

# Sherpa-DP and OAIS

Funded by:

 JISC



Arts & Humanities  
Research Council

© AHDS

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

Funded by:

JISC



# 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

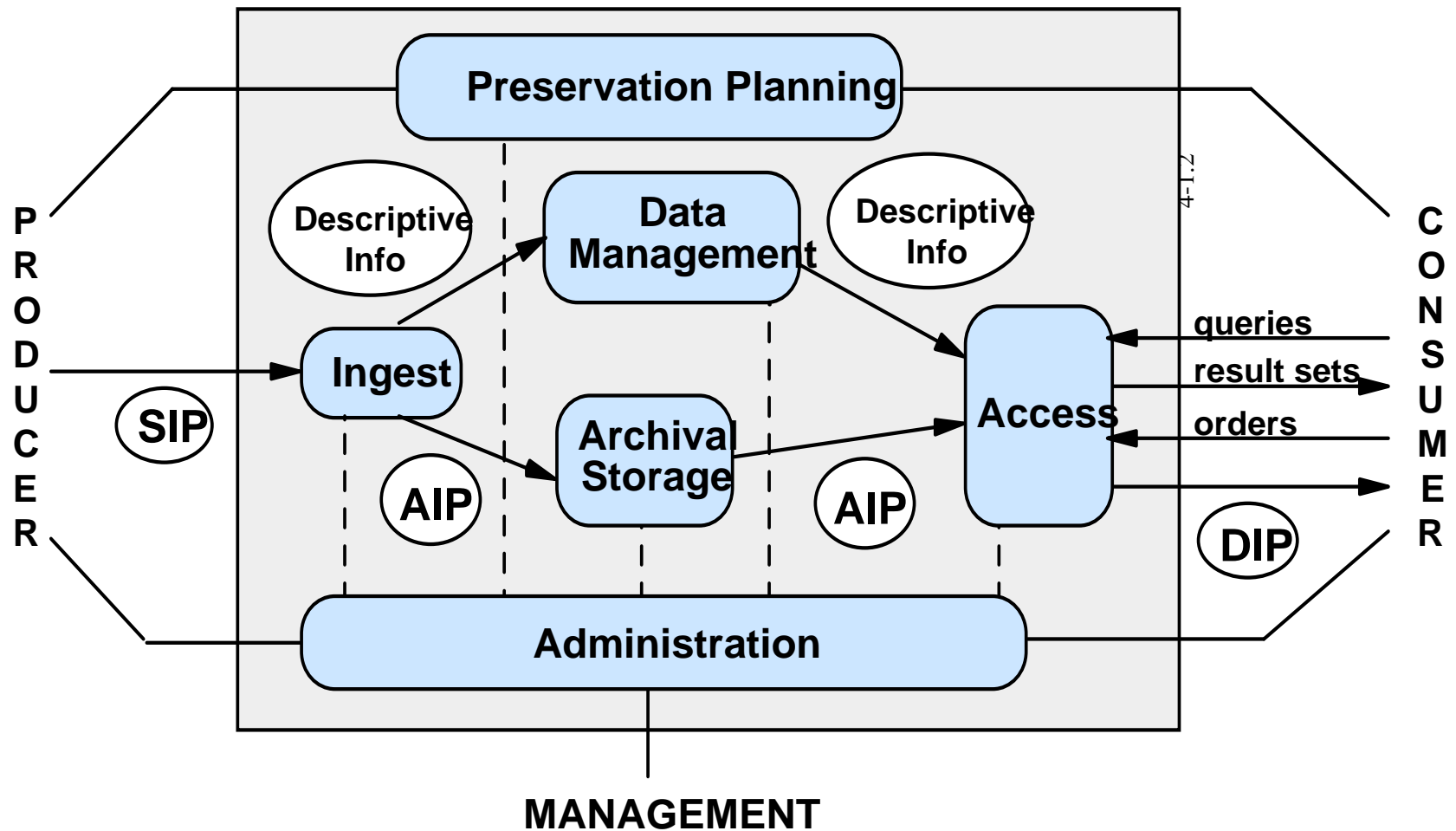
Funded by:

JISC



© AHDS

# OAIS Functional Model



Funded by:

JISC

Arts & Humanities  
Research Council

© AHDS

# 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

Funded by:

JISC



© AHDS

# OAIS Compliance

- What is it?
- How do you achieve it?
- How do you measure it?
- How do you compare across implementations?
- Should there be a audit process to assess/certify compliance?

