

A Shared Preservation model for Institutional Repositories

SHERPA DP Project

Gareth Knight

Arts and Humanities Data Service

JISC Joint Programmes Meeting, 7th July 2005

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

- **Aims:**
 - To develop a prototype preservation environment for SHERPA Partners based on the OAIS reference model including a set of protocols and software tools
 - To establish a workflow & 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.
 - To develop an exemplar for an outsourced 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

Why disaggregate preservation functions?

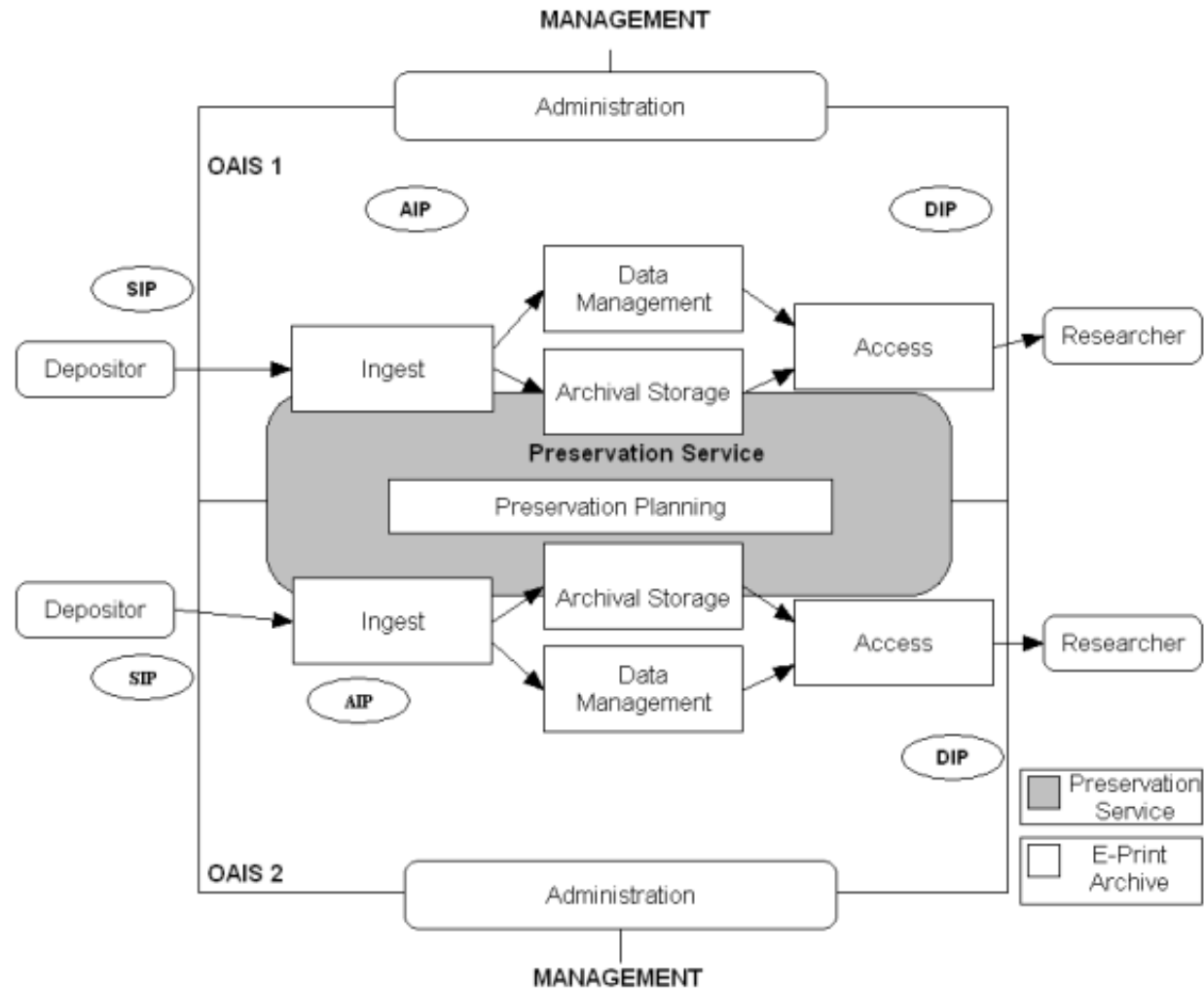
“Institutional arrangements may benefit most from third-party or common services being developed to support preservation planning or remote storage”

JISC Continuing Access and Digital Preservation Strategy 2002-5 (Beagrie, 2002, p. 13).

