

To Preserve and Make Available...



- ‘An archival information system is an organisation of people and systems that has accepted the responsibility to **preserve information** and make it **available** for a **Designated community**’
- ‘Open Archival Information System’



OAIS



- Moving on from simple archival storage
- Issue of standards, compliance and trust
- Responsibilities
- Establishes Common terms and standardisation
- Shared vocabulary
- Reference Model

Overview



- Digital Preservation Coalition & activities
- UK Needs Assessment
- Overview of OAIS
- Practical implementation
- Step-by-step walk through of processes
- Detailed Metadata

Digital Preservation Coalition



1. Promoting Digital Preservation

- Efforts aimed at ensuring digital preservation is on the agenda of key stakeholders. This involves an advocacy and PR campaign, press articles, conference papers, the UK Needs Assessment, etc.

2. Acting to increase funding

- Ensuring the nation's investment in preservation

3. Fostering collaboration and forging strategic alliances

- With relevant agencies nationally and internationally
e.g. National Library of Australia, Library of Congress



Six Work Packages



4. Producing, providing, and disseminating information

- Building expertise in digital preservation e.g. commissioning Technology Watch Reports and holding DPC forums. Developing the DPC website and producing quarterly reports of 'What's New in Digital Preservation', monthly reports to members via Discussion lists

5. Promoting and developing services, technology, standards and training

- E.g. Initiating training workshops, including the Digital Preservation Training Programme and producing Technology Watch Reports, developing intensive digital preservation training programme

6. Continuing to develop the Coalition's activities

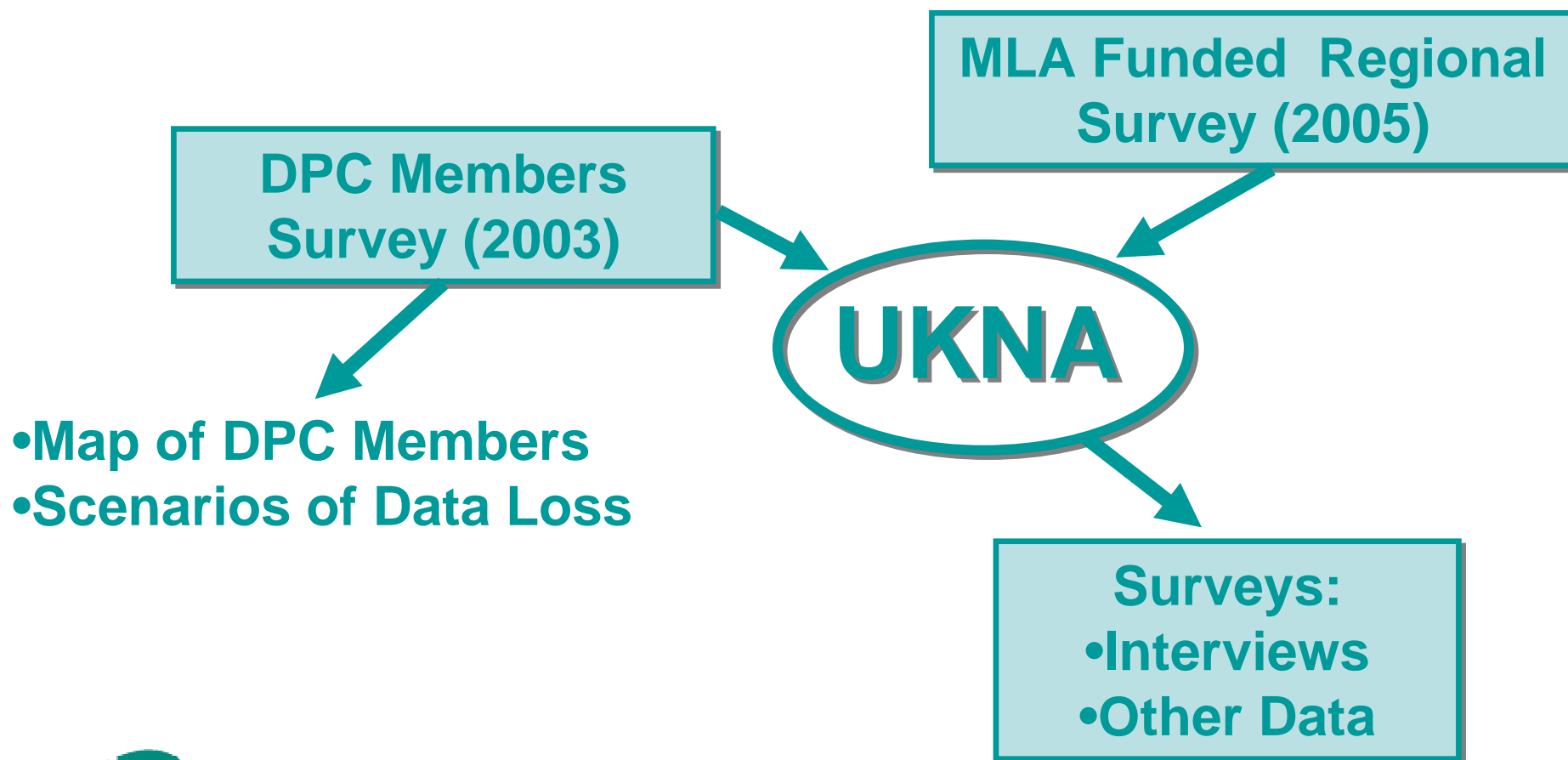
- Recruiting membership and attracting resources to support DPC's goals



Activities – Initiatives



UK Needs Assessment Exercise:



UK Needs Assessment



- Building up a picture of what is (and isn't) happening in the UK
- Gathering data on volumes and formats
- Identifying priorities
- Quantifying the problem
- Building sufficient knowledge base to be able to plan effectively



‘OAIS at Edinburgh University’

Najla Semple

Content Area #1



Edinburgh University

- Administrative services
 - Centrally managed services
 - Official materials
 - Legal considerations
 - Freedom of Information & Data Protection Acts
 - No culture of creating metadata



Content Area #2



- Research materials
 - Key research data
 - Little control over their production
 - Teaching materials, databases, digitised material
 - Survey of departments
 - Great variety of data formats



Pilot



- Test-bed – University Calendar
- Archive the current version
- Exists as HTML/PDF
- Authenticity concerns
- Make a case for Digital Preservation





?

