Sherpa-DP and OAIS

Andrew Wilson
Preservation Services and Projects Manager
AHDS
4 April 2006
SHERPA DP Project

- **Acronym:** Securing a Hybrid Environment for Research Preservation and Access: Digital Preservation
- **Development Partners:** AHDS at King’s College London (Lead), Nottingham, Glasgow, Edinburgh, White Rose Consortium, London Leap Consortium
- **Duration:** 2 years, March 2005 – February 2007
- **Funding:** JISC and CURL
- **Programme:** JISC Digital Preservation and Records Management Programme
Sherpa DP Project

Purpose:
To create a collaborative, shared preservation environment for the SHERPA project framed around the OAIS Reference Model.

Aims:
1. To develop a prototype preservation environment for SHERPA Partners based on the OAIS reference model including a set of protocols and software tools.
2. To establish a workflow & procedures to suit the needs of institutional repositories and the preservation service.
3. Provide guidance on the ingest process, to encourage the deposit of formats that will minimise long-term operational costs.
4. To develop an exemplar for an outsourced preservation service.
5. Create a User Guide that recommends standards, best practice, protocols and processes that may be used in the management, preservation and presentation of e-print repositories.
OAIS Functional Model

- **Preservation Planning**
- **Data Management**
- **Archival Storage**
- **Access**
- **Ingest**
- **Descriptive Info**
- **Administrative**
- **Management**

- **PRODUCER**
  - SIP
  - AIP

- **CONSUMER**
  - DIP
  - result sets

- **queries**
- **orders**

**Archival Storage**

- Descriptive Info
- AIP
- AIP
- AIP
OAIS Functional Entities

- **Ingest:** services and functions that accept SIPs from Producers; prepares AIPs for storage, and ensures that AIPs and their supporting Descriptive Information become established within the OAIS.

- **Archival Storage:** services and functions used for the storage and retrieval of AIPs.

- **Data Management:** services and functions for populating, maintaining, and accessing a wide variety of information.

- **Administration:** services and functions needed to control the operation of the other OAIS functional entities on a day-to-day basis.

- **Preservation Planning:** services and functions for monitoring the OAIS environment and ensuring that content remains accessible to the Designated Community.

- **Access:** services and functions which make the archival information holdings and related services visible to Consumers.
OAIS Compliance

• What is it?
• How do you achieve it?
• How do you measure it?
• How do you compare across implementations?
• Should there be an audit process to assess/certify compliance?
SHERPA-DP OAIS Model
(WP2.2 – 2.12)

Aim: apply the OAIS reference model to the distributed preservation service proposed by the SHERPA DP project

How:
1. Map the six entities of an OAIS-compliant repository (Ingest, Archival Store, Administration, Data Management and Access) onto an existing structure.
2. Ensure that domain-specific terminology can be mapped to the OAIS equivalent
3. Identify rights and responsibilities, services and actions, and apportion these between the IR and preservation repository service
Why disaggregate preservation functions?

Why disaggregate preservation functions?
- institutional repositories lack the time to implement preservation
- scarcity of staff with necessary preservation skills and expertise
- seeking to remove repetition of services
- potential cost savings in terms of staff time and equipment?
- preservation is not inherent in most repository software
- DSpace and EPrints software primarily about submission, basic storage and access
High Level View of Sherpa-DP Repository Landscape
Source: Feasibility and Requirement Study
On the Preservation of E-Prints
Establishing responsibility

- **Who is responsible for creating the AIP?**
  - Preservation service, Institutional repository, both?

