A disaggregated model for preservation of E-Prints

Gareth Knight
SHERPA DP Project
Arts and Humanities Data Service

‘E-prints’

‘E-prints’ = a digital duplicate of an academic research paper that is made available online as a way of improving access to the paper.

Document types:

- pre-prints, post-prints
- Journal articles, conference papers, book chapters or other research output.
- **Properties**
  - Textual format that can be created or converted by word processing software
  - Emphasis placed upon ease of use
- **Formats**

  PDF, HTML, MS Word, Postscript, RTF

ahds

SHERPA

Institutional Repository

An institutional repository "….is a set of services that an institution offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organisational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organisation and access or distribution."

Lynch, C., ARL Bimonthly Report 226, http://www.arl.org/newsltr/226/ir.htm

ahds

SHERPA

Construction of repositories

An initial emphasis placed upon the construction of repositories:

- JISC FAIR (Focus on Access to Institutional Resources) programme funded several projects.
- SHERPA (Securing a Hybrid Environment for Research Preservation and Access) funded for 3 years (November 2002 – November 2005) with the aim of constructing a series of institutional OAI compliant e-print repositories
- “Forget about OAIS for now! The OAI-compliance of the Eprint Archives is enough for now.”

  Stevan Harnad, September98 forum, 13 February 2003

ahds

SHERPA

SHERPA DP Project

- **Acronym:** Securing a Hybrid Environment for Research Preservation and Access: Digital Preservation
- **Development Partners**: AHDS (Lead), Nottingham + Edinburgh Research Archive, Glasgow E-Prints Service, White Rose University Consortium & London LEAP
- **Aims**:
  - To develop a persistent preservation environment for SHERPA Partners based on the OAIS reference model including a set of protocols and software tools
  - To develop an exemplar for an outsourced preservation service
  - To explore the technical and organisational requirements of an outsourced preservation service,
  - use of METS for packaging and transferring metadata and content

SHERPA

A Disaggregated Service

- Digital preservation could be seen as one of these ‘value-added’ services, and may not necessarily be performed by the institution.

  JISC Continuing Access and Digital Preservation Strategy 2002-5 (Beagrie, 2002, p. A13).

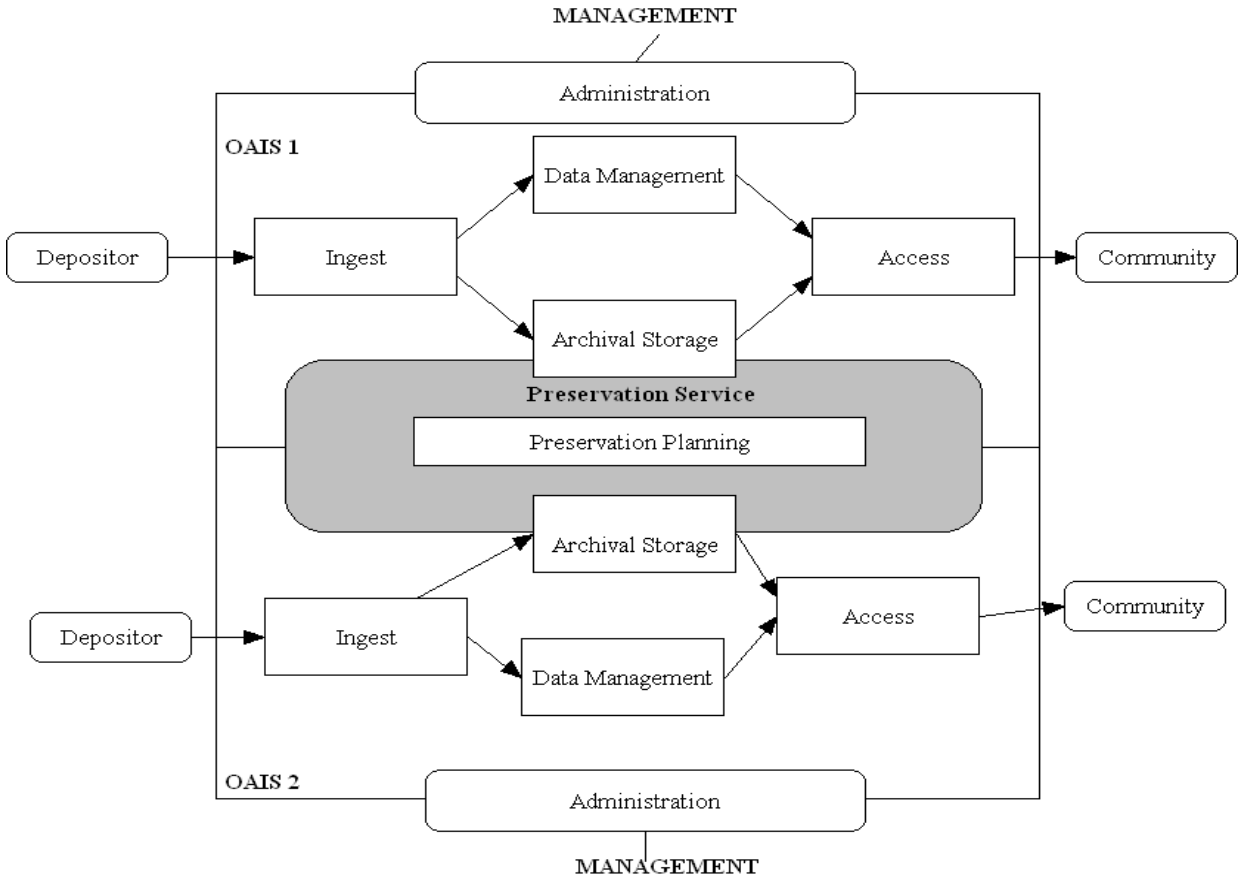
Reasons:

- Preservation is not inherent in most repository software
- DSpace and EPrints software primarily about submission, basic storage and access
- 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?

ahds

SHERPA

OAIS Functional Model



ahds

SHERPA

Technical Infrastructure

- Investigate technologies required to enable changes and update e-print content
- Create services to remotely monitor and report on:
  - Integrity
  - Obsolescence
- Investigate mechanisms for automatic creation of new versions, migration and redeposit

ahds

SHERPA

Transfer Mechanisms

- Investigate and implement automated transfers of data between institutional repositories and preservation repository
- Review DSpace and Eprint APIs, storage layers and module add-on capabilities
- Examine the capabilities of OAI-PMH for complex object formats
- Prototype and test SRB as a common storage medium
- Prototype and test API based access mechanisms
- Test external synchronisation mechanisms

ahds

SHERPA

OAIS Information Packages

- A container that encapsulates Content Information and Preservation Description and other metadata.
- Packages for submission (SIP), archival storage (AIP) and dissemination (DIP)
- AIP = "... a concise way of referring to a set of information that has, in principle, all of the qualities needed for permanent, or indefinite, Long Term Preservation of a designated Information Object"
  - M Day, 2002

ahds

SHERPA

What about metadata?

- Review existing metadata captured by repositories.
  - Discovery metadata
  - Minimal Preservation metadata
- Identify additional metadata required for preservation and capture methods
  - Technical, provenance metadata
- Review the potential for the use of METS within the SHERPA environment
  - As a framework for combining and packaging metadata
  - As a transfer mechanism for metadata and e-prints