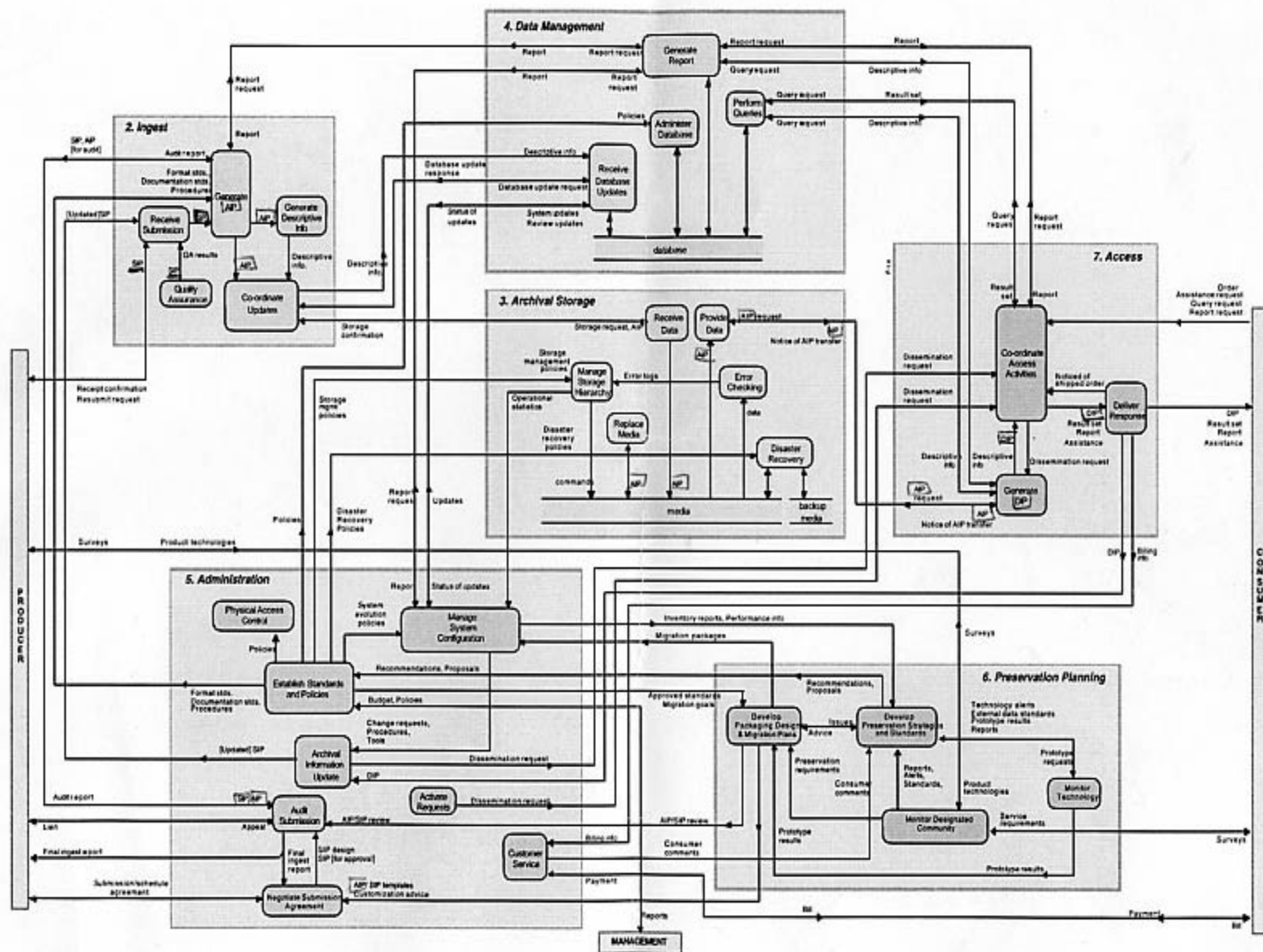
?

?

?

?







?

?

?

?

?