- **What type of information is created?**
  - Descriptive, technical, structural & administrative metadata, migrated resource

- **How will it be used?**
  - Identification of at-risk formats, migration

- **When will they create it?**
  - On ingest, schedule, or when the resource is at-risk
Establishing responsibility: Institutional Repository

- Implement appropriate repository software
- Develop selection, retention and ingest policies
- Develop a rights framework
- Specify a minimum metadata set, and provide details to the Preservation Service
- Quality control for descriptive metadata
- Support mechanisms for metadata harvest
- Support for extension schemes to enable preservation.
- Creation of technical metadata (possibly)
- Alerting mechanisms for updated/additional content?
Establishing responsibility: Preservation Service

Storage:
• Provide a permanent storage facility and disaster recovery capabilities
• Manage storage hierarchy

Preservation Planning:
• Evaluate contents of archive and undertake risk assessment
• Develop recommendations for preservation standards and policies
• Life cycle management. Monitor changes in technology environment, users’ service requests, and knowledge base

Preservation Action:
• Develop and implement migration plans
• Create and manage multiple copies of content, including off-site storage
• Record appropriate information on any changes
OAIS Functional Model

SIP = E-print & discovery MD

AIP = E-print, discovery & preservation MD

DIP = E-print and discovery MD

Legend

- Preservation Service
- E-Print Archive
Information Package & METS

- Information Package (SIP, AIP & DIP) is the primary object within the OAIS model.
- Review existing metadata captured by repositories.
  - Discovery metadata
  - Minimal Preservation metadata
- Identify additional metadata required for preservation and capture methods
  - Technical information, audit trails
- Review the potential for METS within the SHERPA-DP environment
  - As a framework for combining and packaging metadata
  - As a transfer mechanism for metadata and e-prints
What will we store in the AIP?

Legend

- Created by Depositor
- Created by IR
- Created by PS
- Created by IR + PS

Archival Information Package

Content Information

Digital Object

- Deposited data
- Archival data

Technical Metadata

- Deposited data
- Archival data

Resource Discovery

Rights

Relational Metadata

Provenance Metadata

Fixity Information

IR = Institutional Repository
PS = Preservation Service
**Simplified Workflow (preliminary)**

1. **Deposit E-print**
2. **Validation successful**
   - **Yes**
     - **Is metadata complete?**
     - **Yes**
       - **Enhance Metadata**
       - **Copy SIP to repository store**
       - **E-print in appropriate deposit format**
       - **Migrate to dissemination format**
       - **Copy DIP to repository store & dissemination server**
       - **Transfer schedule**
       - **Update availability status in catalogue**
       - **Make available to Researcher**
   - **No**
     - **Request Resubmission**
3. **No**
4. **Risk assessment reveal problems?**
   - **Yes**
     - **Implement migration strategy**
   - **No**
     - **Obsolescence Monitoring Schedule**
5. **Deposit format Obsolete**
6. **Generate AIP**
7. **Transfer AIP to Preservation store**
8. **Create Preservation metadata**
9. **Implement migration strategy**
10. **Generate AIP**
11. **Create new dissemination format**
12. **Transfer AIP to Preservation store**
Data Transfer

- Investigate methods to identify new submissions.
- Implement transfer mechanisms between institutional repositories and Preservation Service (DSpace and Eprint APIs, storage layers and module add-on capabilities)
- Examine the capabilities of OAI-PMH for complex object formats
Sustainable Preservation Actions

Investigate the processes required to enable changes and updates to e-print content that ensures their long-term integrity and preservation

Create/refine automated tools to perform:
• obsolescence checking and migration services
• Mechanism for establish and track versions
• integrity checking and reporting
What’s to come…

• Provide a generic model that may be applied to other Preservation Services
• Establish a workflow and procedures to suit the needs of institutional repositories and the preservation service
• Provide guidance on the ingest process, to encourage the deposit of formats that will minimise long-term operational costs.
• Develop/refine software tools to enable communication between the Institutional Repository and Preservation Service.
• Create a User Guide that recommends standards, best practice, protocols and processes that may be used in the management, preservation and presentation of e-print repositories
Further Information

**URL:**
http://www.ahds.ac.uk/about/projects/

**Contact**
andrew.c.wilson@ahds.ac.uk

gareth.knight@ahds.ac.uk