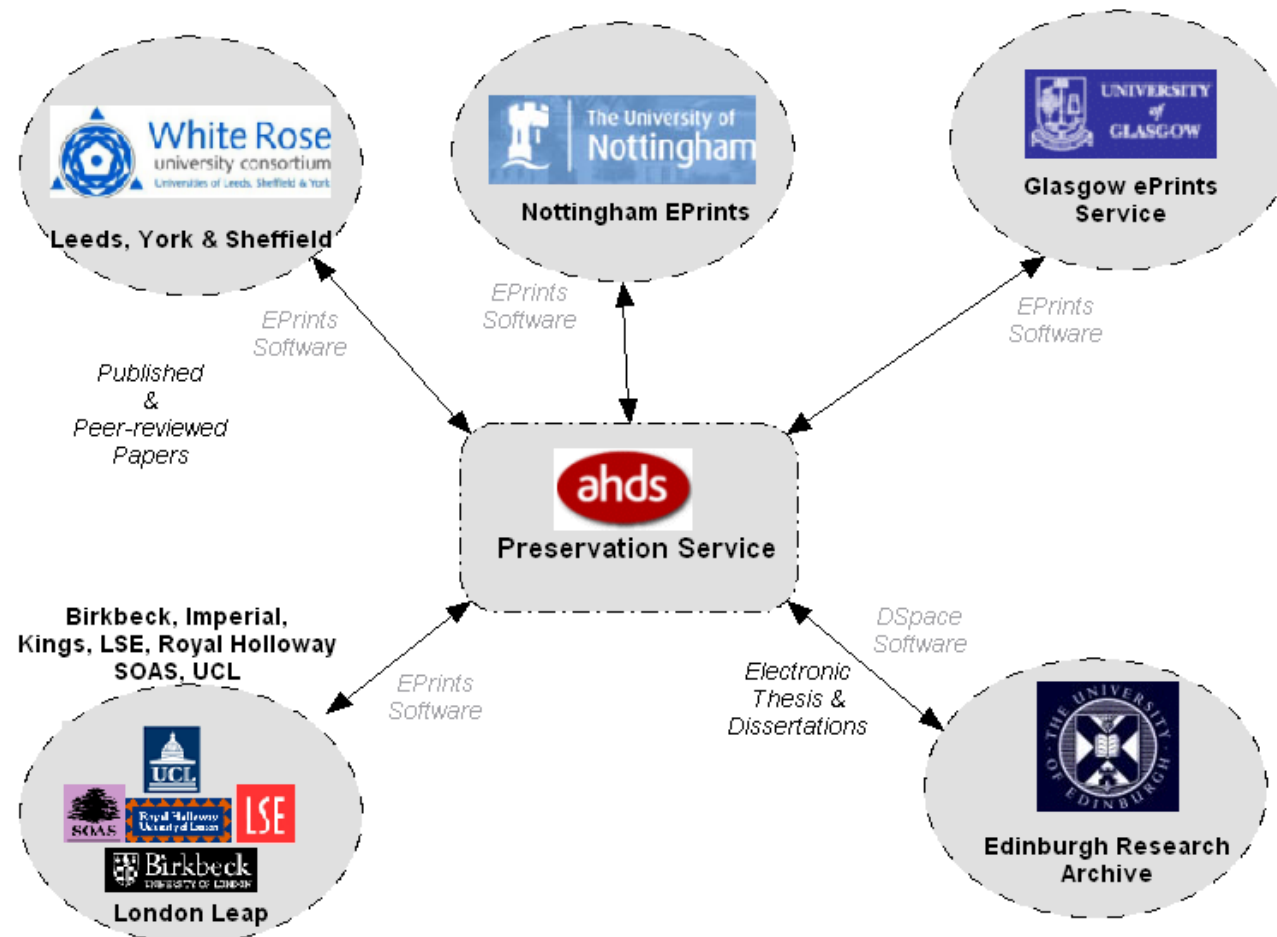
Why disaggregate preservation functions?