Funded by:

JISC



Arts & Humanities  
Research Council

# 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

Funded by:

 JISC

 Arts & Humanities  
Research Council

© AHDS

# 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

Funded by:

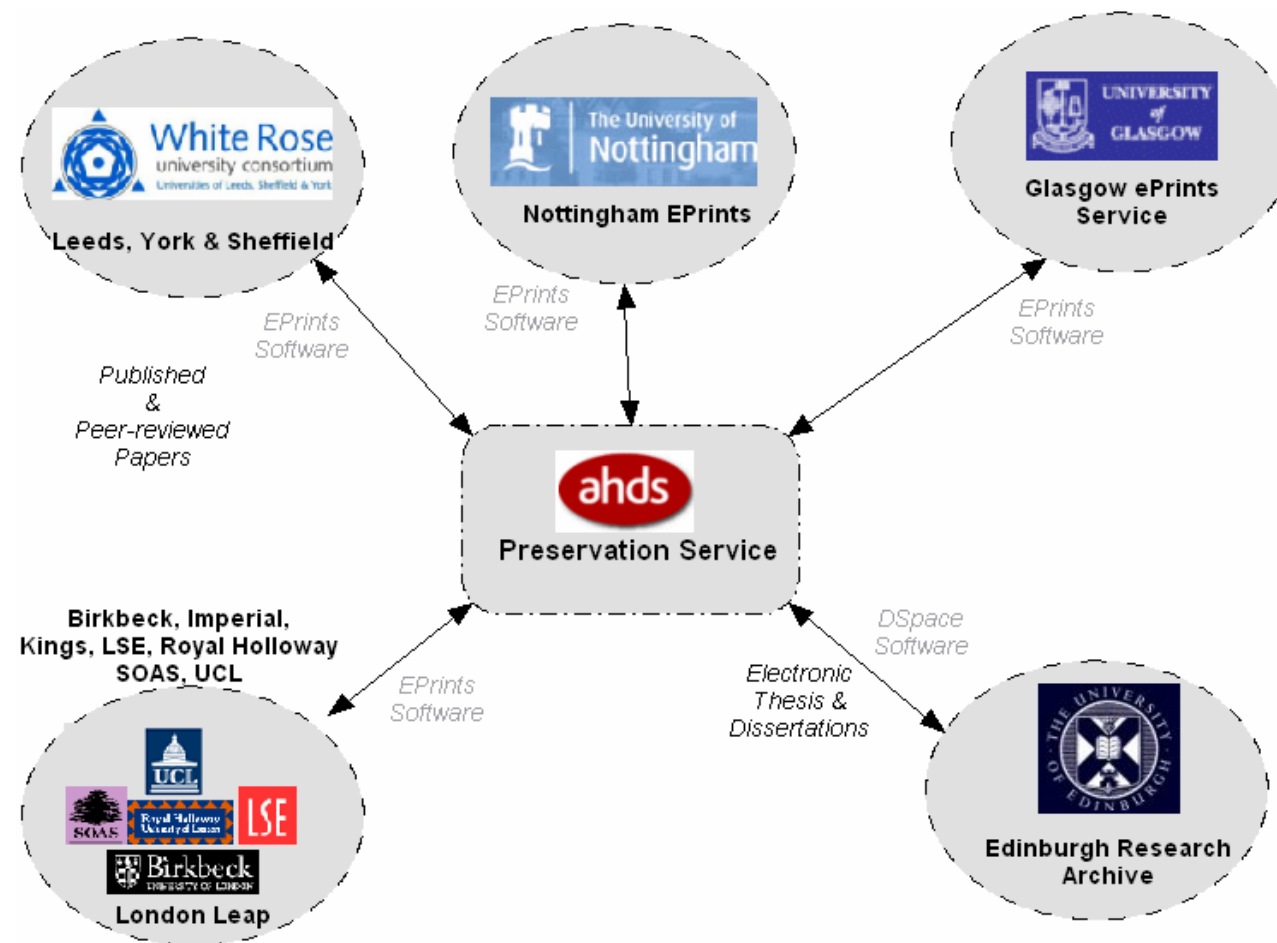
JISC

Arts & Humanities  
Research Council

© AHDS



