in part from the Cornell Digital Preservation Management Workshop in the US; further content was drawn from the UK Digital Preservation handbook (Beagrie & Jones, 2002) or provided by selected experts resident in the UK.[4]

The course was delivered on three occasions, each time as a week-long residential course. Topics were covered using a mixture of lectures, discussions, practical tasks and a class project. Modules included:

- Access
- Costs and risks management
- Digital records management
- File formats
- Institutional repositories
- Legal issues
- Living with obsolescence
- Metadata
- OAIS
- Outsourcing
- Preservation approaches

---

[4] Cornell Digital Preservation Management Workshop: http://www.library.cornell.edu/iris/dpworkshop/
UK Digital Preservation Handbook: http://www.dpconline.org/graphics/handbook/

Course content may be requested from the course convener but is not available online.

Outputs

- [DPTP website](#)
- [DPTP final report](#)
- DPTP promotional poster, DCC 1st Annual Conference (October 2005)
- DPTP training:
    - Pilot week long residential course, Warwick (October 2005)
    - Week long residential course, Birmingham (March 2006)
    - Week long residential course, York (July 2006)

3.1.4 ESPIDA: An effective Strategic model for the Preservation & disposal of Institutional Digital Assets

ESPIDA aimed to help the digital preservation community make sound business cases for sustained funding. Based at the University of Glasgow, the two year project sought to develop and implement a sustainable business-focussed model for digital preservation as part of a wider knowledge management agenda in HE institutions. In particular, the project intended to identify the institutional costs and benefits of developing a coherent, managed and sustainable approach to the preservation of its digital assets in a way transparent to all stakeholders.

The scope of the project later developed from digital preservation in HE institutions out to the wider information environment and including public sector organisations. This extension of the project meant that digital preservation was embedded into the broader life cycle of digital information management and funding applications. Applicability of the final model is therefore significantly more extensive than first planned. A significant impact of this development however, is that costs and benefits associated explicitly with digital preservation have not been explored within the duration of the project.

The final model developed by ESPIDA can help make business cases for proposals that may not necessarily offer immediate financial benefit to an organisation, but rather bring benefit in more intangible spheres. It is being used initially within the area of digital resource management, but has potential for far wider application in articulating the value of many different types of projects where the primary outcomes are intangible, and thus in the creation of business cases and demonstrating the return on those investments.

Outputs

- [ESPIDA project website](#)
- [ESPIDA final report](#)
- [ESPIDA handbook & model](#)
    - Sample outcome scorecards

- Sample cost template
- Case studies
  - Institutional Repository Case Study
  - eTheses Case Study
  - Museum Archival Collection Case Study

- ESPIDA Training exercise – offers training in how to use the ESPIDA handbook and model

- ESPIDA brochure

- ESPIDA bibliography

Articles

- James Currall & Peter McKinney (2006), 'Investing in Value. A perspective on digital preservation', *D-Lib Magazine*, 12:4.

- Peter McKinney (2005), 'ESPIDA and sustainable digital preservation', *Archives, Records Management and Conservation*, 193.

Conference papers

- Claire Johnson (2007), 'Aligning Records Management with business objectives and the strategic agenda' *Records Management Society Conference*, (29 April -01 May, Brighton).

- James Currall (2006), 'I stand, they don't deliver: or why the highwayman approach to funding digital preservation doesn't work', given at *Online Information 2006*, 28-30 November 2006, Olympia Conference Centre, London.

- James Currall, Peter McKinney, Claire Johnson (2006), 'The world is all grown digital.... How shall a man persuade management what to do in such times?', given at *2nd International Digital Curation Conference*, Digital Data Curation in Practice, 21-22 November 2006, Hilton Glasgow Hotel, Glasgow.

- James Currall, Peter McKinney (2006), 'ESPIDA', given at *Preparing for success: foundations of good records management, Meeting of the RMS Scotland Group*, 21 June 2006, Dundee.

- James Currall, Peter McKinney, Claire Johnson (2006), 'Digital Preservation as an Albatross', given at *Archiving 2006*, May 23-26 2006, Ottawa, Canada.

- James Currall, Claire Johnson, Peter McKinney (2006), 'The Organ Grinder and the Monkey. Making a business case for sustainable digital preservation', given at *EU DLM Forum Conference* 5-7 October 2005 Budapest, Hungary.

- James Currall (2005), 'The ESPIDA Project', given at *DCC and DPC Joint Workshop on Digital Curation Cost Models* 26 July 2005 British Library, London.

Posters

- James Currall (2006), '[Pushing the right buttons: the ESPIDA framework for engaging senior management](#)', given at 2nd International Digital Curation Conference, 21-22 November 2006, Glasgow.

- Peter McKinney (2006), '[Repositories: how to show that they are as useful as YOU think they are!](#)', given at the *Open Scholarship Conference*, 2006, University of Glasgow.

- James Currall (2006), 'The *ESPIDA* Approach', given at Joint NDIIPP /JISC Digital Preservation Workshop 7-9 May 2006 Washington DC.

Events

- **Articulating the Value of Digital Assets** ESPIDA Final Conference, Monday 12th February, 2007, British Library Conference Centre, London

- **Sustainable Preservation of Digital Assets in a University** ESPIDA kick-off event, Friday, 11th February 2005, Wolfson Medical Building, University of Glasgow

3.1.5 LIFE: life cycle Information for E-Literature

Despite efforts over the past decade to provide costing information on digital preservation, it has remained difficult for organisations to anticipate such costs, particularly within the wider context of comprehensive digital asset management. Led by University College London (UCL) in collaboration with the British Library, the LIFE project aimed to rectify this situation by exploring and developing a life cycle approach to costing the digital archiving and preservation process within a digital library environment.

The LIFE project drew significantly on an existing life cycle costing model for analogue materials first developed by the British Library in the 1980's. This was adapted and extended into a formula and methodology that allows institutions to calculate the long term costs and future requirements for the preservation and ongoing care of their digital assets. The model was developed and tested on three different types of digital library collections, each of which was published as a case study.

The project established that a life cycle approach to costing is both useful and applicable for a range of digital collections, a conclusion is supported by the British Library's subsequent broad adoption of the LIFE methodology and formula. Similar interest has been shown from other quarters. Case studies highlighted the need for automated tools to reduce the costs associated with digital preservation activities. LIFE2 has recently been funded to further explore the LIFE formula and costing knowledge base using a broader range and variety of digital collections.[5]

Outputs

- [LIFE project website](#)

- [LIFE project summary report](#)

---

[5] LIFE2 website: http://www.ucl.ac.uk/ls/life/2/

- [LIFE project final report](#)
  - LIFE formula & methodology
  - Case studies:
    - Voluntarily Deposited Electronic Publications
    - Web Archives
    - e-Journals
- [LIFE Generic Preservation Model for Voluntarily Deposited Electronic Publications](#) – MS Excel functional spreadsheet
- [LIFE Generic Preservation Model for Web Archives](#) – MS Excel functional spreadsheet
- [LIFE Research Review: mapping the landscape, riding the life cycle](#) (November 2005) – an exploration of existing digital preservation and life cycle costing research
- [LIFE bibliography](#) (May 2006)

Events

- **LIFE Conference**, 20 April 2006, British Library Conference Centre, London

Conference papers/presentations

- Paul Ayris, [UCL e-journals case study](#), LIFE Conference (April 2006)
- Rory McLeod, [Case study: the legal deposit of electronic publications](#), LIFE Conference (April 2006)
- Paul Wheatley, [Modelling the digital preservation costs](#), LIFE Conference, (April 2006)
- Rory Mcleod, [The British Library Digital Preservation Strategy](#), State Library of Queensland (December 2006)
- Paul Ayris & James Watson, [life cycle Information For E-literature](#), DCC/DPC workshop on Cost Models for preserving digital assets (July 2005)
- James Watson, [Why Digital Collections should ride life cycles](#), JISC Joint programmes meeting (July 2005)
- James Watson, [LIFE: life cycle Information For E-literature](#), ESPIDA event (February 2005)

Posters

- [LIFE project poster](#) DCC 2nd Annual Conference (November 2006)
- LIFE project poster Open Scholarship 2006 (October 2006)
- [LIFE project poster](#) JISC joint programmes meeting (July 2005)

Articles

- DPC [Featured project interview - LIFE](#) (June 2006)

3.1.6 MANDATE: Managing Digital Assets in Tertiary Education

The MANDATE project was the only project funded within a college FE environment. John Wheatley College in Glasgow sought to develop a management toolkit to support asset management and preservation with a focus on digital curriculum and learning objects. This included materials associated with curricular administration and quality control processes liable to retention requirements and legal obligations (such as disclosure under the Freedom of Information Act). The toolkit was developed on the basis of internal investigations and testing within the John Wheatley college.

This ambitious but tightly focussed project addressed the management of objects throughout their life cycle. The main objectives of the project were to:

- create a toolkit to support digital asset management and preservation in FE colleges based on research practice to be piloted in John Wheatley College and supported by the Centre for Digital Library Research (CDLR) and the Scottish Library Information Council (SLIC);

- develop a system which would format, index, and store existing unstructured digital materials created in previously disparate processes and from the research practice create and test templates and workflow models for routine use in John Wheatley College and for application across the sector;

- develop appropriate structures for web-based storage of information that will contribute to the end goal of improved digital preservation; and

- examine the roles of various staff – teaching staff, librarians, administrators, technicians - in the creation of appropriate metadata

The toolkit is aimed primarily at managers, administrative and technical staff considering the management of digital assets in colleges, although use of the contents are by no means limited to this group and the contents are relevant for anyone considering implementation of a digital asset management system. The toolkit discusses some of the main issues to be considered in development of a digital asset management strategy and poses questions for consideration based on these issues. It highlights and supports the recommendations in the toolkit through the illustrative case study on implementation at John Wheatley College, in which the digital asset management system is used to organise materials supporting learning and teaching, as well as the publication of papers relating to the College's governance.

One of the key conclusions to be drawn from the project is the need for organisational /cultural issues to be addressed alongside technical issues. Significant conclusions were also drawn with regards to metadata, paticularly that collection from depositors must be as simple as possible, and that automatically captured or generated metadata is the key to both information retrieval and information services. This conclusion is compatible with the findings of other larger preservation and storage initiates.[6] Finally, and again in keeping with the conclusions of other projects, preservation is but part of the life

---

[6] For example, the PARADIGM project.

cycle; organisations must first coherently manage digital assets before attempting to undertake specific preservation activities.[7]

Outputs

- MANDATE website
- MANDATE project final report
- MANDATE toolkit
  - Strategy
  - Asset Types
  - Metadata
  - Workflow
  - Interoperability
  - Preservation
  - Legislation
  - Software
  - Training
  - Case Study
  - Glossary

Articles

- Craig Green, Digital assets: Mandate for the Future in *Information Scotland* April 2006 Vol 4(2)

Conference papers/presentations

- Craig Green, The JISC MANDATE project, at 'Sharing vision planning practice' 1st annual SLIC/FE member conference, John Wheatley College, (November 2006)

3.1.7 Managing Risk: A model business preservation strategy for corporate business assets

This small scale project at Kings College London (KCL) aimed to establish development of a new digital asset management strategy for the college. It addresses a common institutional need for a digital asset management strategy that combines academic and learning resources together with corporate information and thus acts as a valuable case study from which the wider HE community can learn.

The project carried out a survey into existing digital assets in the college. Analysis of the survey results raised some key issues concerning the way digital resources were managed. One of the main findings was that existing digital asset management practices varied widely and were inconsistent across and between different departments. A number of risks and recommendations were

---

[7] For example, the managing Risk and ESPIDA projects.

identified that were relevant at every stage across the life cycle. This led to development of a broader and college-wide work programme and strategy to maximise the long-term value of digital assets and minimise the associated risks. A College-wide Digital Assets Working Party (DAWP) was set up shortly after the report was issued to to take the project's findings forwards and work towards a coherent strategy for the management of corporate digital assetsthat would effectively balance and manage risks.

Outcomes and findings of the case study are detailed in the project final report, although the survey findings and analysis are for internal, college use only.

Outputs

- [Managing Risk: final report](#)
  - Guide to Digital Preservation - a briefing paper published within the final report and issued to KCL staff alongside survey results

Events

- **Managing Risk Conference - Challenges of the e-environment for records managers and archivists in Higher Education**, London, July 20th 2006.

Articles

- 'Managing Risk Conference: an event report', RMS Bulletin issue 134 October 2006 p.36

Posters

- Managing Risk, JISC Joint programmes meeting (July 2005)

3.1.8 MAT: METS Awareness Training

The METS Awareness Training project aimed to raise awareness in the UK HE community of the METS ("Metadata Encoding and Transmission") standard that was devised to act as a generic container for all metadata associated with objects in a digital library environment. This was to kick-start a critical mass of METS-based projects within the UK, so ensuring that UK institutions are fully standards-based in their digital object management. Based at the University of Oxford, the project adapted and expanded existing internal METS training materials and used these in a series of six training events that:

- Raised general awareness of METS and other closely related emerging standards both within the DPAM programme and among the wider community served by JISC;

- Provided attendees with sufficient information to assess how METS and related standards might contribute to their institutions' current and planned digital preservation and asset management activities

- Enabled attendees to find out more about METS for themselves and to prepare them for the METS tutorial workshops

Six non-technical events were held at locations around the UK. All course materials are freely available from the project website.

Seminars were run in the following locations:

i.National Library of Wales, Aberystwyth (25th October 2005)

ii.Open University, Milton Keynes (25th January 2006)

iii.Leeds University (7th April 2006)

iv.British Library, Boston Spa (18th September 2006)

v.Oxford Digital Library, Oxford (25th September 2006)

vi.Edinburgh University (3rd November 2006)

Seminars used the training materials developed for the project combined with interactive breakout sessions to deepen their understanding of METS and its application.

The main conclusion reached at the end of the project was that further training at a more technical level is needed, as is more support for implementers of the standard (including METS software tools). This may be achieved by greater institutional support, possibly through a central agency.

Outputs

- MAT project website

- MAT project final report

- MAT project training materials:

    - Introduction to METS (ppt)

    - Putting together a METS profile (ppt)

    - METS and learning objects (ppt)

    - Case studies (ppt)

        - METS in the Oxford Digital Library: A Case Study (ppt)

        - Case study 1: A slide collection (doc)

        - Case study 2: A collection of online books (doc)

        - Case study 3: A collection of online texts (doc)

        - Case study 4: A collection of videos (doc)

    - Where next? (ppt)

3.1.9 PARADIGM: Personal ARchives Accessible in Digital Media

PARADIGM explored how archivists might select, acquire, process, store, preserve and provide access to the digital archives of private individuals for future users. Library institutions at the Universities of Oxford and Manchester developed a best practice template and guidelines on issues relating to archiving and preservation of such papers, based in part on practical experiences gained during in the project in accessioning and ingesting digital private papers from contemporary British politicians into digital repositories.

Taking a life cycle and curatorial approach to the management and preservation of digital private papers, the project explored issues relating to all stages including creation, accession, ingest, metadata, storage and technical storage systems, access, and re-use. Technical, legal (including Freedom of Information, Data Protection, and IPR), cultural and organisational aspects were covered throughout the exploration, as was the role for standards such as OAIS and METS. The benefits and place of collaborative approaches to this broad challenge were investigated, particularly in light of the increasing number and outputs from international research and preservation initiatives (including work by a wide range of national archival institutions from around the globe). In light of global and potentially overlapping efforts in developing storage in different sectors (e.g. institutional and archival repositories), the project efforts towards harmonising traditional archival processed with digital object management processes have particularly wide reaching applicability in terms of improved communication and less duplication of effort across different sectors.

The project is still underway. Expected outcomes and outputs include:

- A template for ensuring long-term access for institutional holdings of digital personal papers

- Best-practice guidelines in the form of a workbook on issues relating to the archiving of personal papers in digital form, made available in sections as they are completed

- Strengthened local institutional capacity for digital preservation

- Developed templates for institutional policies for collection development, retention, and disposal

- Practical test of digital repository software DSpace and Fedora and related tools

- Investigation and report on the potential of the Archives Hub as a provider of distributed access to digital content preserved by the project

The Workbook is a significant and major output of the project, Aimed at organisations of any type that care for the personal archives of politicians, scientists, writers, journalists, academics or of other individuals, it covers all aspects of the digital life cycle.

Outputs

- PARADIGM project website

- PARADIGM project final report

- PARADIGM workbook on Digital Private Papers (2005 – 2007)

Conference papers/presentations

- Susan Thomas, Experiences with Personal Digital Archives, GLAM Meeting, Cambridge,16 March 2007.