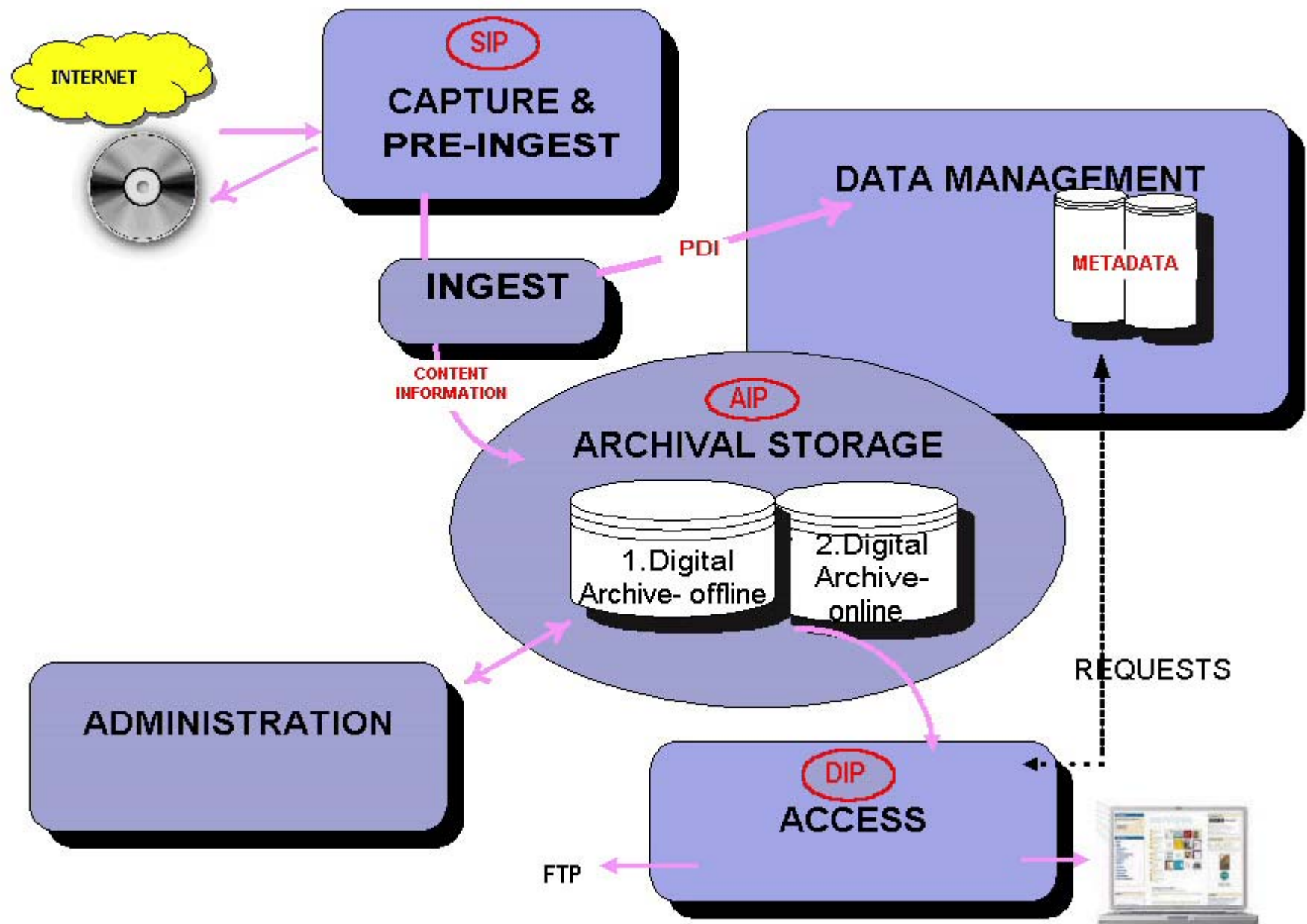
Piecing together the puzzle



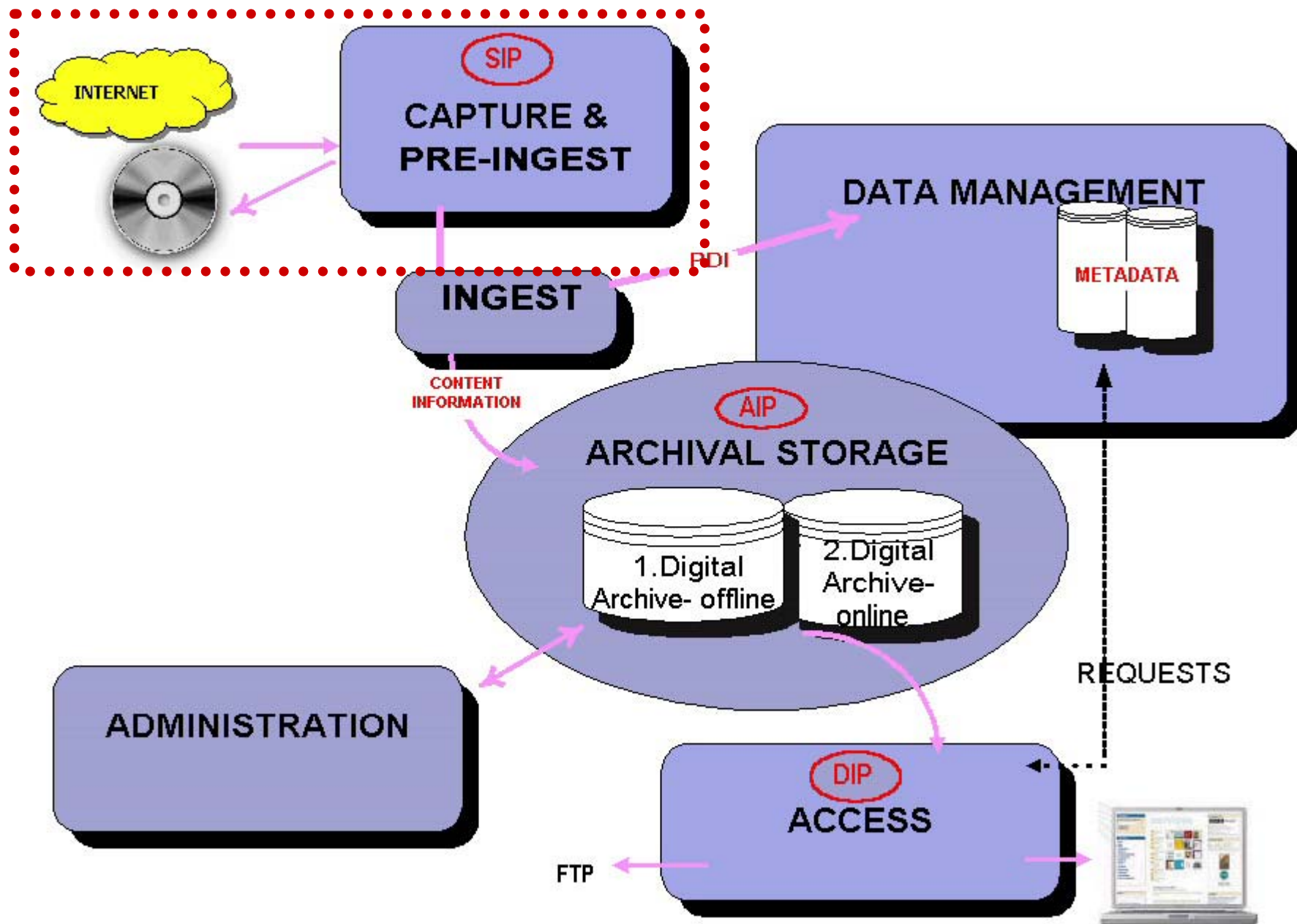
1. Pre-ingest
2. Ingest
3. Archival Storage
4. Data Management
5. Access
6. Administration



EDINBURGH LIBRARY DIGITAL ARCHIVE WORKFLOW



EDINBURGH LIBRARY DIGITAL ARCHIVE WORKFLOW



Pre-Ingest



1. **Delivered to PC via CD-Rom; secure FTP; or harvested from web**
2. **Write copy to CD-Rom**
3. **Deal with IPR issues, QA**
4. **Level of functionality- Transfer data to common file format**
5. **Allocate metadata**