# High Level View of Sherpa-DP Repository Landscape



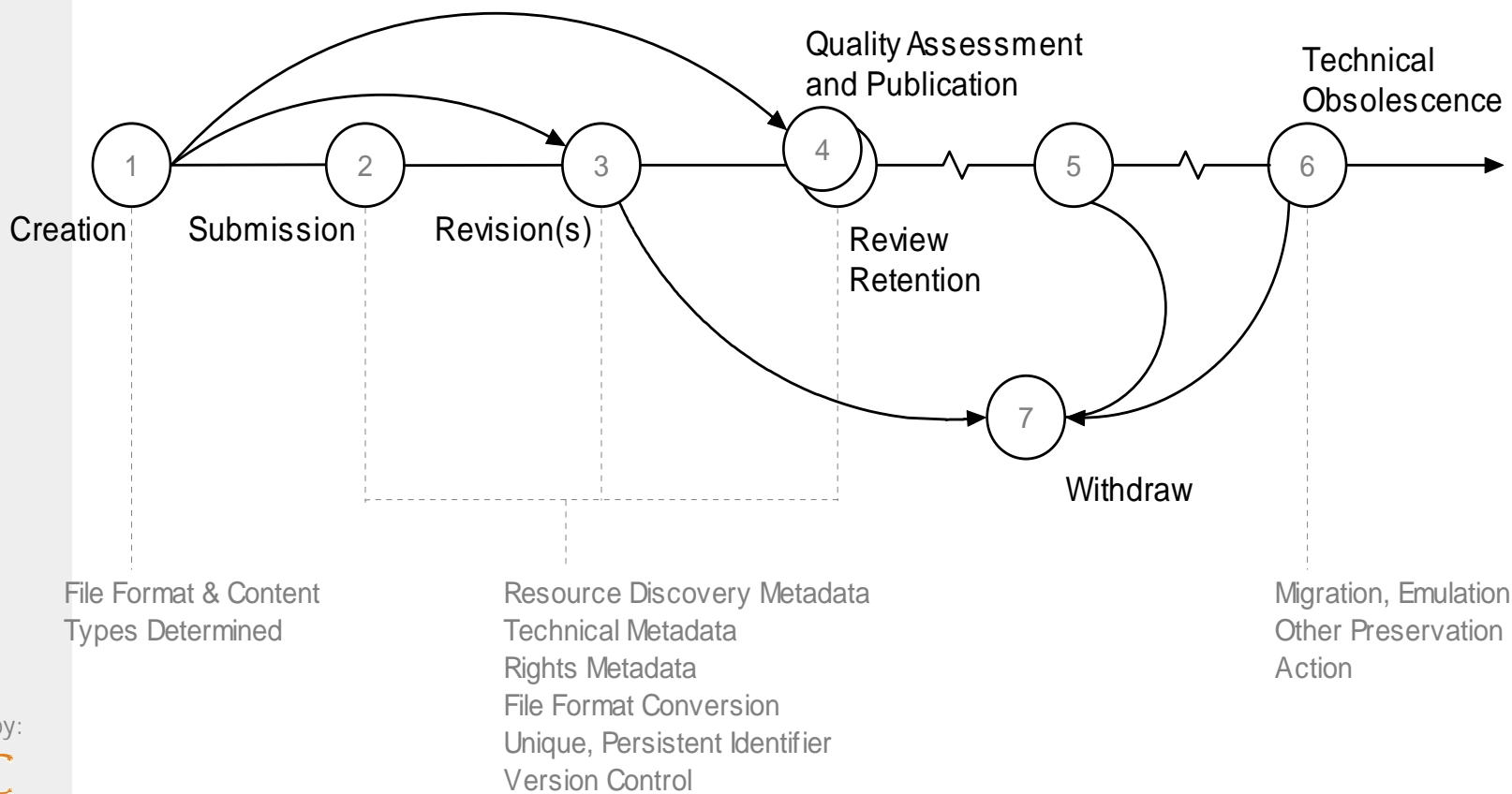
Funded by:



Arts & Humanities  
Research Council

© AHDS

# E-Print Lifecycle



Funded by:



© AHDS

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

Funded by:

JISC



# 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?

Funded by:

JISC



# 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

Funded by:

JISC



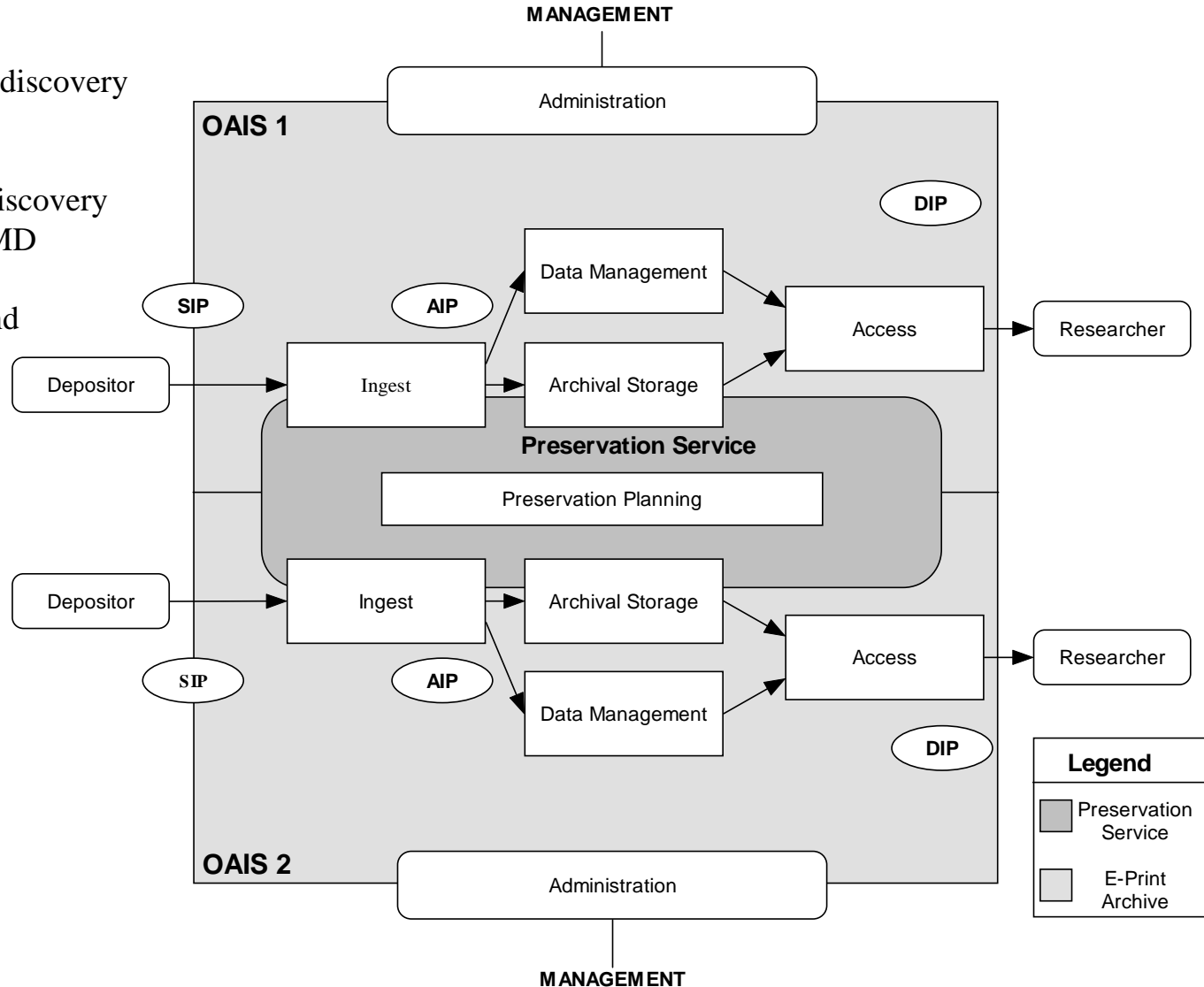
Arts & Humanities  
Research Council

# OAIS Functional Model

**SIP** = E-print & discovery  
MD

**AIP** = E-print, discovery  
& preservation MD

**DIP** = E-print and  
discovery MD



Funded by:

**JISC**

Arts & Humanities  
Research Council

© AHDS

# Information Package & METS

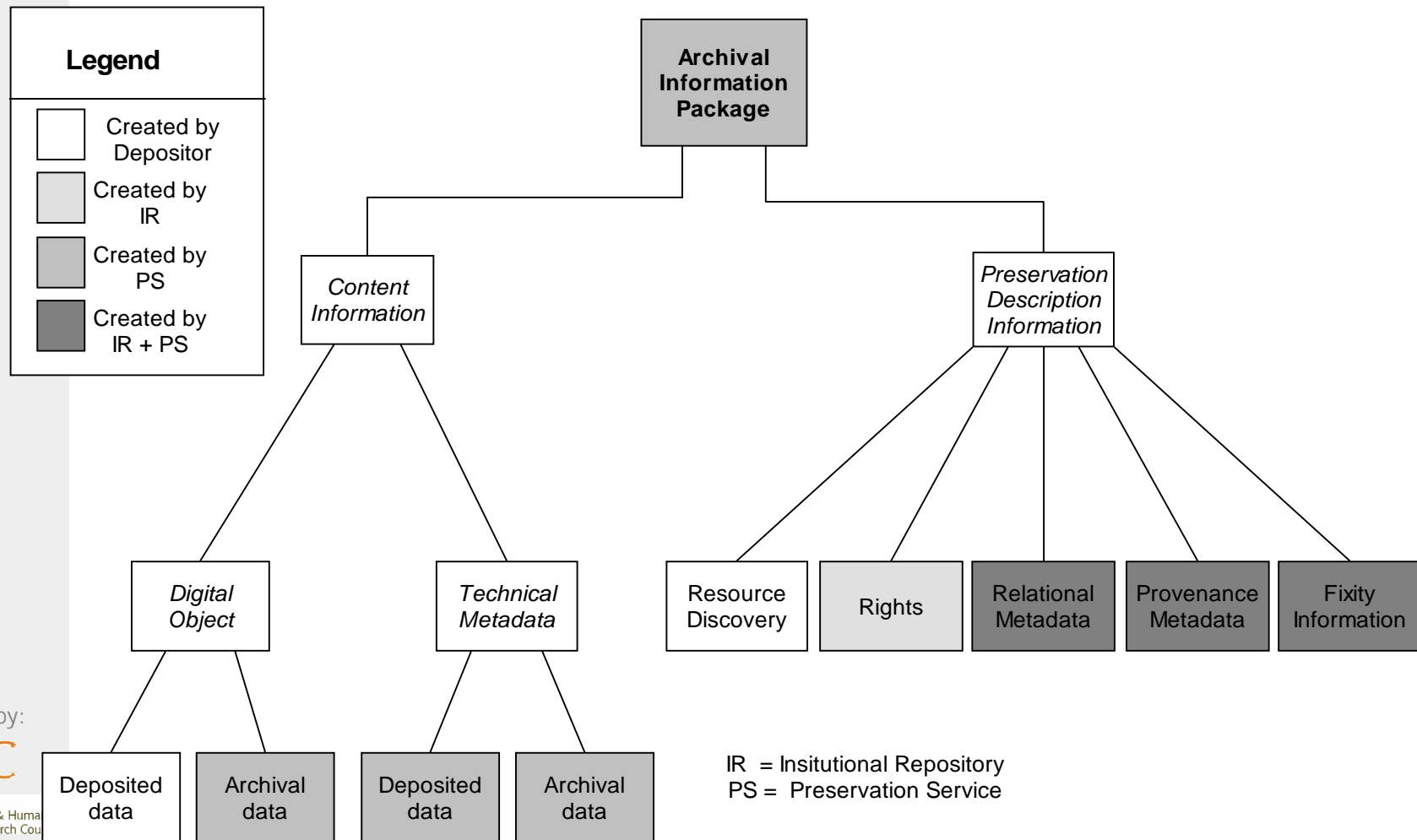
- Information Package (SIP, AIP & DIP) is the primary object within the OAIS model.
- Review existing metadata captured by repositories.
  - Discovery metadata
  - Minimal Preservation metadata
- Identify additional metadata required for preservation and capture methods
  - Technical information, audit 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

Funded by:

JISC



# What will we store in the AIP?

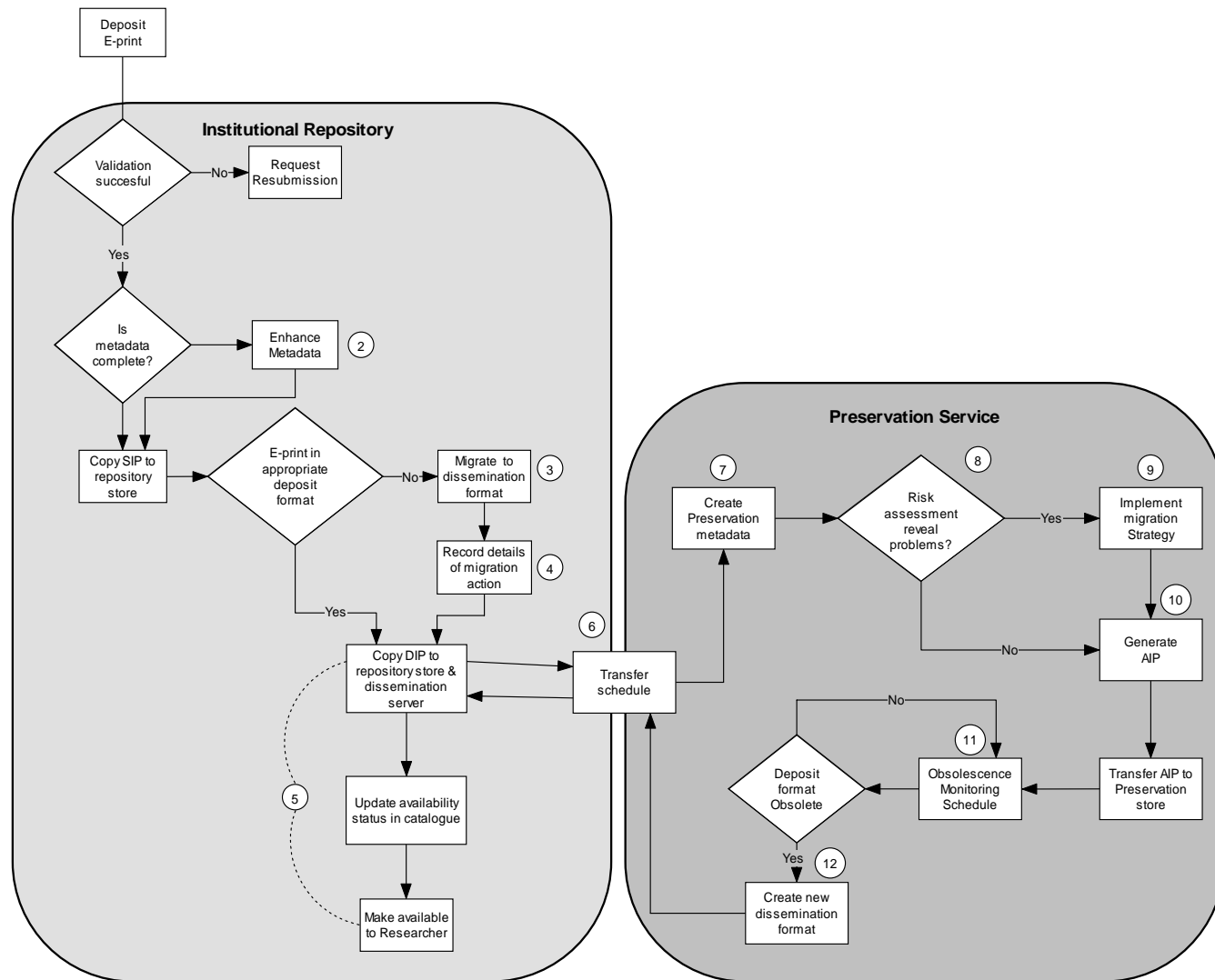


Funded by:





# Simplified Workflow (preliminary)



Funded by:

JISC

Arts & Humanities  
Research Council

© AHDS

# 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

Funded by:

JISC

Arts & Humanities  
Research Council

# 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

Funded by:

JISC



# 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

Funded by:

JISC



© AHDS

# Further Information

## URL:

<http://www.ahds.ac.uk/about/projects/>

## Contact

[andrew.c.wilson@ahds.ac.uk](mailto:andrew.c.wilson@ahds.ac.uk)

[gareth.knight@ahds.ac.uk](mailto:gareth.knight@ahds.ac.uk)

Funded by:

 JISC

 Arts & Humanities  
Research Council

© AHDS