- Susan Thomas, Making Links with Potential Donors, DCC Workshop on Pre- and Post-ingest Activity for Digital Archival Collections, Foresight Centre, Liverpool, 30 November 2006

- Susan Thomas, Practical Experiences with Personal Digital Archives: the Paradigm Project, General Meeting of the Data Standards Group of the Society of Archivists, British Library, London, 16 November 2006

- Susan Thomas, Working with Personal Digital Archives, Manuscripts Matter, British Library, London, 18-19th October 2006

- Susan Thomas, Paradigm: an Overview of the Project, Wellcome Library, London, 10 October 2006

- Susan Thomas, Paradigm: an Overview of the Project, Repository Research and Implementation Workshop, Oxford, 28-29 September 2006

- Janette Martin, The Paradigm Project: Practical Lessons, *Challenges of the e-environment for HE records managers & archivists* Kings College, London, 20th July 2006

- Renhart Gittens, Paradigm project, Fedora UK & Ireland User Group - Inaugural meeting, 9 May 2006

- Renhart Gittens, Paradigm project, Fedora Content Model Workshop, 4 May 2006

- Susan Thomas, Barriers to re-using email over time, DCC Workshop - E-mail Curation: Practical Approaches for Long-term Preservation and Access, 25 April 2006

- Janette Martin, The Paradigm Project, *IALHI Conference*, 7 September 2005

- Janette Martin, The Paradigm Project, *To Boldly Go. Exploring New Archival Frontiers*, Archives Hub day conference, 13 July 2005

- Janette Martin and Susan Thomas 'Comment conserver les archives électroniques des hommes politiques? Les ambitions du projet britannique Paradigm' , *Matériaux pour l'histoire de notre temps*, no. 79, Juillet-Septembre 2005, pp. 99-101.

- Susan Thomas, 'Writing a Workbook for Archivists and Curators of Digital Private Papers, the Paradigm Project', *D-Lib Magazine*, vol. II, no. 5, May 2005

Posters

- Susan Thomas, Paradigm poster JISC Joint Programme Meeting, Homerton College, Cambridge, 7-8 July 2005

- Susan Thomas, Paradigm project, poster presented at the Joint NDIIP/JISC Digital Preservation Workshop, Washington DC, 8 May 2006

Articles

- Susan Thomas, Featured project - interview with the Digital Preservation Coalition, 20 September 2006

- Susan Thomas, Paradigm: hands-on curation and preservation of born-digital personal archives pp32 – 36, the National Archives' Recordkeeping magazine, Summer 2006
- Susan Thomas and Janette Martin, Using the papers of contemporary British politicians as a testbed for the preservation of digital personal archives, *Journal of the Society of Archivists*, 2006

3.1.10 PRESERV: PReservation Eprint SERVices

The PRESERV project investigated long-term preservation for institutional repositories (IRs), by identifying preservation services in conjunction with specialists such as national libraries and archives, and building support for services into popular repository software, in this case EPrints. Led by the University of Southampton in conjunction with the National Archives, the British Library and Oxford University, the project developed a pilot and shared Web accessible preservation-based service and supporting technologies.

The main deliverable of the project is the ROAR-PRONOM service. This links PRONOM-DROID, a web-based service from the National Archives for file format identification, with the Eprints Registry of Open Access Repositories (ROAR) via an Open Archives Initiative (OAI) harvesting service.[8] This provides institutions with a so-called 'PRESERV profile' – an assessment of all formats currently held in their institutional repository. This is a first step towards development and implementation of preservation action and a technology watch service. The project also contributed to development of the widely deployed Eprints repository software, producing:

1. a new history module for the EPrints repository software that records metadata describing changes or actions performed on an object (contributing to an ability to verify an object's authenticity or integrity);

2. application programmes to package and disseminate data for delivery to an external service using either METS or the MPEG-21 DIDL standard;

3. a rights declaration, allowing preservation services to be undertaken upon the content deposited in a repository.

Further outputs included: a survey of repository preservation policies, which although time-specific is a revealing insight into contemporary IR approaches to preservation; a subset of the PREMIS metadata data dictionary that is tailored specifically towards the preservation requirements of IRs; and an extended model of potential preservation services, some of which may be addressed in the project follow-up, PRESERV2.[9]

Outputs

---

[8] PRONOM is an online registry of technical information - http://www.nationalarchives.gov.uk/pronom/. DROID (Digital Record Object Identification) is a software tool developed by The National Archives to perform automated batch identification of file formats - http://droid.sourceforge.net/wiki/index.php/Documentation . The EPrints ROAR service is available at http://roar.eprints.org/. More information about OAI can be found at http://www.openarchives.org/.

[9] PRESERV 2 project proposal: http://preserv.eprints.org/JISC-formal/preserv2-proposal-final-public.pdf.

- PRESERV project website
    - PRESERV project final report
- PRONOM-ROAR web based profiling service
- Preservation support in EPrints 3
    - History module
    - METS and DIDL export plug-ins
    - Creative Commons licensing with preservation rights declaration
- Preservation Metadata for Institutional Repositories: Applying PREMIS
- Survey of repository preservation policy and activity
- Digital preservation bibliography

Conference papers/presentations

- Towards Informed Preservation Choices for Institutional Repositories 2nd International Digital Curation Conference, Digital Data Curation in Practice, Glasgow (November 2006)
- IRs: towards preservation services JISC Repositories & Preservation Programme, New Projects Briefing, London (October 2006)
- Repository models and policies for preservation *DPC Briefing Day, Policies for Digital Repositories: models and approaches*, London, UK (July 2006)
- Preservation for Institutional Repositories: practical and invisible. *Ensuring Long-term Preservation and Adding Value to Scientific and Technical data (PV 2005)*, Edinburgh, UK (November 2005)
- Preservation for IRs *EPrints training course*, Southampton, UK (September 2005)
- Capturing preservation metadata from institutional repositories *DCC Workshop on the Long-term Curation within Digital Repositories*, Cambridge, UK (July 2005)
- PRESERV: a JISC 4/04 project (PPT slides). Pre-project report at *TARDis Project Review Meeting*, Southampton, UK (September 2004)

Posters

- Simple Preservation Services – towards Proactive Support for the Institutional Repository Manager 2nd International Conference on Open Repositories, San Antonio, TX (Jan 2007)
- PRESERV: preservation services for institutional repositories *Joint US-UK (NDIIPP-JISC) Digital Preservation Workshop*, Washington, DC (May 2006)

- PRESERV: Investigating and developing infrastructural digital preservation services for institutional repositories *1st International Digital Curation Conference*, Bath, UK (September 2005)

- Digital preservation for Institutional Repositories: building on a collaborative approach *9th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2005)*, Vienna, Austria (September 2005)

- Enabling Long-term Open Access to Materials in Institutional Repositories (IRs) *JISC Joint Programmes Meeting*, Cambridge, UK (July 2005)

Articles/Videos

- Hitchcock, Steve, Tim Brody, Jessie M.N. Hey and Leslie Carr, Digital Preservation Service Provider Models for Institutional Repositories: towards distributed services. *D-Lib Magazine*, May/June 2007.

- Preserv Us! The story of the Preserv project by the people behind it May 2007

- Hitchcock, Steve, Hold It, Hold It ... Start Again: The Perils of Project Video Production. *Ariadne*, No. 51, (April 2007)

- Steve Hitchcock, Featured project: PRESERV project interview with the DPC, June 2006

3.1.11 SHERPA DP: SHERPA Digital Preservation

The SHERPA DP project was led by the UK Arts and Humanities Data Service (AHDS), a well established centre of practical expertise in digital preservation, in conjunction with the SHERPA programme at the University of Nottingham, a centre of expertise and advice on institutional repository knowledge and advice. It aimed to create a collaborative and shared preservation environment for institutions participating in the SHERPA network of institutional repositories by utilising a preservation repository established within the AHDS. This removed the 'burden' and cost of adding preservation services to each IR in the network and provided a combi-environment that fully addressed all the requirements of the different phases within the life-cycle of digital information.

The project aimed to:

- Use the OAIS reference model to develop a persistent preservation environment for the SHERPA consortium, assigning rights and responsibilities and establishing protocols and work flow processes that will ensure the long-term preservation of the repository content.

- Explore the use of METS as the framework for packaging and transferring metadata held within the institutional repositories, including the preservation metadata created by the preservation service.

- Establish a coordinated set of protocols and software to be implemented as a working preservation service for a group of institutional repositories.

- Explore the use of open source software and tools to add functionality to and extend the storage layer of repository software applications.

- Draw together the experience gained into a practical user guide to implementing this type of preservation environment.

The project has yet to deliver its final deliverables and report.

Outputs

- [SHERPA DP website](#)

- [Requirements for a disaggregated service](#)

- [SHERPA DP OAIS report: An OAIS compliant model for disaggregated services](#)

- SHERPA DP Digital Preservation Handbook

- SHERPA DP final report

Conference presentations/papers

- Gareth Knight, [A problem shared...](#) DCC 2nd Annual Conference (November 2006)

- Gareth Knight, Preservation in Institutional Repositories: Preliminary conclusions, JISC New Projects workshop on 25th October 2006.

- Andrew Wilson, [Sherpa-DP and OAIS](#) Digital Preservation Coalition event (April 2006)

- Gareth Knight, [A Shared Preservation Model for Institutional Repositories](#)
  JISC Joint Programmes Meeting (July 2005)

- Gareth Knight, [A disaggregated model for the preservation of e-prints](#)   Digital Repositories: Interoperability and Common Services workshop in Heraklion, Crete (May 2005)

- Sheila Anderson, Stephen Pinfield [Preserving E-Prints: Scaling the Preservation Mountain](#) (January 2005)

Posters

- [Sherpa DP poster](#) 1st Annual Digital curation conference (September 2005)

Articles

- Andrew Wilson [Featured Project: SHERPA DP interview with the DPC](#)(January 2006)

4 Strand overviews

4.1.1 Institutional Management Support and Collaboration

Seven projects were funded to address institutional management support and collaboration for digital preservation and digital asset management in institutions:

- **Digital Preservation Training Programme (DPTP)**. Led by the University of London Computing centre in conjunction with the UK Digital Preservation Coalition (DPC) and Cornell University in the US, this project developed and delivered week-long residential modular training programmes that explored key issues in digital preservation and examined the need for policies, planning, standards and strategies in digital preservation.

- **Life cycle Information for E-literature (LIFE).** University College London (UCL) Library Services and the British Library (BL) examined the life cycles of key digital collections at UCL and established the individual stages in the cycle. These stages were then costed to show the full financial commitment of collecting digital materials over the long term. A particularly commendable output of this project was the formulae and unique digital preservation equation to cost digital preservation activity within the life cycle model.

- **Managing Digital Assets in Tertiary Education (MANDATE)**. John Wheatley FE college in Glasgow developed a management toolkit to support the implementation of digital asset management and preservation activities, notably digital learning support objects and official college records, within an FE context. The toolkit walks users through the main issues to consider in the development of a digital asset management system, from both a technological and a 'human' perspective, and is supported by case study examples of implementation of the toolkit at John Wheatley College.

- **Managing Risk: a Model Business Preservation Strategy for Corporate Digital Assets**. This case study project at Kings College London (KCL) explored digital asset management practices across a number of distributed sites at KCL, identified risks arising from these practices, and made recommendations to improve practices across the entire data life cycle.

- **METS Awareness Training (MAT)**. Led by Oxford Digital Library, the MAT project ran a series of training events to raise awareness in the UK higher education community of the METS (Metadata Encoding and Transmission) standard for digital object storage within a library environment.

- **Personal Archives Accessible in Digital Media (PARADIGM)**. Research libraries at the University of Oxford and the University of Manchester collaborated on the PARADGM project to explore issues

and develop a workbook with best practice guidelines on managing and preserving private digital papers, based on practical investigation of private papers from politicians.

- An Effective Strategic model for the Preservation & disposal of Institutional Digital Assets (ESPIDA). Based at the University of Glasgow, ESPIDA developed a model to help make business cases for funding to enable management of intangible assets, including digital objects.

Projects commonly addressed three main areas: training, costs & business models, and case studies.

4.1.1.1 Training

Training has been a significant component of this element of the programme, with two projects (MAT and DPTP) focusing exclusively on development and provision of training for key aspects of digital preservation and others (PARADIGM, ESPIDA and MANDATE) providing either localised training or materials that may be used for training purposes.

Training elements:

- Comprehensive digital preservation training
- METS training
- Business case training
- Digital Asset Management System training

Comprehensive digital preservation: The DPTP developed a modular training programme with content targeted at all levels of staff from service or project managers to technical and operational staff. DPTP's broader aim was to equip staff with the skills, tools and confidence to be able to embark on a pragmatic and cost-effective digital preservation programme appropriate to their own institutional needs, taking into account the practical reality that any approach to digital preservation is iterative and that no organisation is able to leap instantly into a fully formed digital preservation programme.

Aimed at multiple levels of attendee, the course brought together technical and archival professions to learn the same standards and methods for digital preservation, examining the need for policies, planning, strategies, standards and procedures in digital preservation. DPTP training materials are provided in hard copy to course delegates. The website provides an overview of the contents of the course and each module. More precise information about course and learning materials is available on request from the DPTP project administrator.

- DTPT training modules: introduction
    - Access: overview
    - Costs and Risks Management: overview

- [Digital Records Management: overview](#)

- [File Formats: overview](#)

- [Institutional Repositories: overview](#)

- [Legal Issues: overview](#)

- [Living with Obsolescence: overview](#)

- [Metadata: overview](#)

- [OAIS: overview](#)

- [Outsourcing: overview](#)

- [Preservation Approaches: overview](#)

METS: The METS Awareness Training project aimed to raise awareness in the UK higher education community of the METS ("Metadata Encoding and Transmission") standard, which had been devised to act as a generic container for all metadata associated with objects in a digital library environment. MAT Training materials are a mixture of graphic presentations and documentary case studies. All are available from the project website.

- [MAT training materials](#)

Business cases: The ESPIDA project developed a training exercise to facilitate understanding and implementation of the model. The exercise accompanies the ESPIDA Handbook and offers training in how to use the Handbook. It is envisaged that the training exercise could be used as a first step in understanding the approach before full implementation within organisations or groups wishing to use it.

- [ESPIDA training exercise](#)

Digital Asset Management Systems: The MANDATE toolkit contains a section on training, highlighting the importance of training for managers approaching digital asset management, including those with responsibility for flexible learning, library services, and quality, and technical managers. In development of the toolkit and in preparation for the development of a digital asset management system at the College, training at conceptual level was provided for both administrative staff and representatives of teaching staff (ILT Champions), based on an early prototype intended to support this purpose and act as a consultation tool. The prototype allowed consultation about interface design and the fit of the system with users' understandings of their role in the digital asset management workflow. The 'Guidance notes for using the asset management prototype' were used to support this training.

- [Guidance notes for using the asset management prototype](#)

4.1.1.2 Costs & Business models

Two projects in this strand focussed on cost and business models for preservation and asset management: LIFE and ESPIDA. Most other projects considered cost issues to some extent or another in the course of the project, but none focussed so explicitly on cost and business models as these two.

Preservation and digital asset management is, in every case, wholly reliant upon one thing: money. Projects were united in their findings that preservation should not be considered as an end in itself: it should be considered within the life cycle of digital object management. This is vital for appropriating the necessary finances to ensure that sustainability is a consideration within life cycle management, and to fund preservation activities.

Models:

- Cost model

- Business case model

Cost model: The LIFE project developed a methodology to calculate the long term costs and future preservation requirements of digital assets. Though the model has been developed in a digital library setting, there is no reason to believe that elements of the model are not transferable to other contexts. The methodology is life cycle based and scalable (though the extent to which this remains true has not yet been established). The preservation aspect of the model is particularly valuable as most other work on preservation costing has approached the subject in isolation, rather than how it is integrated into overall life cycle management.

- [LIFE project final report](#) (breakdown of cost model and approach, pp 9 – 16)

Business case model: The model that ESPIDA developed can help make business cases for proposals that may not necessarily offer immediate financial benefit to an organisation, but rather bring benefit in more intangible spheres. The ESPIDA approach offers advantages to both decision-makers and project proposers in terms of understanding proposals, ensuring the decision-making process is transparent and based on all relevant information, and allows both sets of stakeholders to communicate effectively, ultimately benefiting the organisation. While it was designed initially to be used within the area of digital resource management, it has potential for far wider application (decision making, performance measurement, change management). It can be used for:

- business cases for projects involving intangible benefits.

- impact of change decisions for comparing a change proposal with the current situation.