6. Process: Cataloguing, Acquisitions

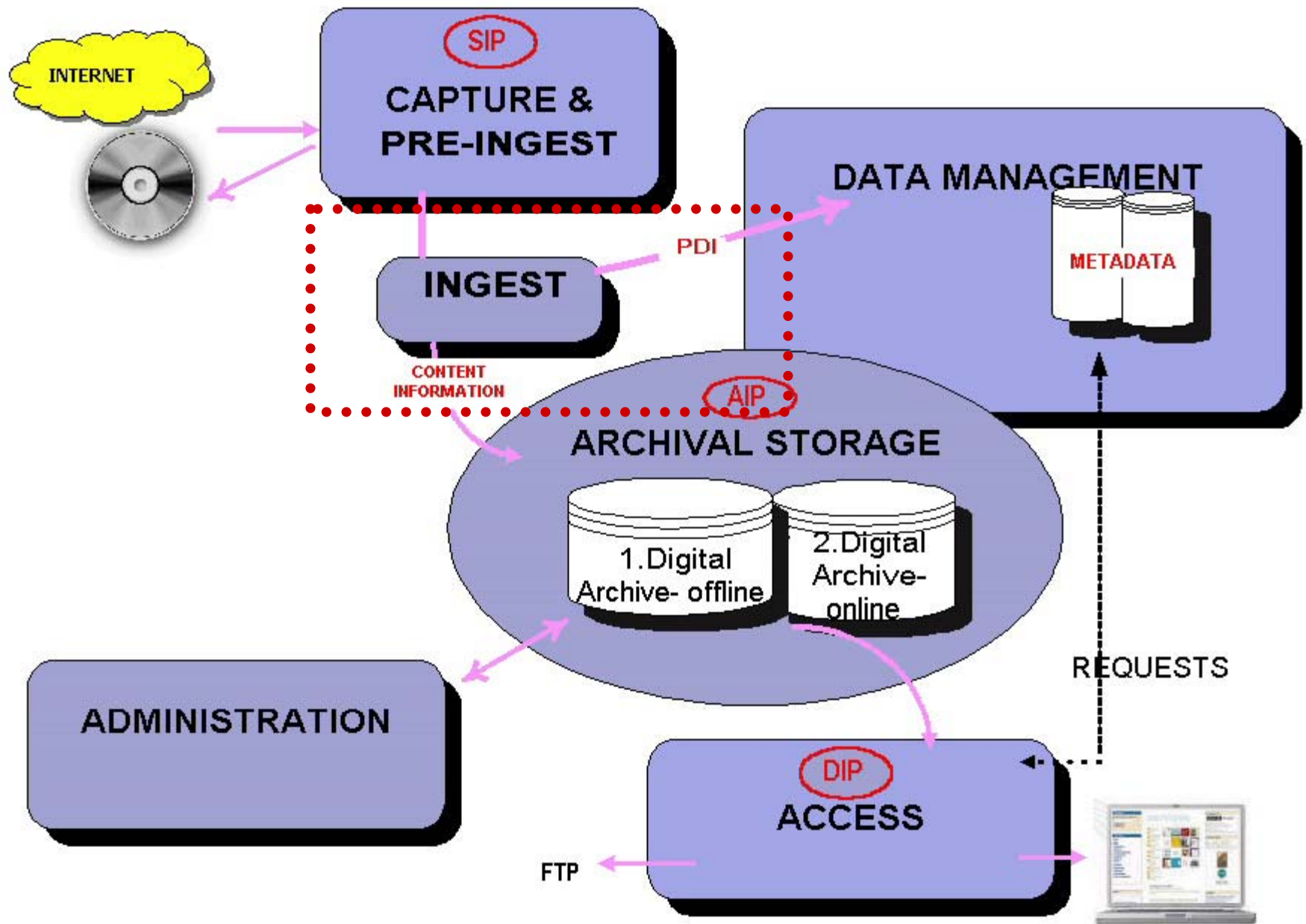


SIP

- 1. Digital Object**
- 2. Bibliographic Metadata**
- 3. Details of file format**
- 4. Details of technical information**
- 5. Retention/accessibility details**