ahds

SHERPA

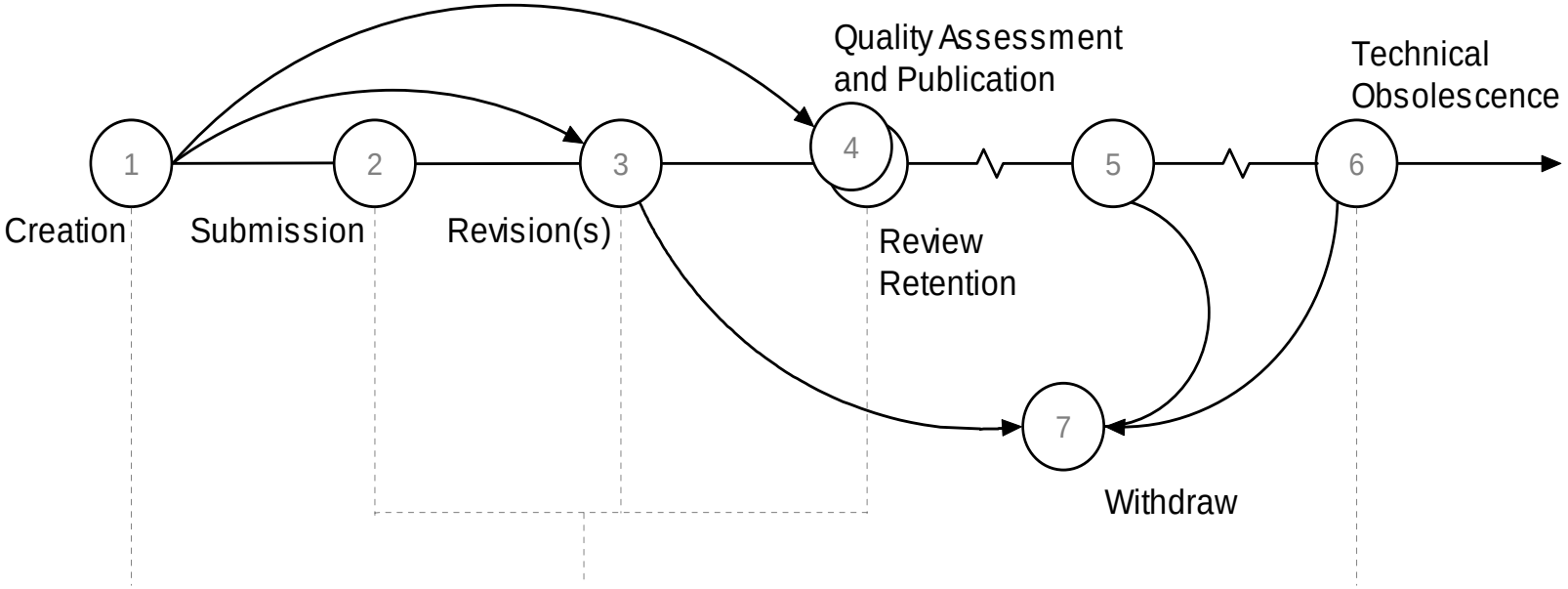
Establishing responsibility

- 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
- When will they create it?
  - On ingest, schedule, or when the resource is at-risk
- How will it be used?
  - Identification of at-risk formats, migration

ahds

SHERPA

E-Print Lifecycle



Source: Feasibility and Requirement Study
On the Preservation of E-Prints

Scenarios: Factors

1) Notification of new or updated resource
   - Repository notify preservation service
   - Preservation Service monitor repository and transfer e-prints (e.g. OAI-PMH SETs)

2) Timetable for transfer to Preservation repository
   - Transfer on ingest/update
   - Scheduled transfers (weekly, monthly transfer of new Information Packages)
   - Transfer when considered to be at-risk

3) Timetable for Migration:
   - Migrate on ingest
   - Generate technical metadata on ingest and migrate when it is considered at-risk.

ahds

SHERPA

Institutional Repository: Responsibilities

Current responsibilities

- Provide a method to accept, store and deliver e-prints.
- Intellectual Property Rights
- Quality control for descriptive metadata

Additional Requirements

- Publish metadata to be harvested
- Support for extension schemes to enable preservation.
- Creation of technical metadata
- One or more methods for transferring content across the network
- Alerting mechanisms for updated/additional content?

ahds

SHERPA

Preservation Service: Responsibilities

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:

- Implement migration plans and convert holdings as appropriate
- Create and manage multiple copies of content, including off-site storage (i.e. Manage version control)
- Record appropriate information on any changes

Moving forward…

- 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.
- 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

ahds

SHERPA

Thank You

Questions?

SHERPA