- evaluation of proposals providing a 'level playing field' for comparing a range of different proposals.

The model and process is highly complex and the immediate relationship to how it may be used for long term preservation is not necessarily clear. However, descriptive and training materials (including case studies) are provided to facilitate implementation.

- ESPIDA handbook (process description, p6)

4.1.1.3 Case studies

Case studies are the final over-riding element of this programme strand, with a number of projects developing content at a local level and sharing results to assist in broader implementation of project outputs by the community. Case studies are incredibly valuable. In some circumstances, they provide a basis on which to test and refine a given theory or approach. In others, they can be used to demonstrate to the wider community how an approach was implemented in a given context. This can give rise to several practical issues that may otherwise not be obvious to new implementers. In many cases, case studies can accomplish both of these outcomes, acting both as test cases AND as learning tools for others. Almost every project in this funding strand drew upon case studies during the course of the project. The prevalence of case studies within this strand illustrates a keen desire on the behalf of participants to encourage and assist in broader implementation of project outputs. Case studies covered:

- Digital asset management in an FE setting

- Digital asset management in an HE setting

- Costing of digital library collections

- Metadata implementation in digital libraries

- Archival collection development

- Developing a business case to support intangible assets

Digital asset management in an FE setting: MANDATE is an outstanding example of how case studies can function both as test scenarios against which to develop and refine an approach and also as practical examples of implementation from which others can glean valuable learning experiences. Each chapter of the MANDATE toolkit is supplemented by a case study description of implementation experiences at John Wheatley college that demonstrates how the toolkit was applied and shows how and why decisions were made to support any given activity. Textual descriptions are supported by graphics and tables, including screen shots of the prototype system and workflow models, and several links to supporting or college documentation are integrated into the study. MANDATE staff at John Wheatley intend to update the system once the system is actually up and running. The study is especially valuable as it is the only instance where research is oriented particularly on an FE environment.

- MANDATE Case Study

  - Purpose of digital asset management

  - Development of digital asset management

  - Asset types

  - Metadata requirements

  - Metadata standards issues

- Metadata quality issues
- Workflow modelling and development
- Interoperability issues
- Metadata mappings
- OAIS model
- Preservation issues
- Compliance with Freedom of Information legislation
- Compliance with Data Protection legislation
- Compliance with Copyright legislation
- Compliance with Special needs legislation
- Software issues
- Training issues

Digital asset management in an HE setting: Similarly, the Managing Risk project provided a local case study that addressed a common and shared institutional need for a digital asset management strategy. The project's findings, although directed very much at an internal audience, are therefore highly relevant to the wider community and institutions facing similar challenges.

- Managing Risk project: final report

Costing of digital library collections: The LIFE project also drew upon case studies as a mechanism for implementing and testing their model and identifying actual costs for different types of digital objects in different scenarios. The case studies were chosen with the precise aim of challenging the LIFE methodology as robustly as possible. Although preservation activities were not undertaken in any of the case studies, the studies were fundamental in shaping the specific and required elements of the preservation aspect of the model.

- LIFE project
  - VDEP Case study (final report, pp 17 – 51)
  - Web archiving case study (final report, pp 52 – 63)
  - UCL e-journals case study (final report, pp 64 – 87)
  - VDEP Preservation model
  - Web Arc preservation model

Metadata implementation in digital libraries: case studies were used as a valuable learning tool in the METS Awareness training project. The course included two types of case studies: the first comprising a presentation on METS implementation in a practical setting, such as at the National Library of Wales or Oxford Digital Library, and the second a fictional case study for participants to use in a practical exercise on implementing METS.

- METS in the Oxford Digital Library: A Case Study (ppt)

- Case study 1: A slide collection (doc)

- Case study 2: A collection of online books (doc)

- Case study 3: A collection of online texts (doc)

- Case study 4: A collection of videos (doc)

Archival Collection Development: The PARADIGM project carried out three case study/pilot projects to test and refine their knowledge of collection development for digital materials. This enabled them to make important recommendations on approaching and developing collection development policies and strategies, though the project ultimately recommended that more case studies are needed, particularly those that test different approaches or which address different domains. One primary case study was carried out, supplemented by two pilots. The primary case study involved identifying, capturing, and transferring records from contemporary politicians. This was supported by two pilot investigations. It led to a series of recommendations on how this valuable historical resource could be targeted for preservation. The study was supported by two complimentary pilot projects, one examining how political web resources could be targeted, the other exploring the issues in accessioning digital deposits from physical storage media. Project staff worked with the archive of Barbara Castle, which included two old PCs and a range of Amstrad disks – this was an archive deposited in the old-fashioned way (most of it is paper – some 500 boxes) and we worked with the BL on using forensic computing techniques to extract faithful images of the hard disks.

- Case study 1 - Hybrid personal archives of working politicians (final report, pp16 - 20)

- Case study 2 - Web archiving pilot (final report, p23)

- Case study 3 - Digital archaeology pilot (final report, p24)

Developing a business case to support intangible assets: A number of case studies were undertaken as part of the development of the ESPIDA approach, and as an aid to deeper understanding of the work and its applicability. The case studies demonstrate how the ESPIDA approach can be used in different scenarios and explore some examples that may hold some resonance for areas within HE/FE. Three case studies were undertaken, testing the model on an e-theses collection, an IR, and a museums collection. The studies were created with the help of stakeholders within and external to the University of Glasgow. Whilst based on real life scenarios, the studies are for explanatory purposes only and are not actual business cases.

For the Institutional Repository study, the project team that had developed the University's own repository helped explore the benefits and so-called 'dis-benefits' that can arise from setting up a repository to manage and disseminate academic outputs. The cost elements of this study are based on an example from Edinburgh University Library.

The eTheses study examines how three different options can be appraised using the ESPIDA Scorecards. By presenting the options together, decision-makers can quickly compare the three options and understand the risks and benefits of each.

The final study looks at the cataloguing of archival holdings of a museum. This was developed with the help of external experts and offers a rich example, where benefits are quite slow to be realised.

- ESPIDA Institutional Repository case study (handbook, pp 36 – 42)

- ESPIDA eTheses case study (handbook, pp 43 – 48)

- ESPIDA museum archival collection case study (handbook, pp 49 - 54)

4.1.2 Digital Asset Assessment Tools

On the basis that it is neither feasible nor desirable to preserve all the digital materials that an institution possesses, funding was made available for projects to explore and develop preservation assessment tools for digital assets. A single project was funded:

- Digital Asset Assessment Tool (DAAT). The University of London Computing Centre together with seven other project partners explored the extensibility and development of the existing analogue-based Preservation Assessment Survey (PAS) tool from the National Preservation Office (PAS) for assessment of preservation requirements of digital materials in a given collection.

Preservation risk assessment: Although no final tool for this task was ultimately released, a pilot database system that reworks the PAS tool for digital assets was delivered. Furthermore, the project carried out a number of comprehensive surveys and assessments of existing assessment tools for other areas of digital asset management, particularly ingest. The risk-based element of the project clearly links to the Managing Risk project funded under the previous strand, although the two projects explored very different mechanisms for identifying risks (one manual, the other automated).

- Pilot D-PAS database tool
  - Guidance on using the tool
- Assessment of risk factors
- File format tools report

4.1.3 Institutional Repository Infrastructure Development

In recognition of the very clear link between digital asset management, preservation, and repository storage, a further three projects were funded to explore institutional repository infrastructure development, particularly standards and shared repository services:

- **Assessment of UK Data Archive and The National Archives compliance with OAIS/METS**. The UK Data Archive at the University of Essex explored compliance of two national repositories with two significant (though with very different foci) standards used in preservation: the Open Archival Information System (OAIS) repository reference model and the METS metadata encoding transmission standard.

- **PReservation Eprint SERVices (PRESERV)**. Led by the University of Southampton, the PRESERV project explored a number of web-based preservation services for institutional repositories (IRs) and provided input to the development of preservation functionality within the institutional repository software package EPrints.

- **SHERPA Digital Preservation: Creating a Persistent Preservation Environment for Institutional Repositories (SHERPA DP)**. The Arts and Humanities Data Service (AHDS) worked closely with the SHERPA project at the University of Nottingham to bring existing repository systems in the SHERPA IR network together with the AHDS preservation repository and create an environment that fully addresses the requirements of digital information objects across their entire life cycle.

The PARADIGM project (see section 2.1) received additional funding under this strand for its activities in practical testing of repository software and related tools, as well as its use of the OAIS work flow model.

Whilst aspects of each project clearly relate to strand activities already discussed (such as assessment and case studies), a key element of this programme strand is the focus on technical and functional repository infrastructures for long term preservation. Both the UKDA and the TNA have archival repositories that cater for long term preservation; the AHDS preservation repository and preservation service has a clear long term preservation objective, and the PRESERV project had an explicit commitment to explore long term digital preservation within the context of institutional repositories.

Main aspects:

- Software

- Standards

- Shared services

4.1.3.1 Open source repository software for long term preservation

None of the projects explicitly set out to assess the suitability of given repository software package for long term preservation. However, PRESERV, SHERPA DP and PARADIGM all explored aspects of certain repository packages within their preservation or life cycle research.

- EPrints

- DSpace

- Fedora

Eprints: Eprints is an established, flexible software infrastructure that is used to store and provide access to a wide range of digital materials in institutions around the world. It was not designed with long term preservation in mind; the emphasis is on Open Access. The PRESERV project was focused particularly on preservation services using Eprints software. It did not aim to make Eprints into a long term preservation infrastructure, but to enable Eprints IRs to deliver materials with the necessary information for preservation specialists to provide those services. The PRESERV project investigated means for improving support for the OAIS Ingest and the Archival Store functions in Eprints-based IRs and developed a small number of plug-ins so that support for these services can be provided in real IRs. Three such features have been integrated into Eprints v3:

- [Complex-Object Export: METS and DIDL plug-ins](#)

- [History Module](#)

- [Preservation Rights Declaration](#)

The SHERPA DP project aimed to develop services for use by a network of institutional repositories, including Eprints installations. The project team carried out an architectural review of the Eprints software that investigated the data model and export interfaces used by Eprints. The primary objective was to identify data migration mechanisms that may be suitable for the SHERPA DP Archive-to-AHDS (Preservation Service) data transfer.

- [Architectural review of Eprints software](#)

DSpace: The DSpace digital repository system was designed to capture, store, index, preserve, and provide access to institutional digital research materials. As such, it is said to offer more immediate support for preservation than the EPrints system, although in practice this is debatable.

The SHERPA DP team carried out an architectural review of the DSpace software that investigated the data model and export interfaces used by DSpace. The primary objective was to identify data migration mechanisms that may be suitable for the SHERPA DP Archive-to-AHDS (Preservation Service) data transfer, four of which were identified.

- [Architectural review of DSpace software](#)

The PARADIGM project workbook contains a detailed description of the project team's experiences in setting up a DSpace repository. This is a useful 'case study' description that others can follow together with the official user installation guidelines. A second chapter in planned that maps the OAIS model against DSpace. This second chapter is more useful in terms of outputs with relevance for long term preservation.

- [PARADIGM workbook section on installing DSpace](#)

- PARADIGM workbook section on OAIS and DSpace ()

Fedora: Fedora is a general-purpose open source digital object repository management system for managing and delivering digital content. It is used in a

range of repository scenarios, from digital asset management and institutional repositories to digital libraries. It is more complex to install and manage than DSpace and Eprints, but is often preferred for more complex activities than commonly undertaken within an institutional repository setting.

The SHERPA DP project adopted Fedora as the basis for the preservation repository and built a technical infrastructure necessary to harvest metadata, transfer data, and perform relevant preservation activities. The team carried out a review of the Fedora software that investigated its digital asset management functionality, focusing on archive management and preservation functionality. This supported development of the preservation services to be offered by the project.

- [Fedora software review](#)

PARADIGM compared FEDORA with DSpace and ultimately selected FEDORA for implementation within the project as a working preservation repository system. Their reasons for this are contained within the PARADIGM final report. The PARADIGM project workbook contains a detailed description of the project team's experiences in setting up a Fedora repository. This is a useful 'case study' description that others can follow together with the official user installation guidelines. A second chapter in planned that maps the OAIS model against Fedora. This second chapter is more useful in terms of outputs with relevance for long term preservation.

- [PARADIGM workbook section on installing Fedora](#)
- PARADIGM workbook section on OAIS and Fedora ()

4.1.3.2 Standards for long term preservation

The use of standards in long-term preservation is a significant component the OAIS/METS Assessment, SHERPA DP, and the PARADIGM projects, particularly in terms of assessment and compliance.

- OAIS
- PREMIS
- METS
- MPEG DIDL

Open Archival Information System (OAIS) Reference Model: The OAIS model establishes a common framework of terms and concepts that describe the necessary functions and elements of an archive intended to provide permanent or indefinite long term preservation of digital information. Standard terminology allows existing and future archives to be more meaningfully compared and contrasted. It also provides a basis for further standardisation within an archival context and it should promote greater vendor awareness of, and support of, archival requirements. The OAIS reference model became an ISO standard in 2003.

The model has been widely seized upon by the archival and preservation communities and several institutions now claim compliance with the model. Yet the reference is model is not a precise specification and functionality of existing systems can be mapped against the OAIS model to varying levels. 'Compliance' can therefore mean different things to different organisations.

The OAIS/METS Assessment project sought to explore in detail what is meant by compliance and to assess whether two existing systems – the UKDA and the system in place at TNA – can be meaningfully deemed 'compliant'. In addition to a comprehensive assessment of compliance, the project delivered a useful checklist for other institutions to use in assessing their own levels of compliance:

- [A set of questions for OAIS compliance self-testing](#) (pp 105 – 107)

SHERPA DP used the OAIS reference model to develop a persistent preservation environment for the SHERPA consortium, assigning rights and responsibilities and establishing protocols and work flow processes that will ensure the long-term preservation of the repository content. An objective of the SHERPA DP project was to ensure that institutional repositories could be considered OAIS compliant. The final SHERPA DP model is essentially a modified version of an OAIS. The services provided by institutional repositories are mapped to the OAIS reference model and notable omissions identified and allocated to the preservation service.

The project delivered a report on an OAIS compliant model for disaggregated services to be delivered as part of the SHERPA DP project, including an assessment of the suitability of the reference model as a practical model. The handbook contains a useful overview of the OAIS framework and discussion of OAIS compliance, based in part on the investigations carried out as part of the OAIS/METS project.

- SHERPA DP Handbook: an OAIS compliant model for disaggregated services ()

The PARADIGM workbook has extensive information on the OAIS model, including some examples of how the model can work in practice. The section does not relate extensively to long term preservation, though it is a useful overview with links to related initiatives including compliance and certification.

- [PARADIGM workbook: section on the OAIS model](#)

PREMIS: PREMIS is a de facto metadata schema standard for preservation metadata. The PREMIS data dictionary specifies core metadata for verifying and tracking the provenance, authenticity and integrity of preserved digital assets.

PARADIGM explored the use of PREMIS metadata within METS files and the workbook contains a section on using PREMIS

- [PARADIGM workbook: section on PREMIS](#)

SHERPA DP adopted the PREMIS data dictionary as the basis for its preservation metadata schema. The project identified refinements necessary to support the operation of the preservation service. The broad requirements of

PREMIS were further supplement by format-specific metadata (e.g. MIX for image metadata), as appropriate.

PRESERV explored the use of PREMIS within an IR context. It found that PREMIS appeared to provide an excellent basis on which assess the needs of IRs with respect to preservation metadata, and it was possible to map the PREMIS elements to an extended model incorporating preservation services and registries. However, more implementation and testing are required, especially to validate the allocation of elements to preservation service providers and environment registries.

- [Preservation Metadata for Institutional Repositories: applying PREMIS](#)

Metadata Encoding and Transmission Standard (METS): The METS schema is a de facto standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using an XML Schema. The METS standard has been proposed as a suitable vehicle to support transfer and dissemination of objects for preservation by an external service provider, particularly in the case of complex objects (i.e. those comprising more than one file).

The OAIS/METS Awareness project considered the METS metadata standard, how it could be used further in a digital archive and its potential role at the National Archives and in the UKDA. The project found that METS documents (i.e. instances of METS in practice) are compatible with the OAIS model but that, particularly for born digital material, there was no clear benefit in implementing METS in the test repositories at this moment in time. The strengths and weaknesses of METS and its applicability within an archival environment are discussed in more detail in the project final report:

- [OAIS/METS Awareness final report: C7 Using the METS metadata standard in a digital archive](#) (pp 71 – 80)

The PARADIGM workbook section on administrative and preservation metadata contains a detailed introduction to METS and its use within a digital archive setting. The project developed specific guidance for users implementing METS in a digital archive setting, including advice on dealing with problems ingesting METS files into Fedora:

- Designing a METS Schema ()

- Sample METS Schema(s) for Personal Papers ()

- [Troubleshooting METS files and Fedora's Directory Ingest Service](#)

PRESERV released a METS export plug-in derived from work done by the Repository Bridge project, who implemented a METS export for EPrints 2. This was updated for the new plug-in architecture and data model in EPrints 3.