EDINBURGH LIBRARY DIGITAL ARCHIVE WORKFLOW



Ingest



1. Assign Unique ID

ed:eucsda:lib:webdoc:03

Name of institution, Name of
archive, What unit , Doc.Type

2. Allocate a checksum- MD5

3. Use of tar to create byte stream

4. AIP creation on Ingest



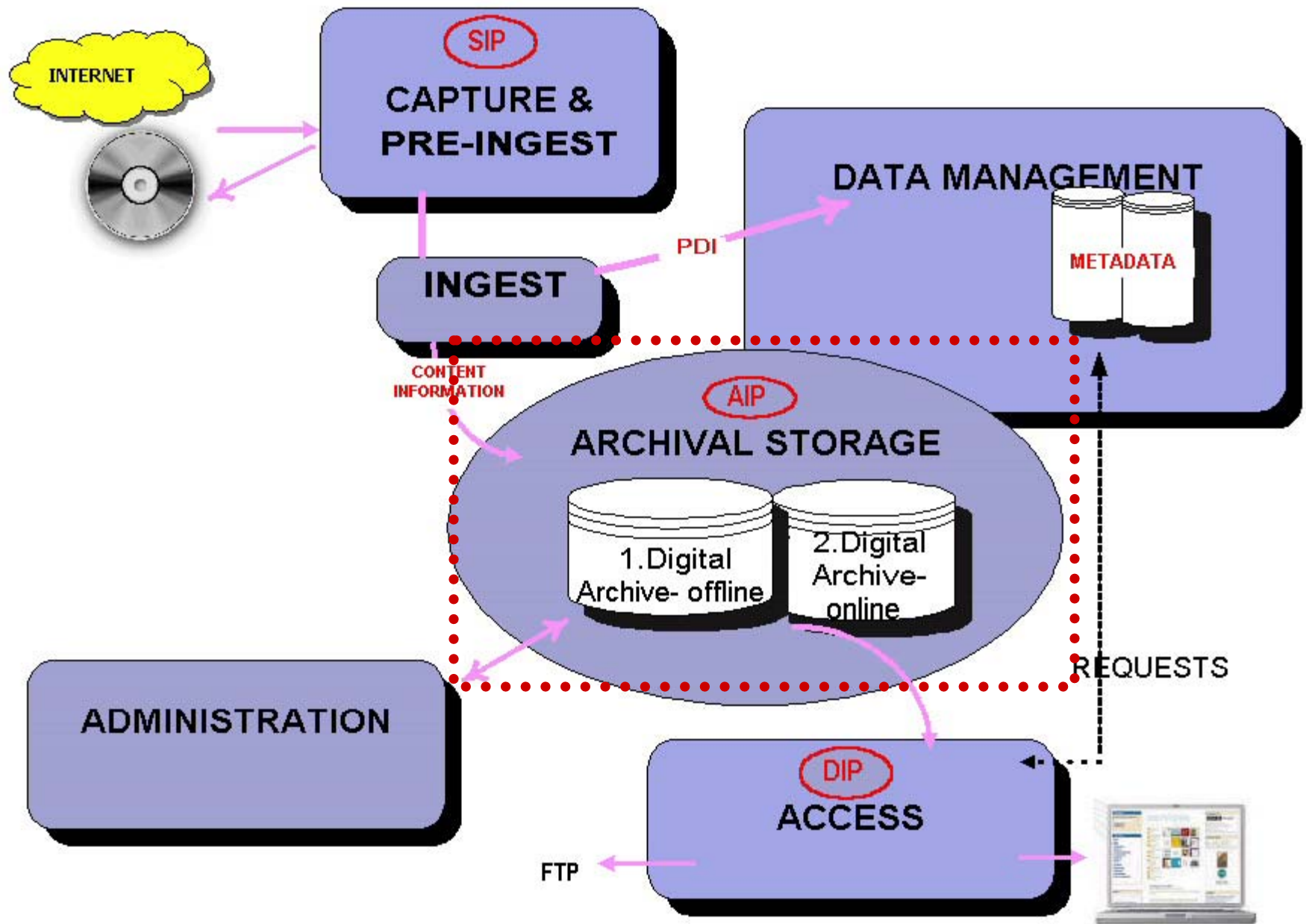


AIP

1. XML file - Content Information: Digital Object & RI
2. XML file- PDI [Copy to Data Management]
3. Text file- Specification of Packaging tar / Zip
4. Text file- Unique ID



