



A problem shared...

Modelling OAIS compliance for distributed services

Funded by:

JISC



Arts & Humanities
Research Council

© AHDS

Gareth Knight
Digital Preservation Officer
AHDS

22 November 2006

Overview

- Introduction to SHERPA DP
- Reasons to distribute services
- Current IR setup & workflow
- Distributed model for IRs & workflow
- Allocation of responsibilities between institutions
- Requirements for collaboration

Funded by:

JISC



SHERPA DP Project

- **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
- **Programme:** JISC Digital Preservation and Records Management Programme
- **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.

Funded by:

JISC



Why distribute essential services?

Institutional reasons

- Staff are too busy - do not have time to do everything.
- Equipment costs – hardware, software costs may be too high
- No existing infrastructure to support functions
- Pre-existing agreement with organisation

Wider issues

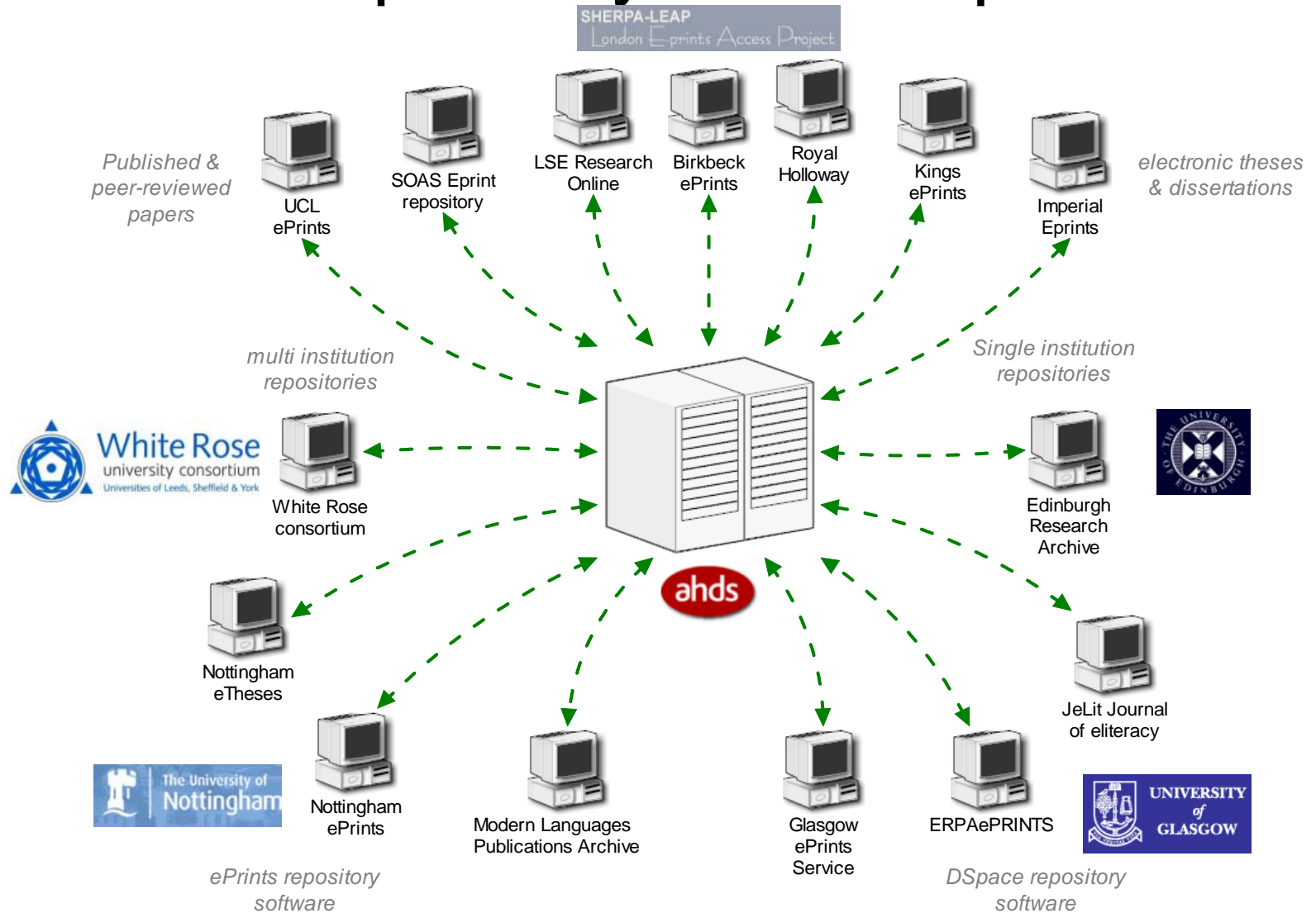
- Compliance with specification/certification agency (e.g. TDR)
- Seeking to remove repetition of services