- [Complex-Object Export: METS plug-in](#)

MPEG-DIDL: Similarly to METS, the de facto standard MPEG-DIDL (Digital Item Declaration Language) has also been proposed as a suitable vehicle to support transfer and dissemination of complex objects for preservation by an

external service provider. PRESERV developed an MPEG DIDL plug-in to package data for delivery to an external service that has since been integrated into Eprints v3.

- Complex-Object Export: DIDL plug-in

4.1.3.3 Shared services for long term preservation

Shared infrastructure services are essential building blocks for an efficient and effective information and communications environment.[10] PRESERV and SHERPA DP both explored or developed shared services for long-term preservation.

- Format profiling
- Preservation Service Provider

Format profiling: The PRESERV project investigated long-term preservation for institutional repositories (IRs), by exploring the capacity for preservation services in conjunction with specialists, such as national libraries and archives and developing a web-based service for identifying the format of repository contents. One of the most widely recognised problems in preserving digital information is format obsolescence: as applications change over time to exploit the capabilities of new technology, digital objects created using old formats, or even old versions of formats, can become unreadable and the content and authenticity of a digital object is therefore at risk. One approach to this problem is to migrate the original format to a current, readable version. By knowing the formats of all objects in a repository, preservation strategies can be planned and action taken at the appropriate time on those objects that may otherwise be at risk of becoming obsolete. The need for accurate identification of the format of original source objects if often therefore at the centre of much digital preservation activity.

PRESERV worked with The National Archives to link the widely renowned PRONOM-DROID tool for file format identification to the widely used Registry of Open Access Repositories (ROAR) through an OAI harvesting service. As a result, file format profiles can be found for over 200 repositories listed in ROAR. This service – that identifies the file formats stored within a repository – is the first step in a more fully fledged technology watch service whereby 'at-risk' formats are subsequently flagged and acted upon and has been termed the PRONOM-ROAR format profiling service.

Continued provision and development of the web-based service is not a certainty: the service is considered a demonstrator, it is not explicitly funded and will require some maintenance. The code for the demonstrator has however been deposited in the local eprints repository.

- Digital Preservation Service Provider Models for Institutional Repositories: towards distributed services
- PRESERV Format Profiling: PRONOM-ROAR: An illustrated guide
- Public PRESERV profiles available via the ROAR website

---

[10] JISC Shared Infrastructure Services Synthesis report, Chapman and Russell (2006) http://www.jisc.ac.uk/media/documents/programmes/capital/jisc-sis-report-final.pdf.

- Institutional repository preservation contexts

Preservation Service Provider: The SHERPA DP project developed a practical and cost effective shared service system for performing preservation activities for a distributed network of institutional repositories. It created a demonstrator for a Preservation Service Provider and established the associated technical infrastructure and business model to support the project.

The model developed by the project identified two types of institution – Content Providers and Service Providers - that perform different tasks in the workflow. The institutional repositories participating in the project served as Content Providers, taking responsibility for accepting data and making it available to their user community. The Service Provider, in turn, takes responsibility for the long-term management of the digital objects, essentially serving as a centralised "dark archive".

The most notable development of the Preservation Service demonstrator was the creation of a reusable service framework that allows the integration of a disparate collection of software tools and standards. The project adopted Fedora as the basis for the preservation repository and built a technical infrastructure necessary to harvest metadata, transfer data, and perform relevant preservation activities. Appropriate software tools and standards were selected, including JHOVE and DROID as software tools to validate data objects; METS as a packaging standard; and PREMIS as a basis on which to create preservation metadata.

The demonstrator established during the project will continue to archive data made available by SHERPA DP partner institutions for a minimum of one year. Further investigation is necessary to establish a broader and more sustainable framework of repository services. The Preservation Service, as defined by the project proposal, performed preservation activities for a small number of institutional repositories operating the EPrints and DSpace repository software. However it found it unlikely that the revenue generated by e-print archives will be sufficient to support a full preservation service. A new need may be identified for the provision of preservation services to repositories that are responsible for a wider range of content types.

- SHERPA DP Preservation Handbook Part II: A disaggregated model for preservation service: the SHERPA DP approach ()

5 Thematic overviews

5.1.1 Asset types

- Corporate & Institutional administrative records
- ePrint deposits
- Datasets
- eJournals
- Web resources
- eLearning materials
- Private papers

Corporate & Institutional administrative records: Managing Risk, MANDATE and ESPIDA considered corporate and institutional records within the wider and more generic focus on corporate digital assets.

Managing Risk surveyed existing approaches to managing corporate digital assets - electronic business records (including meetings' records, student files and office correspondence) and research material in electronic form - within KCL and the risks involved in current asset management practices. MANDATE explored development of a develop a digital asset management strategy supporting flexible access to learning and teaching materials, and publication of documents supporting a pro-active approach to Freedom of Information legislation principles, and; ESPIDA investigated requirements for preservation of institutional records (though this was not the focus of their work). Each project resulted in valuable insight into the requirements of corporate and institutional records.

The MANDATE survey was extended to five departments and covered two types of digital assets:

- Unstructured assets –collections of documents held in digital form (word processed documents, spreadsheets etc.) – (termed 'Record Collections' in survey material provided to users).
- Structured assets –databases of information

It found that only around 20% of record collections were held in digital form and only 5% of these needed permanent preservation, though it recognised that this number may have been higher if the survey had been more comprehensive. The Managing Risk survey found that for administrative records, paper records continued to be the definitive versions. It remained common to print out electronic records – particularly emails – and save them in hard copy. There were recurring problems with version control and currency, especially in relation to research grant applications and ethics forms, and inconsistent application of practices, both good and bad, across different departments. Data was commonly duplicated in central departments and in schools. This may be

appropriate in some cases but in others it led to dispersed records with no single authoritative source. Overall, the project found that in general far more work is needed in identifying the archival needs of electronic records and in supporting their long-term retention and use.

- [Managing Risk final report](#)

MANDATE focussed on the development of a strategy and implementation of a system for digital asset management. It focussed on assets relevant to its core business, and to prioritise learning and teaching materials and documents supporting proactive FoI disclosure. In an effort to provide the widest possible support for the distinct ways of accessing the different types of asset stored, the college decided to develop separate upload interfaces for each type of asset. The specific asset types identified were:

- college publications,
- papers associated with college meetings,
- internal and external learning materials,
- quality process monitoring documents.

Prior to the MANDATE project, college assets were held in a combination of physical and digital formats and in different locations. The digital format used in the creation of documents was and remains Microsoft Word - on the whole these assets were distributed by administrative staff as physical documents. A strategic decision was made to use the PDF format for the publication of text documents (both College publications and learning and teaching materials). The selection of PDF as the preferred format for publishing college documents was partly taken in light of preservation issues. It was felt that such a common standard is very likely to remain accessible for the foreseeable future and issues relating to longer-term access (such as transfer to new formats) will be well catered for because of its widespread use. A survey of college retention requirements revealed that approximately 14% of the record classes required permanent preservation.

- [MANDATE final report](#)

ESPIDA sought to develop and implement a sustainable business-focussed model for digital preservation as part of a wider knowledge management agenda in HE institutions. Part of their initial activities included a survey of college administrative records to identify records of longer term value that required preservation. The results were somewhat surprising, for most respondents indicated they did not hold records of longer term value that required preservation.[11] This finding is consistent with that of the Managing Risk project, though perhaps not as much as the MANDATE project. It was fundamental in shaping the future direction of the project to focus more broadly on a model for developing a business case to support intangible activities.

---

[11] Conversation with project officers, February 2007.

ePrint deposits: SHERPA DP and PRESERV considered management and preservation of ePrint deposits within the broader context of institutional repositories. As such, the focus was on the infrastructure surrounding management and preservation of the assets rather than an exploration of the assets themselves. Both SHERPA DP and PRESERV addressed the subject of preservation metadata for eprints within an IR, but no project explored other aspects of ePrint deposits in detail.

SHERPA DP defined a refined set of metadata elements considered essential for preservation of e-prints provided by institutional repositories. The PREMIS Data Dictionary served as a basis on which to build an appropriate metadata scheme.

- SHERPA handbook: metadata model for ePrints ()

PRESERV explored how PREMIS can be applied within an Institutional Repository setting. As PREMIS is based on the idea of implementation, staff sought to develop an implementation involving institutional repositories (IRs). Their analysis mapped the five entity types identified in the PREMIS Data Dictionary - intellectual entities, objects, events, agents and rights - to potential metadata sources identified in an IR-preservation service provider model. It found that PREMIS appeared to provide an excellent basis on which assess the needs of IRs with respect to preservation metadata, though some elements may need to be adapted or omitted and more implementation and testing is required.

- [Preservation Metadata for Institutional Repositories: applying PREMIS](#)

Datasets: Datasets are commonly found within institutions and have a wide array of uses, including administrative records, and collection/compilation of research data and statistics. The OAIS/METS Assessment and Managing Risks projects both explored issues surrounding the management and preservation of datasets.

OAIS/METS Assessment did not explore in detail the particular requirements of database systems, concentrating instead on assessing the standard infrastructure by which the systems were preserved.

Managing Risk, on the other hand, made a number of discoveries about the way in which datasets were managed whilst in their active phase. The survey revealed that legacy database systems still hold important data and that although software is regularly upgraded and/or superseded, data isn't always migrated onto new systems. In some cases it continues to be trapped in and accessed through older variants. Furthermore, management of research data sets is inconsistent across different departments. The survey found excellent practice in some areas: technically adept teams managed their data assets well, but in other areas data management was poor. In one case personal data was being stored carelessly raising a legal compliance risk. In another case backup

provision was inadequate, threatening the viability of the whole study. Over half of the databases identified were small systems that contained data without long term preservation requirements. A more pressing concern was, in fact, that data was being kept longer than necessary and potentially contravening Data Protection or FoI. Finally, the survey indicated that long term retention of research data remains an issue for a minority of datasets. Some data sets have real archival value and they could be retained indefinitely for ongoing research use, but the data needs to be identified and appraised early to ensure effective and efficient preservation and this is not yet taking place.

- [Managing Risk: Final report](#)

eJournals: the preservation and management of eJournals is a recognised area of concern within the publishing and library communities. Until recently, little substantial and quantifiable work had been done to address the cost of managing and preserving eJournals. The LIFE project contributed extensively to this gap in knowledge, refining and testing its life cycle costing formula against an eJournal collection at UCL (University College London) library. It undertook an exploration of the life cycle stages for eJournals at UCL, from acquisition to preservation and produced two case studies in eJournals management at UCL. It also tested and refined the model against a collection at the British Library of Voluntarily Deposited Electronic Publications (VDEP), which is comprised of monographs and serials. This collection is considered comparable with an eJournal collection: although most are not, strictly speaking, academic journals, the file formats and modes of publication are similar enough for the collection to be an effective comparator for a collection of electronic journals.

The main study at UCL addressed the corpus of works from the Public Library of Science (PLoS). Examination of the file formats in the collection indicated that none of the file formats in existence in the collection had actually become obsolete (i.e. were no longer accessible because they were no longer supported). However, some were certainly aging, if still widely supported by software. 'A good example of this is .gif files. GIF files have been the most-widely used of web graphic formats, but they are widely deprecated because of their proprietary nature. They are also an old file format (1989) and use of the PNG (Portable Network Graphic) [format] has been advocated instead; they have also been used in digital preservation Case Studies as an "obsolete" file format before.[12] GIF files account for about 5.26% of the collection with a total of 455 files.'[13] The age of the collection is not specified, though the study does specify that the collection comprises 30 issues and that issues are published monthly; we can therefore estimate the age of the collection at no more than 3 years old when the study was undertaken (Summer 2005). It is not unreasonable to think that if the collection were more dated then an increased number of obsolete files may have been found.

Based on experiences with the UCL collection, the team found that the cost for the acquisition, ingest, metadata creation, access, storage and preservation of UCL's e-journals over 1 year is a range of £206.87 - £559.19 per title. Further

---

[12] See Rosenthal et al. 2005: http://www.dlib.org/dlib/january05/rosenthal/01rosenthal.html

[13] LIFE final report, pp 79 – 80.

projection of the UCL data over a 5 and 10 year range could not be performed with any certainty due to the lack of robustness of the data from the Case Studies. Desk based research estimated the figure to be a total of £2,189.83 for years 1 – 5 (i.e. for the whole five years, not per year) and £3,052.74 for years 6 – 10. Note however that, as service-led organisation, UCL is not responsible for ingest, storage or preservation functions and that access to the e-journal content is via the remote publisher's server. Consequently, no costs are allocated for these activities. In terms of staffing activity performed on the e-journal content, the most expensive activity in terms of costs are those actions concerned with making the materials accessible to users.

- LIFE final report: Case study on eJournals (pp 64 – 89)

The LIFE VDEP case study is comparable in terms of file formats and obsolescence findings, although costing is significantly more complex as the VDEP collection is comprised on four different types material: hand held monographs; hand-held serials; electronic monographs; and electronic serials.

The average cost for each over periods of 1, 5 and 10 years is reproduced in the table below:

Type & Average cost (£)	Yr 1	Yr 5	Yr 10
Average cost for hand held e-mono	19.49	37.41	48.12
Average cost for hand held serial	19.45	12.53	14.68
Average cost for e-mono	15.01	24.36	30.77
Average cost for e-serial	22.72	8.55	8.99

- LIFE final report: Case study on VDEPs (pp 17 – 51)

Web resources: Archiving and preservation of web resources has been a point of international attention from the cultural heritage community for the past decade. Most activities have concentrated on technical solutions, supplemented by occasional works on the larger life-cycle perspective.[14] The LIFE project contributed substantially to this gap in knowledge by refining and testing its life cycle costing formula against the web archive collection at the British Library. The PARADIGM project also addressed web archiving, though from the perspective of archiving personal web resources such as personal sites and blogs.

The LIFE project findings were developed from the BL web archive, which is part of a wider initiative - the UK Web Archiving Consotrium (UKWAC). Activities at the BL began in 2004 and they currently archive around 1000 website instances per year. The web archiving case study provides an excellent

---

[14] See for example Adrian Brown, 'Archiving Websites: a practical guide for information professionals' (Facet publishing, 2006)

overview of the specific activities involved in the web archiving process and life cycle, which, although broadly consistent across different collections, are tailored to the requirements of each collection type. An assessment of file types in the collection showed that whilst the majority were understandably and expectedly HTML files, a considerable number of image and text files and format types were present, alongside much smaller numbers of multimedia objects. The variety of file fypes and their interdependent structural relationships have obvious implications on any web preservation strategy, in that the strategy must cater for a broad range of file and format types whilst maintaining the structural relationships and dependencies of the collection that, when rendered, allow users to access a complete object.

Applying the LIFE costing model to the Web Archiving Case Study involved a significant amount of estimation (e.g. in terms of staff time, identifying proportions of time spent dealing with web resources alone, and preservation activities) and so the results should be considered only a very rough guide to the potential costs of preserving web materials. The study indicated relatively high costs in the first year (£493.17 per instance), likely due to the high effort required for the initial collection and assessment. This drops with each subsequent measurement, to £183.04 after five years, £128.83 after ten years, and £80.88 after twenty years. It anticipated that costs will continue to fall as more automated preservation tools are developed and source file formats become more stable. The study also suggested that a number of services can be shared (e.g. performing technology watch, selecting appropriate preservation strategies, and recording preservation metadata (representation information)), thus further reducing costs.[15]

- [LIFE final report: Case study on Web Archiving (pp 52 – 63)](#)

The PARADIGM web archiving pilot was run in conjunction with the UK Web Archiving Consortium (UKWAC) and London School of Economics. It involved taking snapshots of websites and blogs daily during the run-up to the General Election in 2005. Staff found that that obtaining permission to archive websites was 'a time consuming, and often fruitless, task. Obtaining permission from the politicians with whom PARADIGM had existing relationships was more successful than approaching politicians with whom there had been no previous contact, though permission was obtained from some additional politicians.'[16] Despite this, acquiring websites was considered less complicated that preserving and providing access to them, as websites are highly complex objects. The project therefore recommended that research libraries consider cooperation with a central web archive rather than attempt the activity themselves.