EDINBURGH LIBRARY DIGITAL ARCHIVE WORKFLOW



Archival Storage



1. Two locations for storage

a) Service Provider- OCLC/DOMS

Day to day usage, server storage

b) EUCS - Optical media

Write Once Read Only, Cannot delete

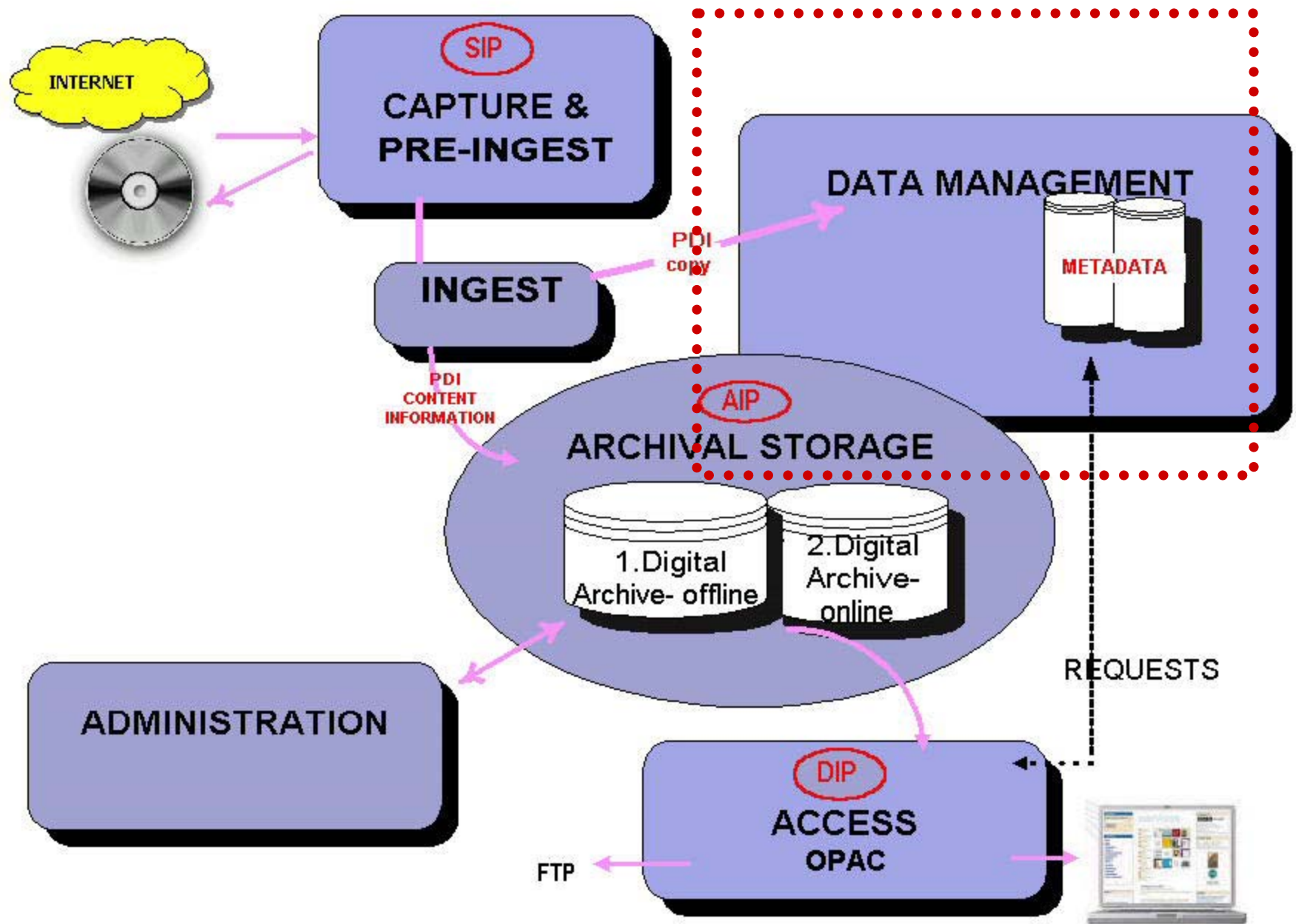
'Dark' vault

Access via FTP

2. Library process: Data Storage



EDINBURGH LIBRARY DIGITAL ARCHIVE WORKFLOW



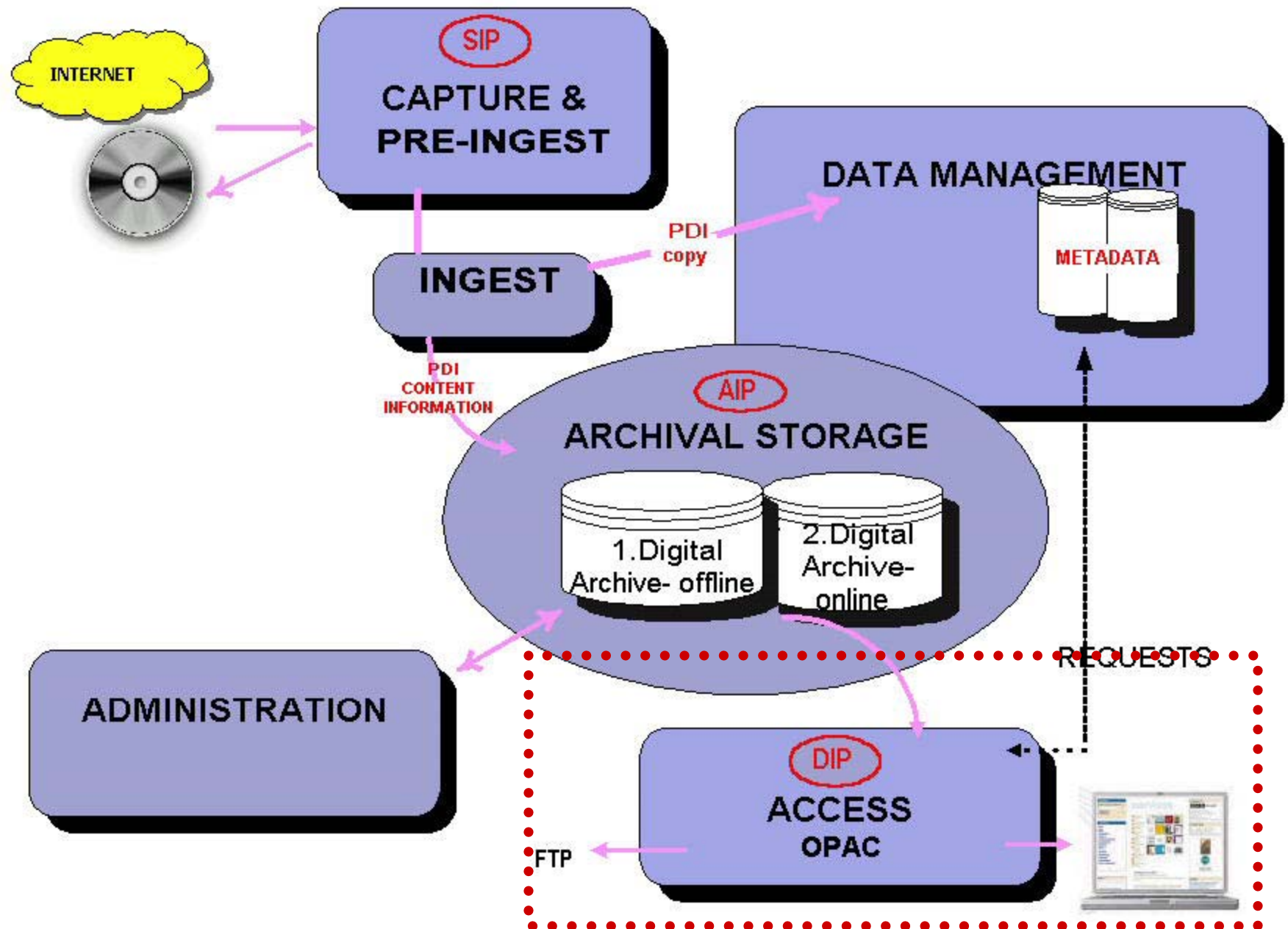
Data Management



1. **AIP-Preservation Description Information**
2. **Managed by library systems team**
3. **Databases**
4. **Retrieval metadata**
5. **Linked to Administration- Requests from users are logged**
6. **Technical conditions to view the object**



EDINBURGH LIBRARY DIGITAL ARCHIVE WORKFLOW



Access



1. **Users receive the DIP**
2. **Records catalogued onto library OPAC**
3. **Dublin Core mapped to MARC records**
4. **'Deep' archive accessible via ftp**
5. **Day to day - OCLC / Service provider**