- 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

OAIS Functional Model



Repository Landscape



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)
- Prototype and test SRB as a common storage medium
- 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:

- integrity checking and reporting
- obsolescence checking and migration services
- Mechanism for establish and track versions

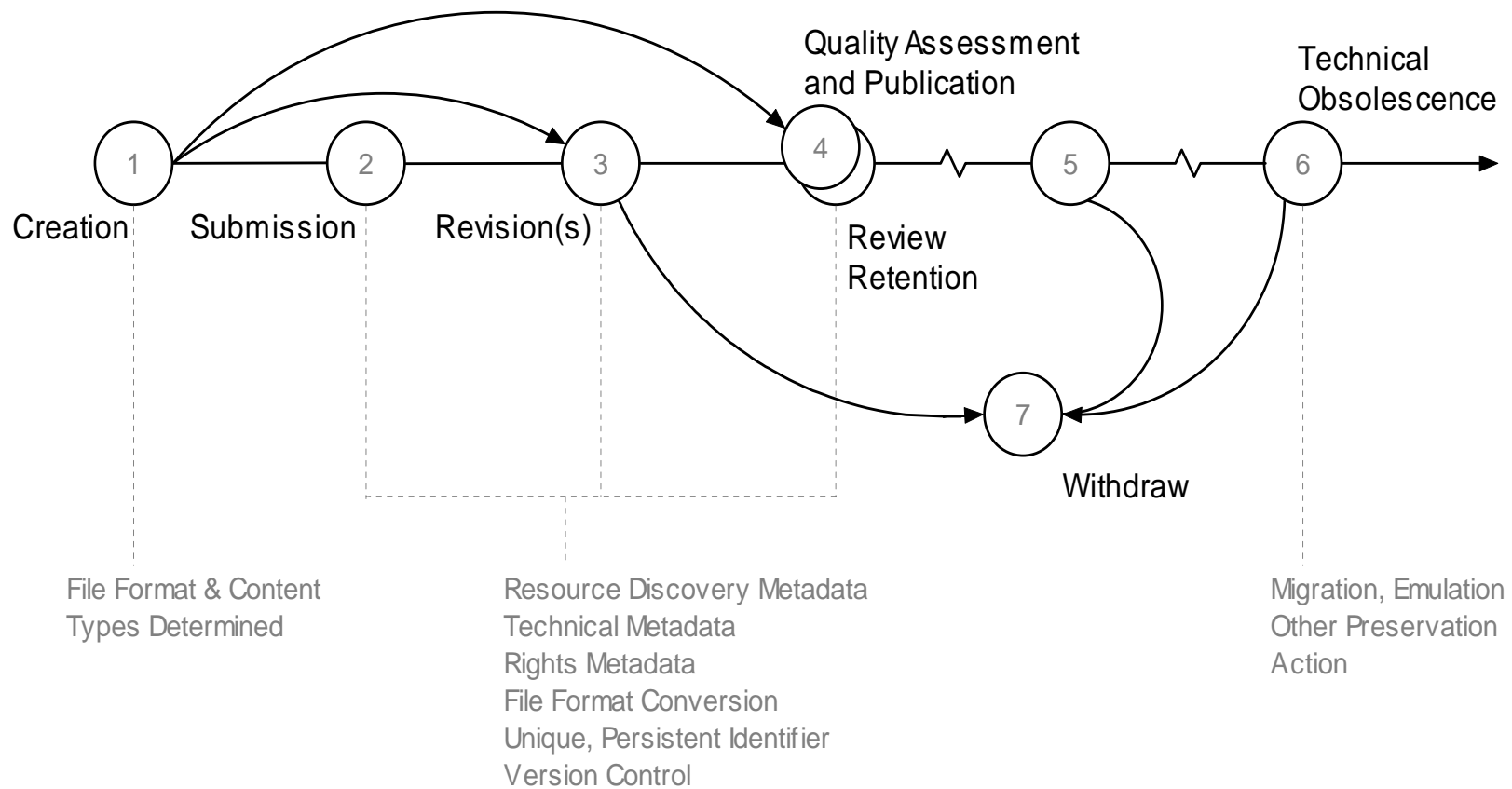
Information Package & METS

- 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 environment
 - As a framework for combining and packaging metadata
 - As a transfer mechanism for metadata and 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

E-Print Lifecycle



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

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

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.
- 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