- PARADIGM final report: Web Archiving pilot

E-learning materials: e-learning materials are as much a type of digital asset as more traditional research and administrative records, yet comparatively little

---

[15] It is important to note that the Web Archiving activities of UKWAC are considered to be experimental. As more experience is developed, software becomes more stable and gathering activity is scaled up, then efficiency should dramatically increase and costs will again fall.

[16]PARADIGM final report.

attention has been paid to their management and preservation. This may be due to the relatively short period for which they remain valuable and re-usable: many learning materials need to be frequently updated in order to remain compatible with cutting edge research and teaching. Despite this, the materials still have value (albeit if not enduring) for the short term and for wider distribution and use by colleagues.[17]

The MANDATE project explored the management of digital earning materials within a broader digital asset management system that also contained documents relating to college governance. Microsoft Word is commonly used to create textual digital documents. On the whole these assets were distributed by administrative staff as physical documents. Some were held and distributed by teaching staff in common network storage areas and using folders in the institutional communications system. A strategic decision was made to use the PDF format for the publication of text documents (both College publications and learning and teaching materials). Other formats used for learning materials are already stored on web servers (e.g. National Learning Network materials, where the system stores metadata records and links to the assets rather than storing them independently). The end system designed for use at John Wheatley caters for the upload of newer file formats for learning materials such as PowerPoint presentations (likely to be used), images, audio, and video (not currently used extensively).

Insofar as preservation is concerned, the study found that 'in a Further Education (FE) setting a commitment to preservation is likely to be assessed in how it relates to the core business of the college and its legal and cultural obligations. A meaningful preservation commitment for an FE college is likely to be less than a decade. Beyond such a period materials are no longer required for business use or legal obligation for the college and the only likely motivation for preservation is to cater for unspecified future historical interest. Materials may not necessarily be deleted (or physically destroyed) after this period, but active curation is beyond the funded remit of the college. Preservation of information assets beyond obligation is likely to be an attractive option for a college only where there is little or no cost involved. Their use and value would now be more the preserve of national memory institutions.' College policy regarding preservation of learning and teaching materials in the new system is consistent with this statement and staff have decided that there is no point in retaining learning materials beyond the life of the teaching unit to which they are attached.

Externally acquired learning materials licensed for use on College servers are available only to enrolled users, although the metadata records about them are more widely accessible via the web interface and, at a later stage, the catalogue. Internally produced learning materials are only available to enrolled users.

- [MANDATE toolkit: Case Study on John Wheatley College](#)

---

[17] The JORUM repository is a typical example of this value. Jorum is a JISC-funded collaborative venture in UK Higher and Further Education to collect and share learning and teaching materials, allowing their reuse and repurposing, and standing as a national statement of the importance of creating interoperable, sustainable materials. See http://www.jorum.ac.uk/.

Private papers: Whilst the management and preservation of official institutional or corporate records has been an issue for several years now, the particular context and preservation requirements of personal archives have attracted little attention. The difference can be generalised across two main themes:

- From a technical and management perspective: personal archives are created and managed by individuals using a variety of standards, tools, and technologies over which the archivist has little or no direct influence and which are often not optimised for preservation. This compounds the preservation challenge, particularly as preservation is not addressed until much later in the life cycle, and preservation is best addressed as early in the life cycle as possible.

- From a legal perspective: permission to preserve or provide access to the collection is not always straightforward to identify and/or obtain. Several parties may have particular privacy or intellectual property rights. This means that different parts of the collection will often have different legal requirements and restrictions, all of which must be identified and addressed.

The project made several findings and recommendations on dealing with digital archives, several of which are generic but some of which pertain particularly to personal papers. For example, it found that digital preservation awareness amongst creators is low, even relating to preservation for their own short to medium term needs. Combined with the fact that many individuals do not even consider their papers to be 'archives', many personal archives are at risk of loss from either deletion or obsolescence. The study therefore recommended that it is important to increase awareness of archival and digital preservation generally, and specifically within domains and with individuals whose personal archives research libraries wish to collect. The project workbook therefore includes guidance for creators of personal archives that includes information on a number of relevant and basic issues like backup, automatically updating fields, and obsolescnce.

- [PARADIGM workbook: Guidance for creators of personal digital archives](#)

Insofar as collection is concerned, project staff found that the storage of personal digital archives is distributed across local devices, storage media, networked servers and web services. Some component parts may therefore be overlooked by a depositor or accessioning archivist, making it difficult to gather or be sure of a 'complete' collection.

Regarding obsolescence, none of the file formats accessioned from contemporary politicians (ranging from days old to five years old) were inaccessible using contemporary computing environments, although some required the use of proprietary software. However, deposits accepted from other, older sources proved more difficult, particularly data stored on Amstrad 3" disks – which required effort on a somewhat more heroic scale to recover.

PARADIGM developed a records survey to characterise the contemporary personal archives of politicians and assist archivists in identifying materials of historical value by assessing functions and roles, the nature of the records that document them, the vulnerability of records and their technical characteristics. Staff found that record surveying techniques are as useful for personal archives

as they are for organisational archives: they allow archivists to record valuable contextual information about material that will be transferred in the immediate and distant future, and therefore to prepare for its arrival. This is especially true in a digital context where more information is needed for initial transfer and processing. A range of material from the survey was identified for collection, including office documents, spreadsheets, presentations, images, emails, and email attachments in various formats.

- PARADIGM workbook: Records survey tool for digital private papers

The project also developed a deposit agreement to cover the placement of exemplar private paper materials for the duration of the pilot project, and a model agreement for the permanent placement of hybrid personal archives at a research library. These agreements help to establish the terms on which digital archives are placed with the research library, and ensure that both parties understand their obligations.

- PARADIGM workbook: Accessioning documentation for digital private papers

Finally, from a legislative perspective, the project recommended that changes to Freedom of Information legislation should be made to allow private archives donated to public institutions to be excluded from the provisions of the legislation for a number of years. This will facilitate collection development, particularly that which aims to acquire personal archival materials sooner after creation.[18]

5.1.2 Life-cycle investigations

Several projects have taken a life cycle approach to preservation and digital asset management, notably PARADIGM, LIFE, Managing Risk and SHERPA DP. Each one refers to a different type of digital asset.

- Life cycles of private papers
- Life cycles of digital library resources
- Life cycles of electronic corporate records
- Life cycle of an e-print deposit

Life cycles of private papers: The PARADIGM project has taken a holistic approach to the preservation of private papers and developed a life cycle model foir private papers that brings together traditional archival practice with the OAIS model. This is described in graphic form in the PARADIGM workbook. The main stages of the model are:

- Record creation
- Record's active life
- Onsite appraisal and selection
- Transfer to archival custody

---

[18] See PARADIGM final report.

- Ingest processing; ingest Submission Information Package (SIP)

- Appraisal and disposal

- Prepare and ingest Archival Information Package (AIP)

- Digital preservation of AIP

- Descriptive cataloguing

- Prepare Dissemination Package (DIP)

- Provide DIP

- Publish DIP

The workbook provides guidance across the entire life cycle of materials from creation to storage management, appraisal and disposal, cataloguing, metadata, strategies and re-use of materials. The project final report notes that preservation awareness and activity should begin as early in the life cycle as possible. To persuade creators that some measure of curation is required of them, an effort to engage with creators earlier in the life cycle of their archive is advocated.

The case studies showed that it is easier to identify and obtain archival records when intervention takes place at an early stage in the life cycle; they particularly showed that pro-active collection within the first five years required less intervention and heroic effort than reactive receipt of deposits. However, 'pro-active engagement with creators earlier in the life cycle places new demands on research libraries in terms of selecting and maintaining relationships with individuals whose archives are now, or will be, historically valuable'.

- PARADIGM final report: Case study on collecting archives of contemporary politicians ()

- PARADIGM final report: Case study on digital archaeology ()

The model includes specific activities relevant for each stage in the life cycle and identifies the approximate location in which each of the stages should take place (eg whilst with the creator, in transit, in a dark archive). The model is available as part of the workbook:

- PARADIGM project life cycle model

Life cycles of digital library resources: The life cycle model used by LIFE is based on work first undertaken by Andy Stephens in 1988, in which a formula for calculating the total cost of keeping an item in a Library throughout its life cycle is introduced. This work was refined by Stephens over the next few years, introducing costs to specific parts of the model, and was refined further by Helen Shenton in 2002/03 when specific focus on the aspects of preservation costs throughout the life cycle was included. This was a key extension and provided the first example of a life cycle cost model with a consideration for preservation. This model was thus taken as the basis for the life cycle costing model adopted and developed by the LIFE project.

The life cycle model consists of the following stages:

- Acquisition

- Ingest

- Metadata

- Access

- Storage

- Preservation

These six main life cycle categories are broken down further into life cycle elements. The project used case study implementation of the model to attribute costs to each stage. These should be taken as a guide, for the true costs may vary in different contexts. The persistent value of the model is in the provision of a framework and identification of the main costing elements that impact on the total life cycle cost of managing and preserving digital library materials.

- life cycle formula (contained within summary report)

- explanation of life cycle stages (contained within final report, pp 10 - 15)

- research review on life cycles and life cycle costing

Life cycles of electronic corporate records: In recognition that planning for preservation of electronic records should begin as early in the life cycle as possible, the Managing Risk project took a life cycle perspective that informed the survey of assets and subsequent recommendations. The project final report identifies a generic high level set of life cycle stages and corresponding management issues. The basic stages are:

- Creation

- Metadata

- Review and Appraisal

- Storage and Preservation

The survey furthermore suggested that a corporate classification scheme such as 'JISC's 2003 Study of the Records life cycle – Function Activity Model' would be sufficient for classifying digital records throughout their life cycle.

- Managing Risk project: digital life cycle management: stages & corresponding issues (pp 24 - 26)

Life cycle of an e-print: As part of its investigations into the preservation requirements of e-prints, the SHERPA DP project outlined an e-prints life cycle model that not only identifies the main stages in the life cycle but also suggests how responsibility may be allocated to different partners in the Sherpa DP disaggregated model. This extended model is based on an original suggestion by James et al in 2004. The main stages are:

- Creation

- Submission

- Revisions

- Quality Assessment and Publication

- End of frequent reading

- Technical Obsolescence

- Withdrawal

This life cycle model is unusual in that the cyclical nature of the model is unclear and the stages are not necessarily sequential. This is perhaps a particular characteristic of an e-print at its current stage of development, as opposed to an archival record or library resource. Archival records and library resources have been around for much longer and tend to follow clearer and better established curatorial processes that are oriented towards a longer time frame. ePrints, in comparison, are a relatively young and new type of digital asset for which the common ground in management processes have yet to be properly established and which are often perceived as having shorter retention requirements.

- SHERPA DP project: life cycle model for an e-print in an institutional repository ()

5.1.3 Assessment & surveys

Surveys and assessment formed a recurring theme in the programme. They can be found in at least five projects and address a number of different topics. Programme assessment and survey activities were of two main types, either examining existing tools or approaches, or developing tools for use in assessment and surveys.

- Assessment/survey of existing tools or approaches

  - File format testing tools

  - Compliance with METS

  - Repository preservation policies and approaches

  - Institutional digital asset management practices

- New assessment/survey tools

  - Digital asset assessment methodology & tool

  - Records survey tool

  - Repository contents profiling service

  - OAIS/METS Compliance assessment checklist

a) Assessments/surveys of existing approaches

File format testing tools: The DAAT project undertook an exploration of existing file format testing tools to explore whether or not they could be incorporated into the DAAT asset assessment tool. It focussed on three currently available tools:

- DROID (Digital Record Object Identification), a software tool developed by The National Archives to perform automated batch identification of file formats.

- JHOVE - JSTOR/Harvard Object Validation Environment, a tool that provides functions to perform format-specific identification, validation, and characterization of digital objects.

- AIHT at Stanford University: Automated Preservation Assessment of Heterogeneous Digital Collections, and the Empirical Walker tool.

The aim was not to make a general report on whether the tools worked, but whether they have useful qualities which will help identify the preservation needs of digital assets.

The team found DROID to be useful: it contains information on a large range of file formats, based on regularly updated information from PRONOM. It can identify the file format type and, for some types of files, also the version. However, the team found it did not provide any additional information that couldn't be found by looking at the file extension or a directory listing. It does not – and was not developed to – look at the entire information environment, which is something that the DAAT project D-PAS tool would need to do.

JHOVE was found to provide more information, particularly in terms of object validation and metadata extraction/reporting. The project test results indicated that the tool had a high success rate, though it was significantly limited by its ability to function on only a small range of file types. The team concluded that JHOVE is going to be essential in a digital preservation context (particularly one which manages to implement the PREMIS model), as it can be used continually to check and recheck the validity and integrity of each digital object stored in the repository.

The DAAT team was unable to obtain a copy of the Empirical Walker tool, so their assessment was based purely on a description of the tool in a published article. The article indicated that this was a significantly superior tool to either DROID or JHOVE, with significantly more functionality across both individual files and structured collections. The assessment is particularly useful for demonstrating an increasing maturity in the tools developed to assess digital objects.

- [DAAT file format tools report](#)

Compliance with OAIS/METS: The OAIS/METS Assessment of Compliance project assessed whether two existing systems – at the UK National Archives and the UK Data Acrhive – were compliant with the OAIS model and whether the metadata stored by each system could be successfully and usefully mapped against the METS standard.

The project found that, based on its assessment, both organisations exceeded the requirements of the OAIS reference model. Assessment of practical implementations of the model identified potential discrepancies inherent in the current model, particularly an overstated strength in the relationship between the OAIS concepts of producer and Designated Community, and the separation of the stored Archival Information Package from the stored Dissemination Information Package.

The project assessment of the use of METS within an OAIS environment concluded that although there may be potential for its use in the future, there was little advantage to be gained at this moment in time by using METS over other more formal metadata standards. The flexibility of the schema was on the one hand deemed useful as it could be adapted for preservation purposes in environments with different object and preservation requirements. On the other hand, this very flexibility had potential to limit interoperability in a network of repositories that do not use the same tools and metadata content.

- [Assessment of UKDA and TNA compliance with OAIS and METS standards: final report](#)

Repository preservation policies and approaches: PRESERV carried out a sample survey to determine exactly what repositories – particularly Institutional Repositories – were doing about digital preservation. The survey was carried out in June 2006 via email and was targeted at selected repository managers. Repositories were selected from the Registry of Open Access Repositories (ROAR). The availability of a PRESERV profile was a guiding factor in the selection process: since a knowledge of the file formats of objects in a repository is a prerequisite for preservation planning, the availability of a PRESERV profile provided a reference point for repositories included in the survey.

The survey found that none of the repositories surveyed had a formal preservation policy to guide decisions on the questions raised in the survey, yet *de facto* policies were being applied to specific areas such as file formats and transformations, for example. This is problematic, for the absence of comprehensive and forward thinking policies meant that original source data was frequently transformed without documentation, often prior to deposit in the repository, and adversely impacted on an ability to validate authenticity and integrity of files and their contents.

The survey questions and answers are detailed in the two documents below:

- [Survey of repository preservation policy and activity](#) (paper)
- [Repository models and policies for preservation](#) (presentation)

Digital asset management practices: As part of its investigation into corporate digital asset management at KCL and in recognition that preservation planning necessarily begins at the start of the digital records life cycle, the Managing Risk project carried out a survey of digital assets and management practices in a number of college units.

The survey covered both structured (databases) and unstructured (collections of individual files) digital assets. It revealed that practices were inconsistent across different schools and departments, and that duplication – particularly between digital and paper versions of files – was commonplace. A number of risks or inefficiencies were evident as a result of these practices. The inventory and report produced by this exercise enabled the project to make recommendations on reducing the risks associated on poor digital asset management, and to work towards a viable, embedded, long-term College-wide strategy for digital asset management.

The survey was carried out by external consultants who retained IPR over their survey tool, so the tool itself is not available. An overview of the survey results, which are considered to be indicative of practices at many other similar institutions, is contained in the final report.

- Managing Risk: final report

b) New survey & assessment tools/services

Digital asset assessment methodology & tool: Although the DAAT project were unable to deliver a final functional tool for digital asset assessment, the project did produce three deliverables that contribute to the development of such a tool.

The first of these is an *assessment of functional attributes* of a digital asset assessment tool, particularly the anticipated tool from the project, D-PAS. The report states that the D-PAS tool requires sampling, assessment, and analysis attributes:

- Sampling is necessary for the assessment to produce an accurate assessment of the overall collection without expending effort on investigating the entire collection, which in some contexts could consist of entire libraries.