Funded by:

JISC



Repository Landscape



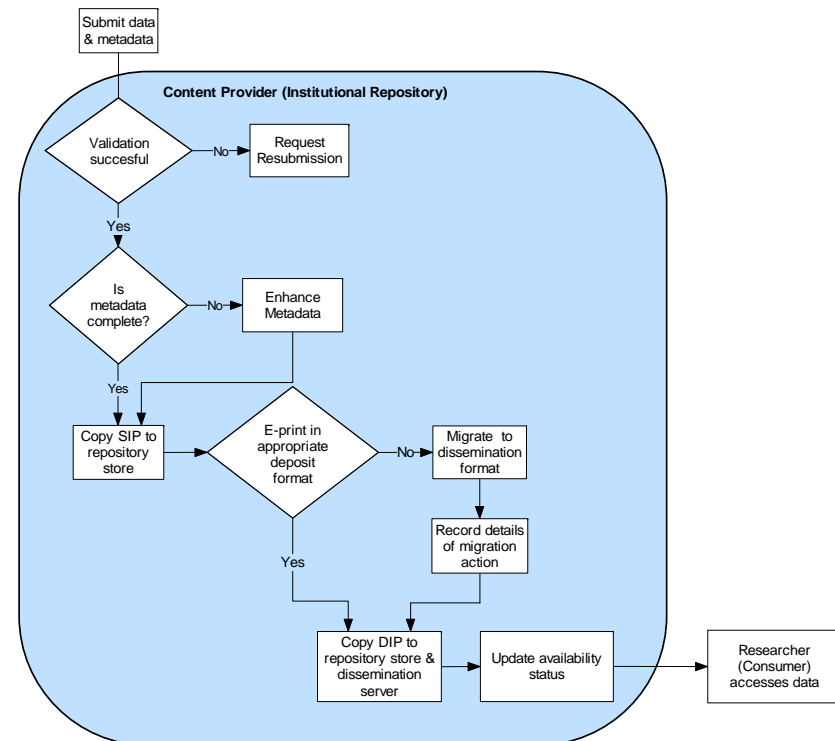
Funded by:



Simplified IR Workflow

IR staff typically perform four tasks:

1. Document validation
2. Metadata validation & enhancement
3. Migration of data formats
4. Release



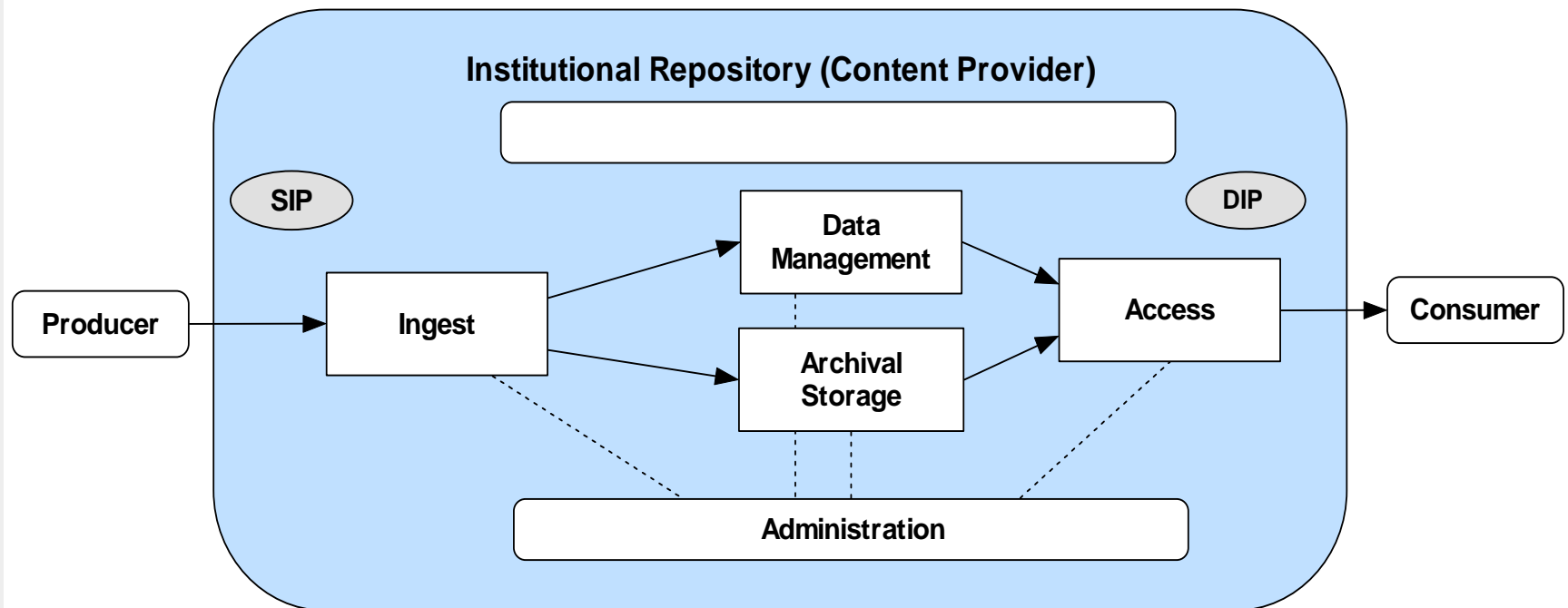
Funded by:

JISC

Arts & Humanities
Research Council

© AHDS

OAIS compliance in the IR



SIP = E-print & discovery MD

DIP = migrated e-print and discovery MD

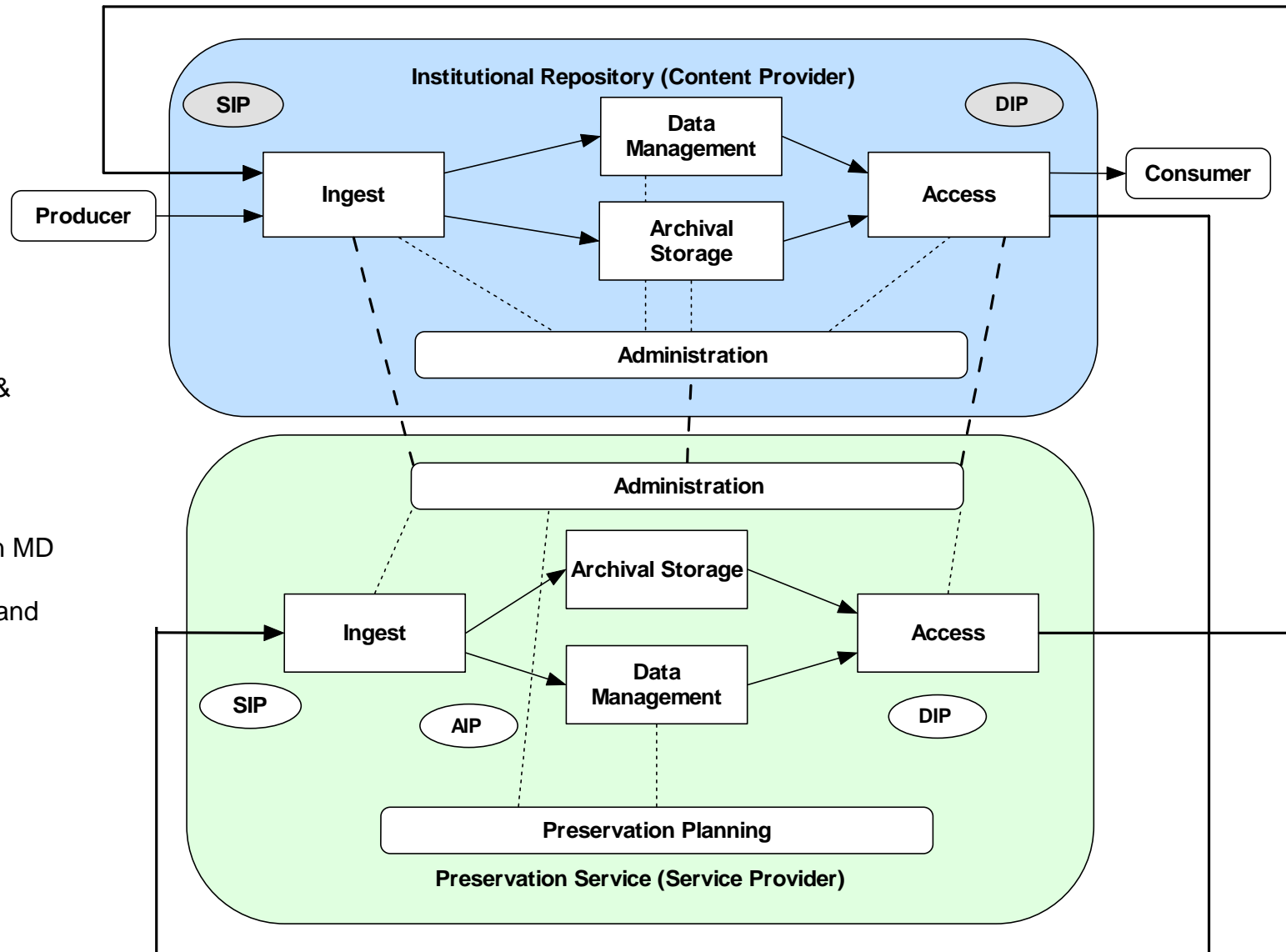
Funded by:

JISC

Arts & Humanities Research Council

© AHDS

Distributed OAIS Model



SIP = E-print & discovery MD

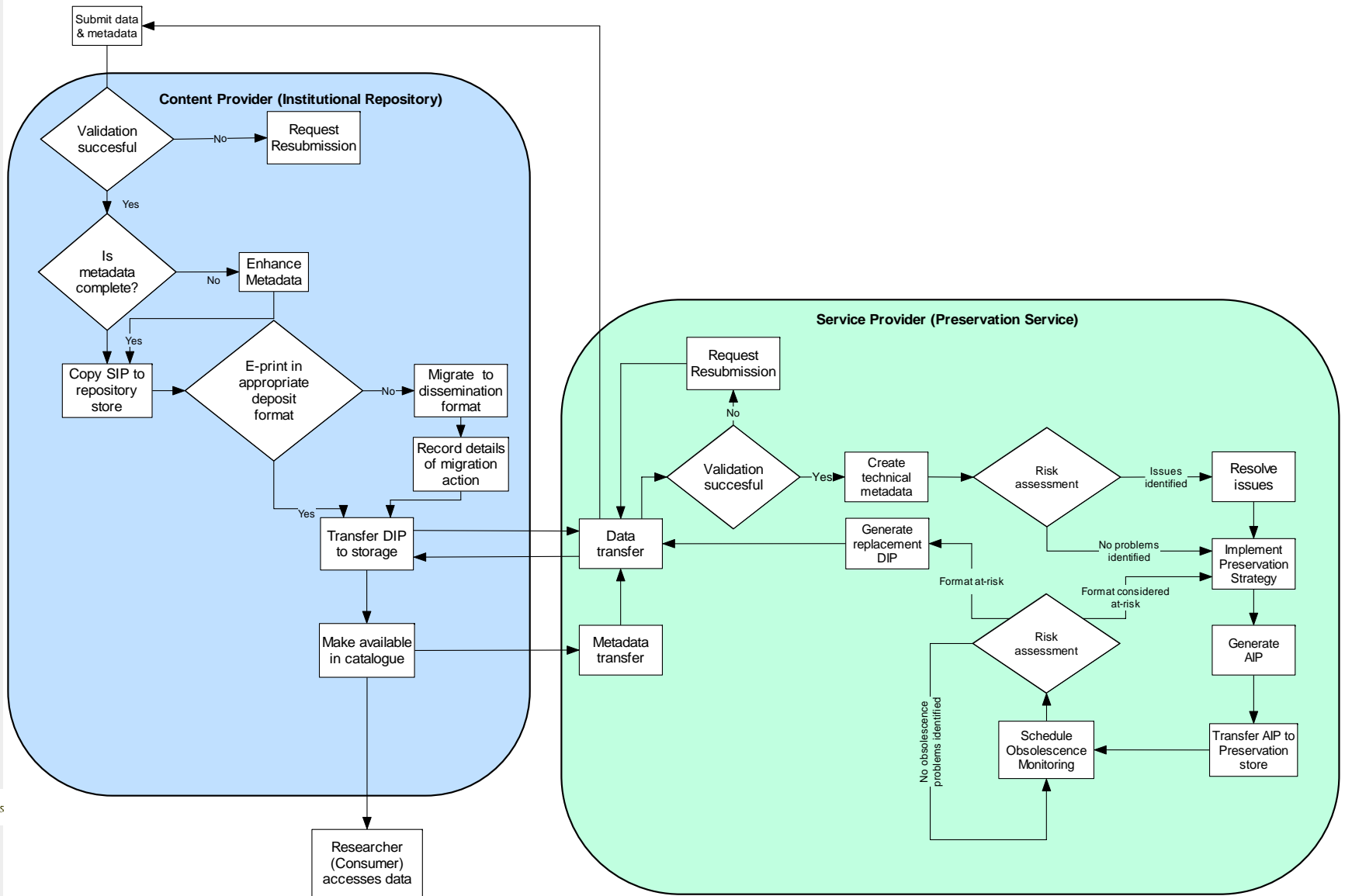
AIP = E-print, discovery & preservation MD

DIP = E-print and discovery MD

Funded by:



Distributed Workflow



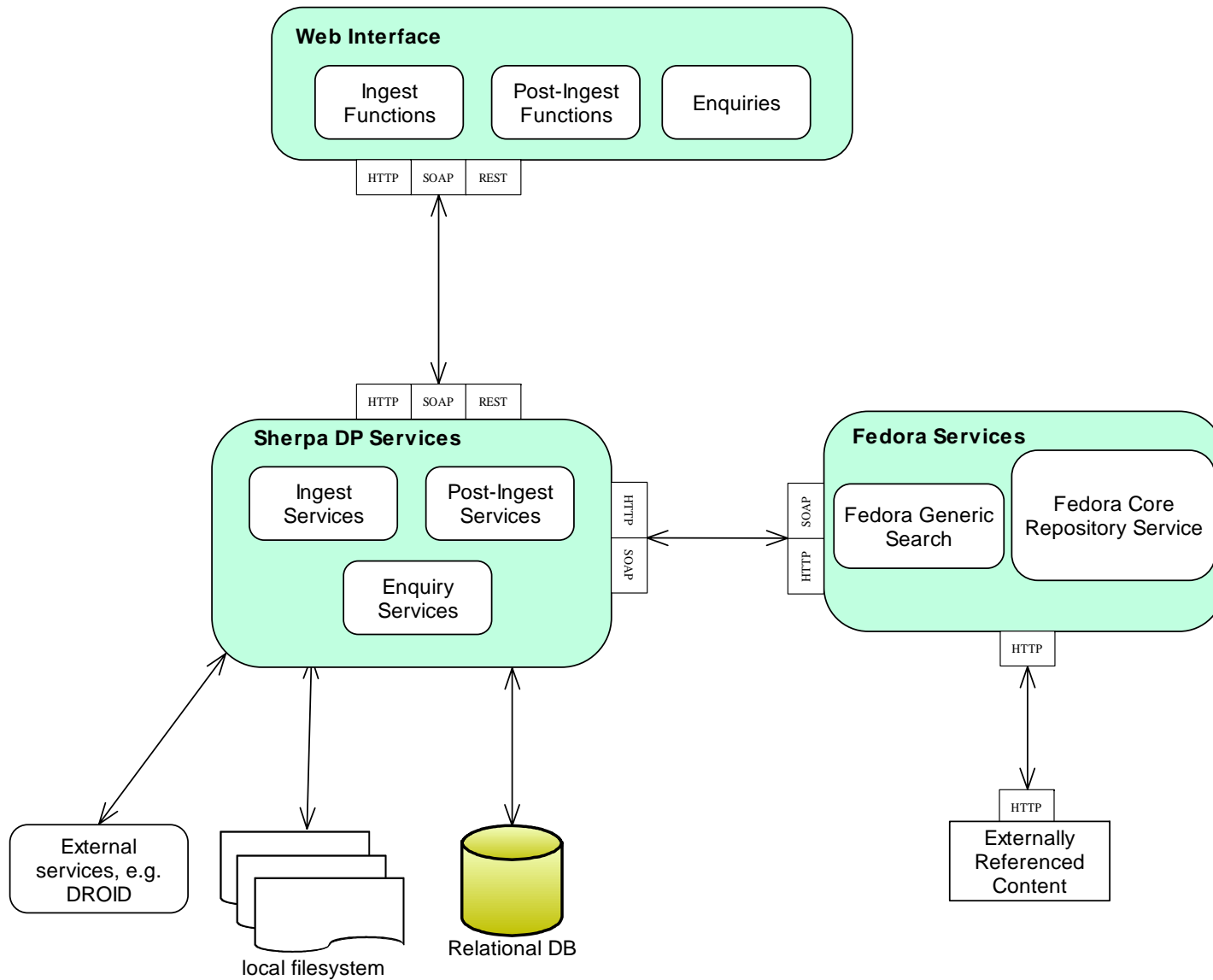
Funded by:



Arts & Humanities Research Council

© AHDS

System Architecture



Funded by:



Arts & Humanities
Research Council

© AHDS

Other examples of distributed services

- **AHDS** – distributed centres
- **Florida Center for Library Automation & Florida colleges/universities** – distributed ingest and preservation functions
- **SDSC, Uni. of Maryland, & NARA** – distributed archive connected through SRB
- **Cornell University Library and Göttingen State and University Library** – repository for digital journals managed through a shared admin interface

Funded by:

JISC



Minimum Requirements for Preservation

Technical

1. Expose basic metadata to identify new submissions.
2. Provide some method of identifying data objects associated with a metadata record
3. Provide some method of authenticating data objects associated with a metadata record

Policy

4. Policies to identify preferred file formats for deposit and inform the Producer (depositor) and preservation service provider of these requirements;
5. Create and implement a deposit licence that:
 - Establishes permission to transform the submitted resource (e-print) for the purpose of preservation and accessibility.
 - Establish permission for the Content Provider to allocate responsibility for preservation to a third-party.

Funded by:

JISC



Best Practice Requirements for preservation

Technical

6. Expose a full record of all metadata stored by the IR, including desc, admin, preservation.
7. Provide a detailed description of the metadata schema implemented, including a list of elements and vocabulary.
8. Co-operate with the partner institution to identify methods that may be used to return metadata and data to the institutional repository

Policy

9. Co-operate with the Preservation Service Provider to review and potentially revise ingest policies to ensure SIPs are deposited in formats appropriate for preservation.

Funded by:

JISC



Service Provider 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:

- 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



Conclusion

- Must have a reason to distribute services.
- The location of activities is unimportant. However, appropriate repository services must exist.
 - Repository interoperability is possible using appropriate standards.
 - There is no out-of-the-box solution to preservation.
- There must be a clear understanding of the roles & responsibilities for each institution.
- Further investigation on OAIS-compliant models to represent distributed services is necessary

Funded by:

JISC



Thank You!

URL:

<http://www.sherpadp.org.uk/>

Contact

gareth.knight@ahds.ac.uk

Funded by:

JISC