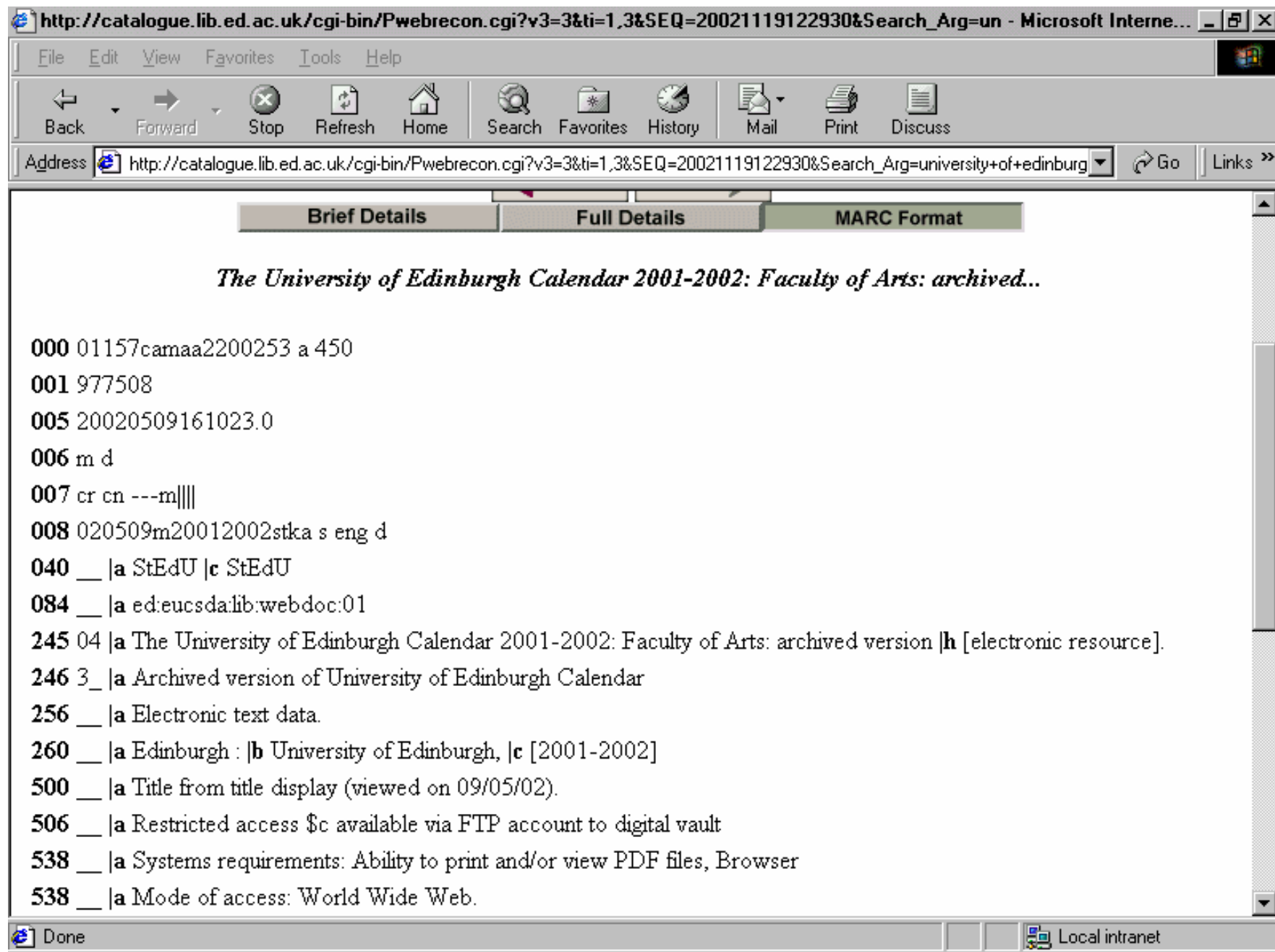




DIP

1. Digital Object
2. Software to view it
3. PDI- some fields





Administration



- 1. Review of IPR**
- 2. Access controls**
- 3. Long-term preservation management**
- 4. Based with digital archivist**
- 5. Process: Library systems**
- 6. Usage statistics**