- In order to make an accurate assessment of preservation priorities, the tool will need to assess such factors as use, storage, condition/usability, value/importance, item condition, file format, and damage.

- The tool will need to have analytical qualities so that factors assessed can have weighted scores returned which can then be used to generate reports on the preservation needs of a collection.

The project also identified a number of risk factors that may affect the survival of digital assets and thus should be measured by an automated assessment tool. These include physical risks for storage media and bit streams, format risks, and organisational risks.

Finally, the project released a preservation assessment method and prototype tool. The method comprises a manual survey, the results of which are entered into a tool – currently an Access database – and automatically scored. Guidance on using the method and tool are available. The project final report recognises that the survey tool requires significant additional work before it was usable in a wide variety of settings, but the reports which the project has made available are of use not only to those embarking on similar ventures, but also to those looking at the wider issues of automation in digital curation.

- Assessment of Risk Factors

- Assessment of Attributes

- Assessment of file format testing tools

- Database and data entry tool: D-PAS

   - D-PAS Guidance Manual

Records Survey Tool: PARADIGM developed a records survey for use in the offices on politicians, but most of the questions may be applied to the private papers of other individuals. The survey is designed to elicit an overview of the paper and electronic records being created, with a view to their subsequent appraisal and preservation.

The records survey tool assisted the surveying archivists in identifying materials of historical value by assessing functions and roles, the nature of the records that document them, the vulnerability of records and their technical characteristics. It also prompted creators to think about the historical value of their traditional and digital materials, as well as preservation-related issues. The information that the records survey covers is described in the following table:

Information	*Purpose*
Series of historical interest	Identify location of material for accession; descriptive overview assists in basic intellectual control of materials prior to cataloguing
Quantity of material for accessioning	Informs archivist's approach to accession
Specification of hardware and software used to create records	Informs archivist's approach to accession; Recorded as preservation metadata
Formats used	Informs archivist's approach to accession; Recorded as preservation metadata
How/by whom records are created	Recorded as useful contextual information for descriptive purposes
Third party rights present in the material	Inform administration of material; recorded as metadata
Access restrictions required by creator	Informs administration of material
How records are managed by their creator	Informs timing of accessions

Paradigm supplemented the records survey tool with simple measures such as taking screenshots of file management tools and computer hardware details or generating textual files representing creators' arrangement of digital materials. The final report recommends that research libraries use record

surveying techniques when talking to creators and surveying materials, as this information can be useful for a number of digital, and general, archive administration activities.

- [PARADIGM records survey tool](#)

Repository contents profiling service: PRESERV worked with The National Archives to link the widely renowned PRONOM-DROID tool for file format identification to the widely used Registry of Open Access Repositories (ROAR) through an OAI harvesting service and deliver a web-based assessment tool known as the PRONOM-ROAR format profiling service. As a result, file format profiles can be found for over 200 repositories listed in ROAR. This assessment of file formats stored within a repository is the first step in a more fully fledged technology watch service whereby 'at-risk' formats are subsequently flagged and acted upon.

Continued provision and development of the web-based service is not a certainty: the service is considered a demonstrator, it is not explicitly funded and will require some maintenance. The code for the demonstrator has however been deposited in the local eprints repository.(shared service)

- [PRESERV Format Profiling: PRONOM-ROAR: An illustrated guide](#)

OAIS/METS Compliance assessment checklist: The OAIS/METS Assessment project carried out an in-depth assessment of system compliance at the UKDA and TNA with the OAIS and METS standards. The UKDA and TNA developed a set of questions to help carry out their assessment of OAIS compliance. This question set can be extended and tailored for use as a self-assessment tool for OAIS compliance testing by other institutions. Answers to these questions can help when comparing the functions and workflows in an archive with those recommended by the OAIS standard.

The question set covers the functions of the archive, OAIS mandatory responsibilities, functional entities, and the OAIS model. It can be found in the back pages of the project final report:

- [Assessment of UKDA and TNA compliance with OAIS and METS standards: final report](#) (pp 105 – 107)

5.1.4 Costs & business models

A number of projects addressed issues relating specifically to costs and business modelling, namely the DPTP, LIFE, SHERPA DP and ESPIDA.

- Costs training
- Costs model: digital library
- Business model: preservation service
- Business case model: intangible assets

Costs training: the DPTP included a module on costs that linked the costs of digital preservation closely with good risk management. The module is intended to provide trainees with an increased awareness of the factors affecting the costs of digital preservation and an ability to set these costs within a wider life cycle and digital asset management strategic framework.

- DPTP module on costs and risk management (outline)

Digital library cost model: The LIFE project developed a life cycle based, scalable methodology to calculate the long term costs and future preservation requirements of digital assets. Though the model has been developed in a digital library setting, there is no reason to believe that elements of the model are not transferable to other contexts. The preservation aspect of the model is particularly valuable as most other work on preservation costing has approached the subject in isolation, rather than how it is integrated into overall life cycle management.

- LIFE project final report (breakdown of cost model and approach, pp 9 – 16)

Preservation service business model: SHERPA DP built on the outputs of both the LIFE project to develop a cost and business model for the SHERPA DP preservation service. The model uses the life cycle cost model developed by the LIFE Project as a basis, indicating how it may be applied to Service Providers offering a service to institutional repositories, or indeed anyone, who wish to assess the cost of preservation on a year-by-year basis. Costs are calculated on a three-tier basis – set-up costs; service costs; and exit costs that occur at different time period during the Service contract. Service costs serve as the focus of investigation, with the cost of storage, preservation and limited technical support being provided as examples of the method in which different levels of service will incur varying costs. The business model may prove useful for any institution that is seeking to establish a Preservation Service.

Further information on the business model is contained within the project final report:

- SHERPA DP final report ()

Intangible assets business case model: The business case model that ESPIDA developed can help make business cases for proposals that bring benefit in more intangible spheres than immediate financial benefits. While it was designed initially to be used within the area of digital resource management, it has potential for far wider application, including decision making, performance measurement, and change management.

- ESPIDA handbook & model

5.1.5 Training

The importance of training for preservation and digital asset management cannot be underestimated. Many of the practices in digital asset management and preservation are fundamentally different to those for digital materials. Staff working towards implementing solutions therefore often require training in new technologies, approaches, and even ways of thinking. Training possibilities within DPAM were diverse, in keeping with the broad and diverse nature of the programme.

- Comprehensive preservation training

- METS training

- Business case training

- Digital Asset Management System training

Comprehensive preservation training: The DPTP developed a modular training programme with content targeted at all levels of staff from service or project managers to technical and operational staff. DPTP's broader aim was to equip staff with the skills, tools and confidence to be able to embark on a pragmatic and cost-effective digital preservation programme appropriate to their own institutional needs, taking into account the practical reality that any approach to digital preservation is iterative and that no organisation is able to leap instantly into a fully formed digital preservation programme.

Aimed at multiple levels of attendee, the course brought together technical and archival professions to learn the same standards and methods for digital preservation, examining the need for policies, planning, strategies, standards and procedures in digital preservation. DPTP training materials are provided in hard copy to course delegates. The website provides an overview of the contents of the course and each module. More precise information about course and learning materials is available on request from the DPTP project administrator.

- DTPT training modules: introduction

  - Access: overview

  - Costs and Risks Management: overview

  - Digital Records Management: overview

  - File Formats: overview

  - Institutional Repositories: overview

  - Legal Issues: overview

  - Living with Obsolescence: overview

  - Metadata: overview

  - OAIS: overview

  - Outsourcing: overview

  - Preservation Approaches: overview

METS: The METS Awareness Training project aimed to raise awareness in the UK higher education community of the METS ("Metadata Encoding and Transmission") standard, which had been devised to act as a generic container for all metadata associated with objects in a digital library environment. MAT Training materials are a mixture of graphic presentations and documentary case studies. All are available from the project website.

- MAT training materials

Business cases: The ESPIDA project developed a training exercise to facilitate understanding and implementation of the model. The exercise accompanies the ESPIDA Handbook and offers training in how to use the Handbook. It is envisaged that the training exercise could be used as a first step in understanding the approach before full implementation within organisations or groups wishing to use it.

- ESPIDA training exercise

Digital Asset Management Systems: The MANDATE toolkit contains a section on training, highlighting the importance of training for managers approaching digital asset management, including those with responsibility for flexible learning, library services, and quality, and technical managers. In development of the toolkit and in preparation for the development of a digital asset management system at the College, training at conceptual level was provided for both administrative staff and representatives of teaching staff (ILT Champions), based on an early prototype intended to support this purpose and act as a consultation tool. The prototype allowed consultation about interface design and the fit of the system with users' understandings of their role in the digital asset management workflow. The 'Guidance notes for using the asset management prototype' were used to support this training.

- Guidance notes for using the asset management prototype

5.1.6 Repositories

Repositories were a significant element of the programme, particularly in terms of preservation infrastructures and institutional repository infrastructure development. Four main areas of activity are discernible:

- Software

- Preservation services for IRs

- Standards

- Training

1. Software

Selection of software: There are many repository software options or packages to choose from. The big three – Eprints, Fedora and DSpace – are the most widely known open source solutions, particularly in the Institutional Repository sector, but there are several alternatives. The MANDATE toolkit contains information on the issues surrounding selection of a 'digital asset management' system. Its inclusion (albeit somewhat limited) of proprietary solutions and supporting tools is particularly valuable.

The College ultimately chose to develop a system based on proprietary database management and web server software for the implementation of the *mandate toolkit*, reflecting concerns about the total cost of open source software and the skill set of its technical support teams (which are based on Microsoft platforms.)

- [MANDATE toolkit: Section on Software](#)

PARADIGM compared FEDORA with DSpace and ultimately selected FEDORA for implementation within the project as a working preservation repository system. This was for several reasons, including its extensibility and its ability to support complex objects and their hierarchical relationships. A comparison of the two will be published as part of the PARADIGM workbook.

- [PARADIGM workbook](#)

The DPTP course included a section on institutional repositories that introduced a number of systems and described some of their characteristics. The website only provides an overview of the contents of module. More precise information about course and learning materials is available on request from the DPTP project administrator.

- [DPTP IR module overview](#)

Eprints: Eprints is an established, flexible software infrastructure that is used to store and provide access to a wide range of digital materials in institutions around the world. It was not designed with long term preservation in mind; the emphasis is on Open Access. The PRESERV project was focused particularly on preservation services using Eprints software. It did not aim to make Eprints into a long term preservation infrastructure, but to enable Eprints IRs to deliver materials with the necessary information for preservation specialists to provide those services. The PRESERV project investigated means for improving support for the OAIS Ingest and the Archival Store functions in Eprints-based IRs and developed a small number of plug-ins so that support for these services can be provided in real IRs. Three such features have been integrated into Eprints v3:

- [Complex-Object Export: METS and DIDL plug-ins](#)

- [History Module](#)

- [Preservation Rights Declaration](#)

The SHERPA DP project aimed to develop services for use by a network of institutional repositories, including Eprints installations. The project team carried out an architectural review of the Eprints software that investigated the data model and export interfaces used by Eprints. The primary objective was to

identify data migration mechanisms that may be suitable for the SHERPA DP Archive-to-AHDS (Preservation Service) data transfer.

- Architectural review of Eprints software ()

DSpace: The DSpace digital repository system was designed to capture, store, index, preserve, and provide access to institutional digital research materials. As such, it is said to offer more immediate support for preservation than the EPrints system, although in practice this is debatable.

The SHERPA DP team carried out an architectural review of the DSpace software that investigated the data model and export interfaces used by DSpace. The primary objective was to identify data migration mechanisms that may be suitable for the SHERPA DP Archive-to-AHDS (Preservation Service) data transfer, four of which were identified.

- Architectural review of DSpace software

The PARADIGM project workbook contains a detailed description of the project team's experiences in setting up a DSpace repository. This is a useful 'case study' description that others can follow together with the official user installation guidelines. A second chapter in planned that maps the OAIS model against DSpace. This second chapter is more useful in terms of outputs with relevance for long term preservation.

- PARADIGM workbook section on installing DSpace
- PARADIGM workbook section on OAIS and DSpace ()

Fedora: Fedora is a general-purpose open source digital object repository management system for managing and delivering digital content. It is used in a range of repository scenarios, from digital asset management and institutional repositories to digital libraries. It is more complex to install and manage than DSpace and Eprints, but is often preferred for more complex activities than commonly undertaken within an institutional repository setting.

The SHERPA DP project adopted Fedora as the basis for the preservation repository and built a technical infrastructure necessary to harvest metadata, transfer data, and perform relevant preservation activities. The team carried out a review of the Fedora software that investigated its digital asset management functionality, focusing on archive management and preservation functionality. This supported development of the preservation services to be offered by the project.

- Fedora software review ()

PARADIGM compared FEDORA with DSpace and ultimately selected FEDORA for implementation within the project as a working preservation repository system. The PARADIGM project workbook contains a detailed description of the project team's experiences in setting up a Fedora repository. This is a useful 'case study' description that others can follow together with the official user installation guidelines. A second chapter in planned that maps the OAIS model against Fedora. This second chapter is more useful in terms of outputs with relevance for long term preservation.

- [PARADIGM workbook section on installing Fedora](#)

- PARADIGM workbook section on OAIS and Fedora ()

2. Services

Format profiling: The PRESERV project investigated long-term preservation for institutional repositories (IRs), by exploring the capacity for preservation services in conjunction with specialists, such as national libraries and archives and developing a web-based service for identifying the format of repository contents. One of the most widely recognised problems in preserving digital information is format obsolescence: as applications change over time to exploit the capabilities of new technology, digital objects created using old formats, or even old versions of formats, can become unreadable and the content and authenticity of a digital object is therefore at risk. One approach to this problem is to migrate the original format to a current, readable version. By knowing the formats of all objects in a repository, preservation strategies can be planned and action taken at the appropriate time on those objects that may otherwise be at risk of becoming obsolete. The need for accurate identification of the format of original source objects if often therefore at the centre of much digital preservation activity.

PRESERV worked with The National Archives to link the widely renowned PRONOM-DROID tool for file format identification to the widely used Registry of Open Access Repositories (ROAR) through an OAI harvesting service. As a result, file format profiles can be found for over 200 repositories listed in ROAR. This service – that identifies the file formats stored within a repository – is the first step in a more fully fledged technology watch service whereby 'at-risk' formats are subsequently flagged and acted upon and has been termed the PRONOM-ROAR format profiling service.

Continued provision and development of the web-based service is not a certainty: the service is considered a demonstrator, it is not explicitly funded and will require some maintenance. The code for the demonstrator has however been deposited in the local eprints repository.

- [Digital Preservation Service Provider Models for Institutional Repositories: towards distributed services](#)

- [PRESERV Format Profiling: PRONOM-ROAR: An illustrated guide](#)

- [Public PRESERV profiles available via the ROAR website](#)

- [Institutional repository preservation contexts](#)

Preservation Service Provider: The SHERPA DP project developed a practical and cost effective shared service system for performing preservation activities for a distributed network of institutional repositories. It created a demonstrator for a Preservation Service Provider and established the associated technical infrastructure and business model to support the project.

The model developed by the project identified two types of institution – Content Providers and Service Providers - that perform different tasks in the workflow. The institutional repositories participating in the project served as Content Providers, taking responsibility for accepting data and making it available to

their user community. The Service Provider, in turn, takes responsibility for the long-term management of the digital objects, essentially serving as a centralised "dark archive".

The most notable development of the Preservation Service demonstrator was the creation of a reusable service framework that allows the integration of a disparate collection of software tools and standards. The project adopted Fedora as the basis for the preservation repository and built a technical infrastructure necessary to harvest metadata, transfer data, and perform relevant preservation activities. Appropriate software tools and standards were selected, including JHOVE and DROID as software tools to validate data objects; METS as a packaging standard; and PREMIS as a basis on which to create preservation metadata.

The demonstrator established during the project will continue to archive data made available by SHERPA DP partner institutions for a minimum of one year. Further investigation is necessary to establish a broader and more sustainable framework of repository services. The Preservation Service, as defined by the project proposal, performed preservation activities for a small number of institutional repositories operating the EPrints and DSpace repository software. However it found it unlikely that the revenue generated by e-print archives will be sufficient to support a full preservation service. A new need may be identified for the provision of preservation services to repositories that are responsible for a wider range of content types.

- SHERPA DP Preservation Handbook Part II: A disaggregated model for preservation service: the SHERPA DP approach ()

3. Standards

Open Archival Information System (OAIS) Reference Model: The OAIS model establishes a common framework of terms and concepts that describe the necessary functions and elements of an archive intended to provide permanent or indefinite long term preservation of digital information. Standard terminology allows existing and future archives to be more meaningfully compared and contrasted. It also provides a basis for further standardisation within an archival context and it should promote greater vendor awareness of, and support of, archival requirements. The OAIS reference model became an ISO standard in 2003.

The model has been widely seized upon by the archival and preservation communities and several institutions now claim compliance with the model. Yet the reference is model is not a precise specification and functionality of existing systems can be mapped against the OAIS model to varying levels. 'Compliance' can therefore mean different things to different organisations.

The OAIS/METS Assessment project sought to explore in detail what is meant by compliance and to assess whether two existing systems – the UKDA and the system in place at TNA – can be meaningfully deemed 'compliant'. In addition to a comprehensive assessment of compliance, the project delivered a useful checklist for other institutions to use in assessing their own levels of compliance:

- A set of questions for OAIS compliance self-testing (pp 105 – 107)

SHERPA DP used the OAIS reference model to develop a persistent preservation environment for the SHERPA consortium, assigning rights and responsibilities and establishing protocols and work flow processes that will ensure the long-term preservation of the repository content. An objective of the SHERPA DP project was to ensure that institutional repositories could be considered OAIS compliant. The final SHERPA DP model is essentially a modified version of an OAIS. The services provided by institutional repositories are mapped to the OAIS reference model and notable omissions identified and allocated to the preservation service.

The project delivered a report on an OAIS compliant model for disaggregated services to be delivered as part of the SHERPA DP project, including an assessment of the suitability of the reference model as a practical model. The handbook contains a useful overview of the OAIS framework and discussion of OAIS compliance, based in part on the investigations carried out as part of the OAIS/METS project.

- SHERPA DP Handbook: an OAIS compliant model for disaggregated services ()

The PARADIGM workbook has extensive information on the OAIS model, including some examples of how the model can work in practice. The section does not relate extensively to long term preservation, though it is a useful overview with links to related initiatives including compliance and certification.

- PARADIGM workbook: section on the OAIS model

The DPTP included a module on OAIS. This was intended to provide trainees with a better understanding of the OAIS model and its applicability. The website provides an overview of the contents of the module. More precise information about course and learning materials is available on request from the DPTP project administrator.

- DPTP module: OAIS

PREMIS: PREMIS is a de facto metadata schema standard for preservation metadata. The PREMIS data dictionary specifies core metadata for verifying and tracking the provenance, authenticity and integrity of preserved digital assets.

PARADIGM explored the use of PREMIS metadata within METS files and the workbook contains a section on using PREMIS

- PARADIGM workbook: section on PREMIS

SHERPA DP adopted the PREMIS data dictionary as the basis for its preservation metadata schema. The project identified refinements necessary to support the operation of the preservation service. The broad requirements of PREMIS were further supplement by format-specific metadata (e.g. MIX for image metadata), as appropriate.

PRESERV explored the use of PREMIS within an IR context. It found that PREMIS appeared to provide an excellent basis on which assess the needs of IRs with respect to preservation metadata, and it was possible to map the PREMIS elements to an extended model incorporating preservation services and registries. However, more implementation and testing are required,

especially to validate the allocation of elements to preservation service providers and environment registries.

- [Preservation Metadata for Institutional Repositories: applying PREMIS](#)

Metadata Encoding and Transmission Standard (METS): The METS schema is a de facto standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using an XML Schema. The METS standard has been proposed as a suitable vehicle to support transfer and dissemination of objects for preservation by an external service provider, particularly in the case of complex objects (i.e. those comprising more than one file).

The OAIS/METS Awareness project considered the METS metadata standard, how it could be used further in a digital archive and its potential role at the National Archives and in the UKDA. The project found that METS documents (i.e. instances of METS in practice) are compatible with the OAIS model but that, particularly for born digital material, there was no clear benefit in implementing METS in the test repositories at this moment in time. The strengths and weaknesses of METS and its applicability within an archival environment are discussed in more detail in the project final report:

- [OAIS/METS Awareness final report: C7 Using the METS metadata standard in a digital archive](#) (pp 71 – 80)

The PARADIGM workbook section on administrative and preservation metadata contains a detailed introduction to METS and its use within a digital archive setting. The project developed specific guidance for users implementing METS in a digital archive setting, including advice on dealing with problems ingesting METS files into Fedora:

- Designing a METS Schema ()
- Sample METS Schema(s) for Personal Papers ()
- [Troubleshooting METS files and Fedora's Directory Ingest Service](#)

PRESERV released a METS export plug-in derived from work done by the Repository Bridge project, who implemented a METS export for EPrints 2. This was updated for the new plug-in architecture and data model in EPrints 3.

- [Complex-Object Export: METS plug-in](#)

MPEG-DIDL: Similarly to METS, the de facto standard MPEG-DIDL (Digital Item Declaration Language) has also been proposed as a suitable vehicle to support transfer and dissemination of complex objects for preservation by an external service provider. PRESERV developed an MPEG DIDL plug-in to package data for delivery to an external service that has since been integrated into Eprints v3.

- [Complex-Object Export: DIDL plug-in](#)

4. Training

Preservation: The DPTP course included numerous sections relevant to establishing and managing a repository, such as institutional repositories, file formats, OAIS, legal issues, metadata, outsourcing, and preservation approaches. The website only provides an overview of the contents of module. More precise information about course and its learning materials is available on request from the DPTP project administrator.

- [DPTP IR module overview](#)

METS: The METS Awareness Training project delivered a series of training events around the country that aimed to raise awareness in the UK higher education community of the METS standard. METS was devised to act as a generic container for all metadata associated with objects in a digital library environment. MAT training materials are a mixture of graphic presentations and documentary case studies. All are available from the project website.

- [MAT training materials](#)

5.1.7 Legal issues

Legal challenges are present across the life cycle, from active use to archiving, preservation, destruction and future access. The legal challenge of managing and preserving digital information was addressed in a number of projects. The most comprehensive of these was undoubtedly PARADIGM, which examined legal issues across the entire life cycle of digital private papers and records. Other projects were more focussed in their explorations: MANDATE explored legislation relevant to digital asset management systems and strategies; Managing Risk looked at the legal risks of failure to manage assets in line with legislation, and; the DPTP covered a number of fundamental legal issues as they relate to preservation. In all cases, UK legislation was taken as the basis for legal discussions.

- Public Records Acts
- Human rights
- IPR
- Access
- Defamation
- Environmental regulations

Public Records Acts. The PARADIGM project notes that Public Record Acts are important to repositories collecting private papers because the boundaries between public and private records are often blurred, and it briefly explores what this means in practical terms of accessioning and providing access:

- [PARADIGM workbook: Public Records Acts](#)

Human Rights: Data Protection; Human Rights, Privacy & Electronic Communications.

The PARADIGM workbook explores the impact of several pieces of legislation relating to human rights, including the 1998 Human Rights Act, the 1998 Data Protection Act, and the 2003 Privacy and Electronic Communications Regulations. Archives relate to Human Rights in two ways:

- As primary sources documenting the evolution of Human Rights and the memory of their abuse.

- The misuse of archives could potentially violate Human Rights, including Article 8 - Right to Respect for Private and Family Life.

Most personal archives, digital or otherwise, will contain personal data that is subject to the provisions of the Data Protection Act 1998. Finally, privacy and confidentiality concerns can affect how digital materials can be managed within the repository and how and when they can be made accessible to researchers.

- [PARADIGM workbook: Legal issues - Human rights](#)

The MANDATE toolkit explores the impact of the Data Protection Act on the design and operation of a digital asset management strategy and system. The Act can have a direct impact on the storage and management of information about assets when details about individuals involved in the production of an asset are stored. At John Wheatley College, personal data are not held in the system and the only area of potential concern relates to references to individuals in personnel and staffing committee meeting minutes: This is dealt with by existing practice which publishes selectively as appropriate. Issues of confidentiality arising in other minutes are monitored through general staff awareness of data protection issues, and specifically through training of administrative staff responsible for minutes.

- [MANDATE toolkit: Legislation for Data Protection](#)

The DPTP training module on legal issues includes information on the requirements of the Data Protection Act.

Intellectual Property Rights: Copyright, Patents, Trademarks, Design, Confidence & Trade Secrets

The PARADIGM workbook contains detailed explorations of how a number of Intellectual Property Rights can impact on the management and preservation of digital materials. Intellectual Property Rights (IPRs) are composed of a group of rights all relating to the protection of intellectual property. The main types of IPR are considered to be Copyright, Patent, Trademark, Design, Confidence & Trade Secrets. and Database Right. Copyright is a concern for all types of archives, whereas Patent, Trademark, Design and Confidence issues are most likely to be encountered when dealing with the archives of business and science.

The workbook asserts IPR as one of the main 'stumbling blocks' in the collection, curation, and provision of access to digital records and advocates an effective rights management policy and published 'take down' policy as the best way to address IPR challenges.

- [PARADIGM workbook: Legal issues - IPR](#)

The MANDATE toolkit explores the impact of Copyright legislation on the design and operation of an institutional digital asset management strategy and system. For most colleges the major issue in the application of this act to the digital asset management system will be in managing access to their assets and ensuring that they respect the rights of other asset producers.

The toolkit asserts that copyright for many college-produced assets (such as minutes) clearly lies with the college. However, the situation is not so clear cut for other assets, such as learning materials. The understanding of who owns learning materials and similar assets that are produced in connection with employment may be unclear and have to be explicitly addressed. Furthermore, when the system contains learning materials, there is the possibility that material belonging to an external party may have been included. Permission may be required for subsequent storage and/or re-use of such material.

- [MANDATE toolkit: Legislation for Copyright](#)

The DPTP training module on legal issues includes information on the requirements of the Data Protection Act.

Access: Freedom of Information Act (2000), Special Educational Needs & Disability Act (2001), Disability Discrimination Act (1995), Council of Europe Policy on Access to Archives

Access is a major driver in the preservation of digital materials and certain pieces of legislation must be addressed before access is granted. The PARADIGM workbook notes that recent policy and legislation marks a trend towards increasing access. Archivists must therefore strike a balance between such pieces of access legislation as the Freedom of Information Act, the Special Educational Needs Act, and Disability Discrimination Act, whilst simultaneously catering to privacy regulations.

- [PARADIGM workbook: Legal issues - Access](#)

The MANDATE toolkit contains information on FoI legislation, the Disability Discrimination Act, and the Special Needs and Disability Act. An overview of all three is provided, plus information on how the legislation impacts on the design or operation of an institutional digital asset management strategy and system. The toolkit contains a series of questions that institutions should consider in order to effectively and efficiently comply with all three pieces of legislation. It also lists a number of tools that may be suitable for checking compliance of the systems interface with the W3C Web Accessibility Initiative.

- [MANDATE toolkit: Legislation for Special Needs](#)
- [MANDATE toolkit: Legislation for Freedom of Information](#)

The DPTP training module on legal issues includes information on the requirements of the Freedom of Information Act.

Defamation: The UK Defamation Act 1996

The PARADIGM workbook explores how defamation legislation may impact on digital archives, particularly when access to such materials is granted. It recommends that institutions have a 'notice and take down policy' whereby

offending material can be immediately withdrawn from public access upon notice of a potentially libellous statement.

Environmental regulations: Environmental Information Regulations (2004)

The DPTP training module on legal issues covers the UK's Environmental Information Regulations and the impact they have on access and dissemination of environmental information.

Legal deposit

The DPTP training module on legal issues covers the UK's Legal Deposit Act and the impact this has for deposit of digital materials..

5.1.8 Case studies

Case studies are incredibly valuable. In some circumstances, they provide a basis on which to test and refine a given theory or approach. In others, they can be used to demonstrate to the wider community how an approach was implemented in a given context. This can give rise to several practical issues that may otherwise not be obvious to new implementers. In many cases, case studies can accomplish both of these outcomes, acting both as test cases AND as learning tools for others. Almost every project in this funding strand drew upon case studies during the course of the project. The prevalence of case studies within this strand illustrates a keen desire on the behalf of participants to encourage and assist in broader implementation of project outputs. Case studies covered:

- Digital asset management in an FE setting

- Digital asset management in an HE setting

- Costing of digital library collections

- Metadata implementation in digital libraries

- Archival collection development

- Developing a business case to support intangible assets

Digital asset management in an FE setting: MANDATE is an outstanding example of how case studies can function both as test scenarios against which to develop and refine an approach and also as practical examples of implementation from which others can glean valuable learning experiences. Each chapter of the MANDATE toolkit is supplemented by a case study description of implementation experiences at John Wheatley college that demonstrates how the toolkit was applied and shows how and why decisions were made to support any given activity. Textual descriptions are supported by graphics and tables, including screen shots of the prototype system and workflow models, and several links to supporting or college documentation are integrated into the study. MANDATE staff at John Wheatley intend to update the system once the system is actually up and running. The study is especially

valuable as it is the only instance where research is oriented particularly on an FE environment.

- MANDATE Case Study
  - Purpose of digital asset management
  - Development of digital asset management
  - Asset types
  - Metadata requirements
  - Metadata standards issues
  - Metadata quality issues
  - Workflow modelling and development
  - Interoperability issues
  - Metadata mappings
  - OAIS model
  - Preservation issues
  - Compliance with Freedom of Information legislation
  - Compliance with Data Protection legislation
  - Compliance with Copyright legislation
  - Compliance with Special needs legislation
  - Software issues
  - Training issues

Digital asset management in an HE setting: Similarly, the Managing Risk project provided a local case study that addressed a common and shared institutional need for a digital asset management strategy. The project's findings, although directed very much at an internal audience, are therefore highly relevant to the wider community and institutions facing similar challenges.

- Managing Risk project: final report

Costing of digital library collections: The LIFE project also drew upon case studies as a mechanism for implementing and testing their model and identifying actual costs for different types of digital objects in different scenarios. The case studies were chosen with the precise aim of challenging the LIFE methodology as robustly as possible. Although preservation activities were not undertaken in any of the case studies, the studies were fundamental in shaping the specific and required elements of the preservation aspect of the model.

- LIFE project
  - VDEP Case study (final report, pp 17 – 51)
  - Web archiving case study (final report, pp 52 – 63)

- UCL e-journals case study (final report, pp 64 – 87)
- VDEP Preservation model
- Web Arc preservation model

Metadata implementation in digital libraries: case studies were used as a valuable learning tool in the METS Awareness training project. The course included two types of case studies: the first comprising a presentation on METS implementation in a practical setting, such as at the National Library of Wales or Oxford Digital Library, and the second a fictional case study for participants to use in a practical exercise on implementing METS.

- METS in the Oxford Digital Library: A Case Study (ppt)
- Case study 1: A slide collection (doc)
- Case study 2: A collection of online books (doc)
- Case study 3: A collection of online texts (doc)
- Case study 4: A collection of videos (doc)

Archival Collection Development: The PARADIGM project carried out three case study/pilot projects to test and refine their knowledge of collection development for digital materials. This enabled them to make important recommendations on approaching and developing collection development policies and strategies, though the project ultimately recommended that more case studies are needed, particularly those that test different approaches or which address different domains. One primary case study was carried out, supplemented by two pilots. The primary case study involved identifying, capturing, and transferring records from contemporary politicians. This was supported by two pilot investigations. It led to a series of recommendations on how this valuable historical resource could be targeted for preservation. The study was supported by two complimentary pilot projects, one examining how political web resources could be targeted, the other exploring the issues in accessioning digital deposits from physical storage media. Project staff worked with the archive of Barbara Castle, which included two old PCs and a range of Amstrad disks – this was an archive deposited in the old-fashioned way (most of it is paper – some 500 boxes) and we worked with the BL on using forensic computing techniques to extract faithful images of the hard disks.

- Case study 1 - Hybrid personal archives of working politicians (final report, pp16 - 20)
- Case study 2 - Web archiving pilot (final report, p23)
- Case study 3 - Digital archaeology pilot (final report, p24)

Developing a business case to support intangible assets: A number of case studies were undertaken as part of the development of the ESPIDA approach, and as an aid to deeper understanding of the work and its applicability. The case studies demonstrate how the ESPIDA approach can be used in different scenarios and explore some examples that may hold some resonance for areas

within HE/FE. Three case studies were undertaken, testing the model on an e-theses collection, an IR, and a museums collection. The studies were created with the help of stakeholders within and external to the University of Glasgow. Whilst based on real life scenarios, the studies are for explanatory purposes only and are not actual business cases.

For the Institutional Repository study, the project team that had developed the University's own repository helped explore the benefits and so-called 'dis-benefits' that can arise from setting up a repository to manage and disseminate academic outputs. The cost elements of this study are based on an example from Edinburgh University Library.

The eTheses study examines how three different options can be appraised using the ESPIDA Scorecards. By presenting the options together, decision-makers can quickly compare the three options and understand the risks and benefits of each.

The final study looks at the cataloguing of archival holdings of a museum. This was developed with the help of external experts and offers a rich example, where benefits are quite slow to be realised.

- ESPIDA Institutional Repository case study (handbook, pp 36 – 42)
- ESPIDA eTheses case study (handbook, pp 43 – 48)
- ESPIDA museum archival collection case study (handbook, pp 49 - 54)

6 Appendices: Related initiatives

6.1 Digital Repositories programme (2005 – 2007)

DPAM is not the only programme to address preservation. The JISC Digital Repositories programme (also known as the 3-05 programme) included a number of projects that addressed preservation as part of a wider investigation into repository establishment and management. Furthermore, a number of international initiatives complement the work undertaken as part of the DPAM programme. These have not been included in the thematic overview as the DPAM programme is already incredibly diverse. Introduction of a second series of projects with a far less immediate application to preservation and asset management runs the risk of significantly diluting the coherence of the synthesis. These projects and initiatives are instead listed below.

The 2005 – 2007 Digital Repositories programme funded a number of clustered projects. Projects in the Repositories and Preservation cluster are as follows:

- SherpaPlus
- MIDESS
- EThOS
- Repository Bridge
- Spectra

SherpaPlus

SherpaPlus is an extension of the SHERPA (Securing a Hybrid Environment for Research, Preservation and Access) project funded under the FAIR (Focus on Access to Institutional Repositories) programme (2002 – 2005). The SHERPA project worked with twenty partner institutions to set up institutional open access e-print repositories that complied with the Open Archives Initiative (OAI) Protocol for Metadata Harvesting (OAI PMH). SherpaPlus is an extension of the SHERPA network to benefit the wider HE community and support implementation of institutional repositories outside of the original Sherpa partner institutions. It also explores issues arising as more experience is gained with institutional repositories and they become more mature.

SherpaPlus has a strong emphasis on advocacy and training, particularly in terms of events. A number of SherpaPlus road shows are planned for early 2007 in which speakers from SHERPA explore various key issues surrounding Open Access, institutional repositories, scholarly publications and authors rights. Through providing a centralised information point for advocacy, SHERPA Plus can facilitate development of advocacy work in the large number of institutions without repositories.

For the most part, preservation is the domain of another member of the SHERPA suite; Sherpa DP (see main report). However, Sherpa, and by extension SherpaPlus, have helped raise awareness within the repository

community of preservation issues, particularly the work undertaken by Sherpa DP, and promoted SHERPA advice and research into preservation prior to the formation of Sherpa DP. This includes:

- Requirements and Feasibility Study on Preservation of ePrints (Report commissioned by JISC, October 2003)

- The Digital Preservation of ePrints (D-Lib magazine, Sept 2003)

- Selection Criteria for the Preservation of ePrints (Project report, Feb 2004)

- Report on Preservation Standards (Project report, March 2004)

The team has also developed deposit licenses that give repositories the right to copy works for preservation purposes:

- Report on a deposit license for ePrints (Project report, June 2004)

MIDESS - Management of Images in a Distributed Environment with Shared Services

MIDESS explores the management of digitised content, particularly image and multimedia material, in institutional and cross institutional contexts. A pilot infrastructure has been established that can serve as a model for future national distributed repository activities, distributed across three locations (University of Leeds, University of Birmingham and London School of Economics). This enables partners to investigate opportunities for sharing and re-using data across institutions, interoperability issues within a network of different repository software implementations, and requirements for centralised metadata harvesting services.

Work package 5 explores the key issues in digital preservation and investigates how these can be practically addressed by digital repositories specialising in the storage of media such as sounds, images, and moving images. Specific metadata elements for the practical preservation of digitised multimedia material are developed or identified. Preservation features and functionality within different repository software are evaluated. The work package 5 report contains information on preservation strategies, formats, automatic metadata extraction, preservation functionality within the three selected implementations, and minimum preservation metadata.

- WP5: Digital Preservation Requirements Specification

Work package 2 comprised an evaluation of a broader range of repository software in order to investigate potential solutions to requirements identified as part of the WP3 User Needs Analysis. These documents also provide insight into preservation and asset management issues:

- WP2: Functional and Technical Requirements Specification

- WP3: User Needs Analysis

A broad overview of the project is available in a presentation given to ARLIS (Arts and Libraries Society) by Stephen Charles in July 2006

- MIDESS Presentation

EThOS – Electronic Theses Online Service

EThOS developed a prototype system to allow individuals to find, access, and archive electronic theses produced in UK HE institutions, using a UK database of theses with a central hub at the British Library. The full service will enable users to download the full text of the selected items in secure format and in the form intended by the authors.

Within the envisaged full service, institutions can submit their e-theses metadata or full content directly to the e-theses repository (Ethos central hub repository). Alternatively, EThOS can harvest the metadata or full content form institutional repositories. EThOS, courtesy of the British Library, also provides a cost-effective digitisation service for theses produced on paper.

Where thesis content is supplied to EThOS, the thesis and the metadata will be preserved by the British Library. This includes taking a copy of the thesis, storing it on a secure server, and applying any preservation techniques developed as part of the British Library's mandatory role in e-legal deposit for other document types e.g. e-journals, e-books, etc. There is no guarantee of preservation in cases where metadata only is supplied.

Explicit research into how this level of preservation can be achieved was not within the scope of the project, although preservation is mentioned in numerous sections of the EThOS toolkit. The toolkit shows how theses at a given institution can be accessed via EThOS, from the British Library or from an institutional (or consortium) repository.

- EThOS Toolkit

Sustainability of an EThOS service is covered within the Business Case section of the Toolkit. The need for cultural change to appreciate the asset value of e-theses is also discussed, highlighting the fact that an increase in the level of use of e-theses material is of significant benefit to the institution in terms of promoting its research profile.

Repository Bridge

The Repository Bridge project examined the interaction between a regional theses repository based at the National Library of Wales and pilot institutional repositories at University of Wales Aberystwyth and University of Wales Swansea. It developed repository software and tools allowing for full interaction between the DSpace or EPrints repositories and FEDORA open source systems to migrate and store items in a persistent manner.

The project did not address preservation per se but facilitated the transfer of materials to the National Library of Wales, after which preservation could be implemented. This is very similar to the EThOS project approach. The immediate relevancy of both projects is thus more in terms of asset management and capture than preservation.

Repository Bridge successfully demonstrated a preservation service approach for a specific type of content, electronic theses, showing how these materials

could be harvested using OAI and METS from Welsh IRs to a preservation repository at the National Library of Wales.

SPECTRa

SPECTRa - the Submission, Preservation and Exposure of Chemistry Teaching and Research Data – has developed a set of customized software tools to enable chemists to routinely deposit experimental data in Open Access digital repositories. Data is captured in an open format (wherever possible). Context specific metadata and persistent identifiers are added before the data is deposited in an escrow then institutional repository. Deposit in an OAI-compliant institutional repository is considered by the project as potentially an effective means of capturing, preserving, and disseminating the data in accordance with Open Access principles.

The project has contributed to preservation knowledge through exploration of the practical use of persistent identifiers (PIs) and transference of PIs between repositories. The cultural challenge of persuading researchers to deposit and thus preserve potentially sensitive or commercially valuable data is addressed through the idea of an escrow or dark archive; this enables researchers to deposit data straightaway whilst preserving exclusive access to it for a designated period of time.

6.2 Institutional Records Management programme (2003)

DPAM was part of a wider Digital Preservation and Records Management (DPRM) programme. DPRM funded another programme – the Institutional Records management programme – that aimed to help both FE and HE institutions implement institutional records management programmes that would meet the requirements of the Freedom of Information Act (2000) and conform to established good practice for the management of records and digital assets throughout their life cycle. Several of the projects included a digital preservation and asset management component.

A Flexible Framework for Institutional Records Management (AFFIRM)

Led by the City of Bristol College, this project developed a workbook to assess records management practices in an institution and how they may impact on an institution's ability to meet the government's compliance targets.

Developing records management programmes in FEIs

This project investigated how records management programmes can be developed to ensure the good management of records in all formats within all the FEIs in the North East of England through collaboration with the Northern Colleges Network.

Electronic records management training package

This project designed, developed and tested an ERM training package for all staff in HEIs and FEIs to provide a tool to help inculcate good practice. It also developed a web-based training package which provides a step-by-step assessment tutorial approach as well as the ability to search, pick and choose individual topics.

Establishing good practice in FE records management

Led by John Wheatley College, this project used the Model Action Plan (for achieving compliance with the Lord Chancellor's code of practice on the management of records) as an audit tool to audit institutional progress in establishing good practice for records management (paper and digital)

Managing employee records

This pilot study explored the management of employee records, defined the key record series in human resource management activities, and refined retention guidelines for the JISC *Study of the Records life cycle* model.

Managing primary research data & records

This project investigated practical implementation of the generic guidelines for managing the records of the research function from the *Study of the Records life cycle*. It also examined and documented the issues arising from this investigation and any solutions/strategies for addressing them.

Model action plan in a FE context

Led by Loughborough College, this project looked at the range of issues surrounding compliance with the Model Action Plan by analysing it and examining the degree to which it could realistically be applied to a typical FE college.

North Wales Records Management Consortium

This project researched the needs of Human Resources (HR) Departments' clients and produced best practice guidelines for the types of records which HR departments should be keeping and for how long.

Records life cycle for planning & strategy

This project tested the practical application of the planning and strategy-related themes from the Study of the Records life cycle by developing and commencing the implementation of a retention schedule for planning and strategy records created throughout the University of Edinburgh.

Records management and email

Led by Loughborough University, this project investigated email management practices within the University and developed generic policies for retention and disposal of institutional email.

Records management project for UWE

This pilot used the revised Study of the Records life cycle report (2003) as a model to produce an institutional retention schedule for the areas of Widening Participation and Governance.

Records management project for Warwick

This pilot used the revised Study of the Records life cycle report (2003) as a model to produce an institutional retention schedule for personnel management records.

Research Records

This project at the University of East Anglia piloted the research records retention section of the *Study of the Records life cycle* and explored issues surrounding the management and preservation of data in the electronic environment, particularly electronic research datasets.

Retention of electronic research data

This project created a retention schedule for primary research data in paper and electronic form, created at the London School of Hygiene and Tropical Medicine.

Student assessment records

This pilot case study tested the applicability of the generic retention findings relating to student assessment in the *Study of the Records life cycle* at Northumbria University.

Web records management project

The MERIT – Managing Electronic Records in Teaching – project identified good practice for managing web-based electronic records at Brunel University and developed a supporting retention and disposal schedule for records, regardless of their formats.

What is a student record?

This case study project explored the concept and content of a student record and made a number of recommendation surrounding the destruction and preservation of student record in data in paper and electronic form.

6.3 Related International initiatives

Projects had links with a number of international experts in their respective fields, most of who are acknowledged in project final reports. Within an international context, several projects reported awareness of similar initiatives being taken elsewhere in the world. The following list includes a number of initiatives project managers mentioned during interviews.

- NDIIPP – National Digital Information Infrastructure and Preservation Programme (USA)
- Cornell Digital Preservation Management workshops (USA)
- DELOS annual summer school: Digital Preservation in Digital Libraries (EU)
- HATII Information Management and Preservation Msc (UK)
- Stuttgart Academy of Fine Arts Conservation of New Media and Digital Information (MA) (DL)
- ASPR – Australian Partnership for Sustainable Repositories (AUS)
    - APSR/NLA RIFF – Repository interoperability Framework project
    - APSR PRESTA/PREMIS requirements specification project
- ARROW – Australian Research repositories Online to the World (AUS)
- JHOVE – JSTOR/Harvard Object Validation Environment (USA)
- DROID – Digital Record Object Identification (UK)

7 Appendix B: Project Questionnaire

7.1 4-04 projects

Complete list of deliverables & outputs

Please provide a full list of project outputs and deliverables, including publications and presentations, with date and location/publication details

1. How successful do you think the project has been in terms of:

   a) Achieving your project plan (1 – 5, 5 being highest)

   b) Reaching your intended audience

2. What would you say are the main themes/subjects addressed by your project? How have these themes been addressed?

   - Interoperability
   - Information management
   - Preservation (in what way, eg technical)
   - Legal issues
   - Policy issues
   - Cultural issues
   - Advocacy
   - Costs
   - Metadata
   - Software
   - Best practice
   - Training
   - Repositories
   - Assessment
   - Other (please specify)

3. What (or how) has your project contributed to preservation services and knowledge?

4. What do you feel are your most useful project outputs? Why?

5.What feedback have you received from your audience/stakeholders? How has this feedback been received/compiled?

6.Can you identify any areas where work did not go to plan, you experienced problems, or where you have not met your objectives?

7.What lessons (if any) did you learn in this project with regards to:

      a) project planning

      b) project management

      c) your subject area

8.Do you feel you had sufficient support from

      a) JISC

      b) other supporting bodies/projects (please name)

9.What steps have been taken to secure the longevity and sustainability of your main project outputs?

10.How effectively do you feel you contributed to the overall aim of the 4-04 programme to support institutional digital preservation and asset management (with a specific focus on long-term strategies and procedures)

11.Do you know of any similar work being done elsewhere, nationally or internationally? Are your staff involved in any of there projects?

Next steps

12.What further action, if any, is needed to further advance the issues explored in your project?

13.What further action, if any, do you feel is needed to take the advance the aim of the programme further?

7.2 3-05 projects

Please provide a full list of project outputs and deliverables, including publications and presentations, with date and location/publication details

1.What would you say are the main themes/subjects addressed by your project? How have these themes been addressed?

- Interoperability
- Information management
- Preservation (in what way, eg technical)
- Legal issues
- Policy issues
- Cultural issues
- Advocacy
- Costs
- Metadata
- Software
- Best practice
- Training
- Repositories
- Assessment
- Other (please specify)

2.What (or how) has your project contributed to preservation services and knowledge?

3.What steps have been taken to secure the longevity and sustainability of your main project outputs?

4.How do you feel you contributed to the overall aim of the 4-04 programme to support institutional digital preservation and asset management (with a specific focus on long-term strategies and procedures)

5.Do you know of any similar work being done elsewhere, nationally or internationally? Are your staff involved in any of there projects?

Next steps

6.What further action, if any, is needed to further advance the digital preservation or digital asset management issues explored in your project?

8 Appendix C: Interview schedule

8.1 4-04 projects

Project	Interviewee	Location	Date	Interviewer
DAAT	Kevin Ashley (PM)	ULCC	09-01-07	MP
DPTP	Kevin Ashley (PM)	ULCC	09-01-07	MP
ESPIDA	James Currall (PD)	U. Glasgow	12-01-07	BR
LIFE	Rory McLeod (PM)	BL	17-01-07	MP
Managing Risk	Patricia Methven (PM)	KCL	16-01-07	MP
MANDATE	Craig Green (PM)	JWCC	11-01-07	BR
MAT	Richard Gartner (PM)	U. Oxford	29-01-07	BR
OAIS/METS	Hilary Beedham (PM) Matthew Woollard	Telephone	13-02-07	BR
PARADIGM	Susan Thomas (PM)	U. Oxford	29-01-07	BR
PRESERV	Steve Hitchcock (PM) Jessie Hay (PM)	U. Southampton	12-01-07	MP
SHERPA DP	Gareth Knight (PM) Mark Hedges	AHDS	18-01-07	MP

BR – Bridget Robinson, DCC Research Officer.

MP – Maureen Pennock, DCC Research Officer.

8.2 3-05 projects

Project	Interviewee	Location	Date	Interviewer
ETHOS	Colin Galloway (PM)	Telephone	26-01-07	MP
MIDESS	Stephen Charles (PM)	Telephone	29-01-07	MP
REPOSITORY BRIDGE	Stuart Lewis (PM)	Telephone	07-03-07	MP
SHERPA PLUS	Bill Hubbard (PM)	Telephone	30-01-07	MP
SPECTRA	Alan Tonge (PM)	Telephone	29-01-07	MP

8.3 Other interviewees

Interviewee	Reason	Location	Date	Interviewer
Helen Hockx-Yu	Programme Manager: initial interview	London	18-01-07	MP
Helen Hockx-Yu	Programme Manager: follow-up	Telephone	23-03-07	MP

9 Author contact details

Maureen Pennock
UKOLN
University of Bath, Bath BA2 7AY

Tel: 01225 386 711
Email: m.pennock@ukoln.ac.uk