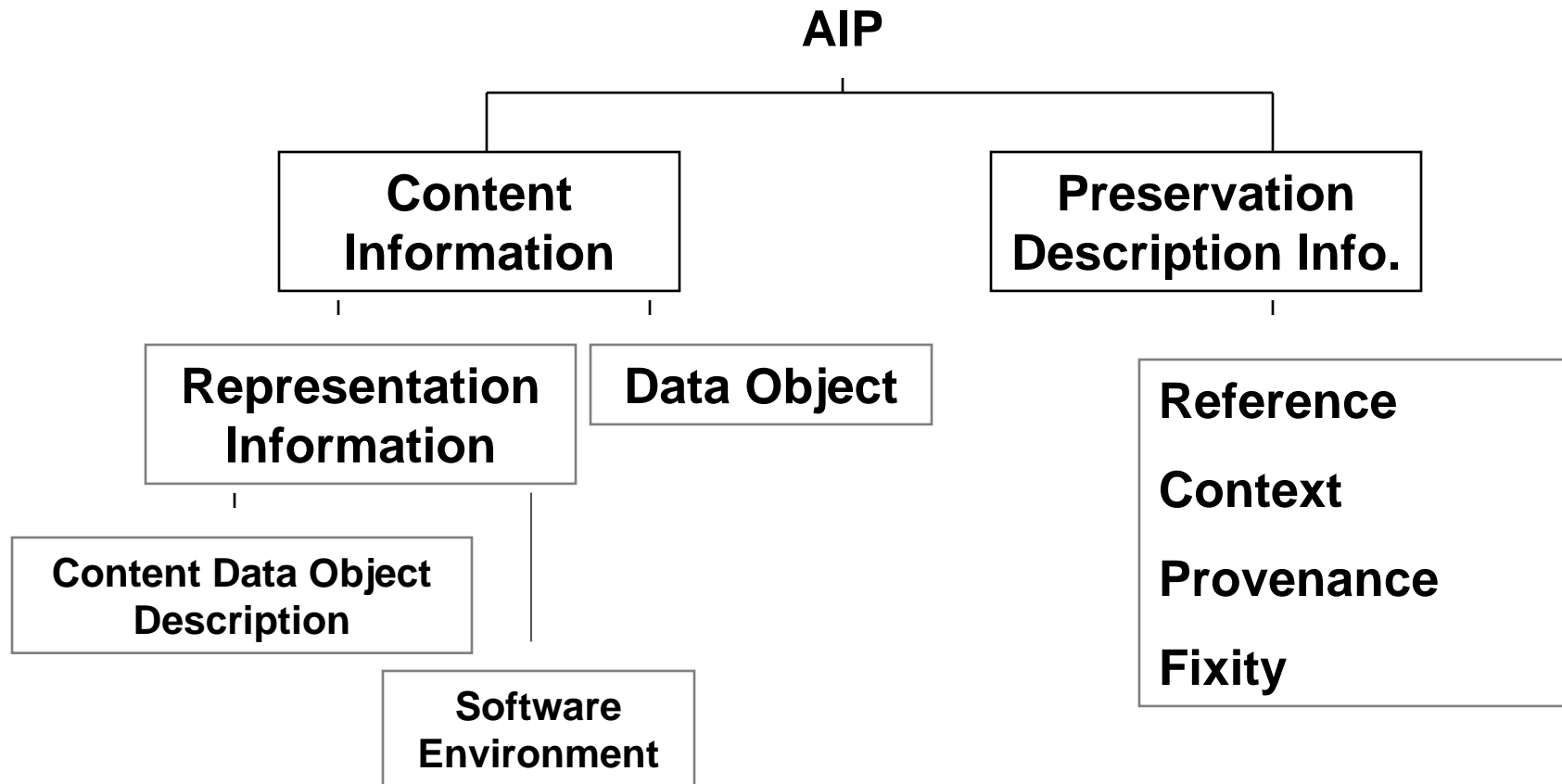
Metadata



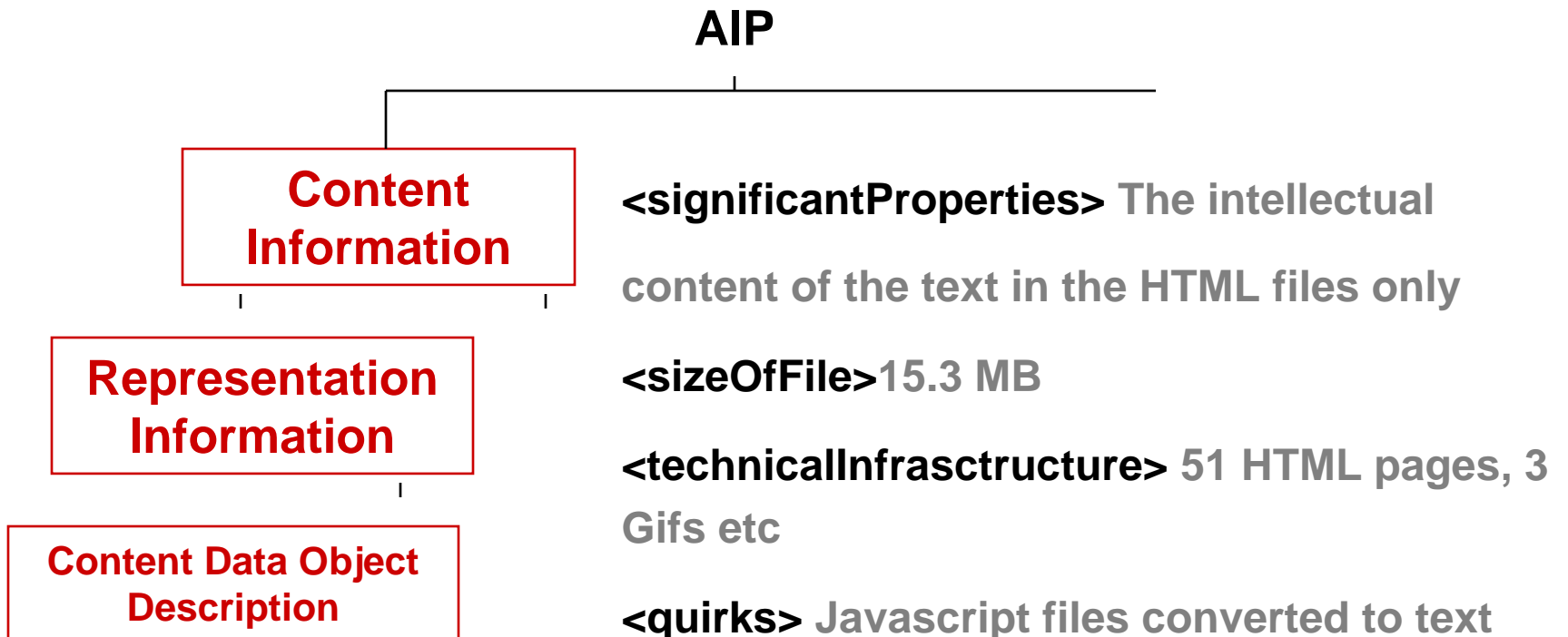
- Created manually or harvested
- Schemas: Cedars, National Library of New Zealand, OCLC/RLG WG
- XML- Cedars DTD
- Level of granularity



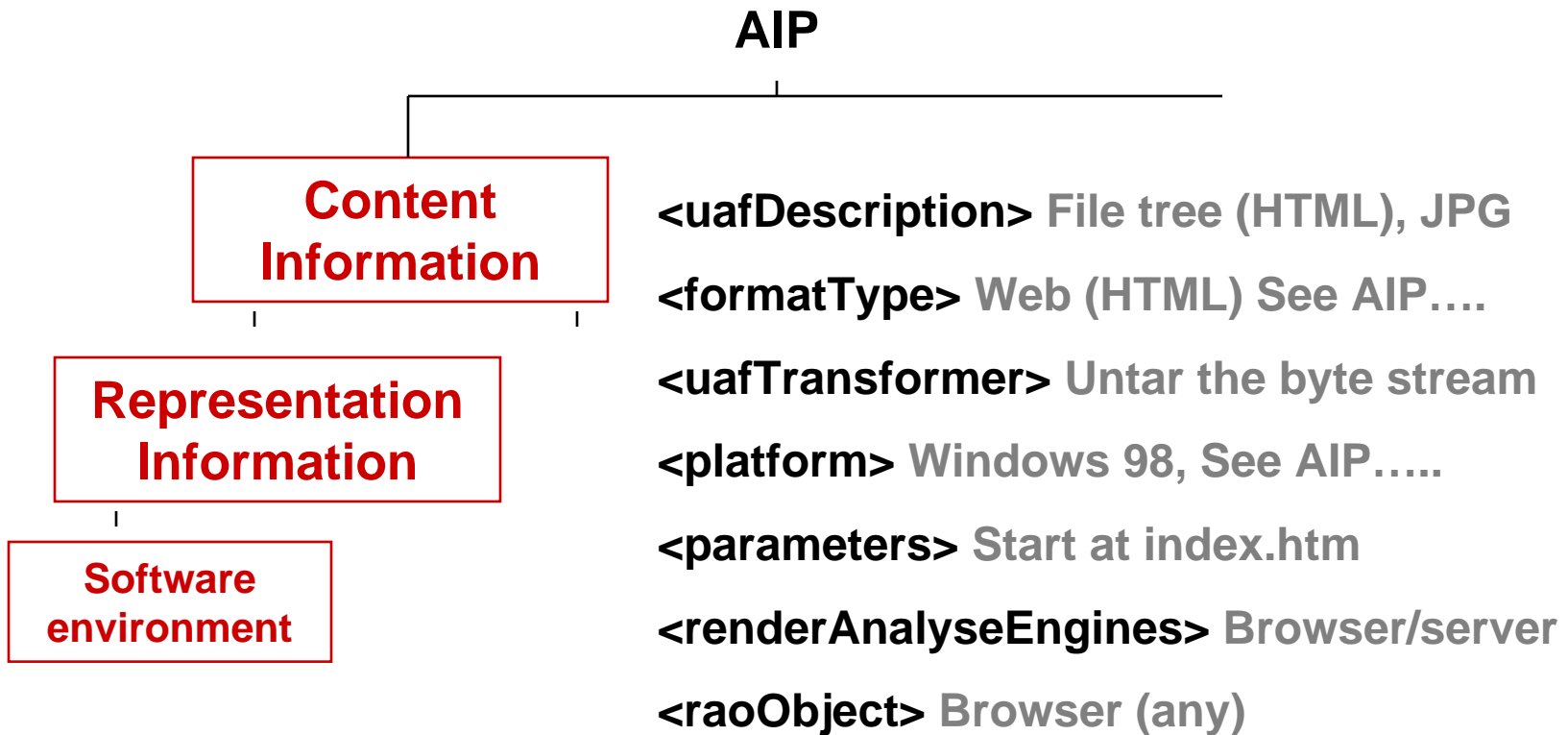
Metadata



Representation Information



Representation Information





PDI: Context



AIP

<contextInformation>

<relatedInformationObject>

<relationship> University Calendar- current
analogue version

<reference> Available at:

DS 198.1

**Preservation
Description Info.**

Reference

Context

Provenance

Fixity



PDI: Provenance: History & Environment

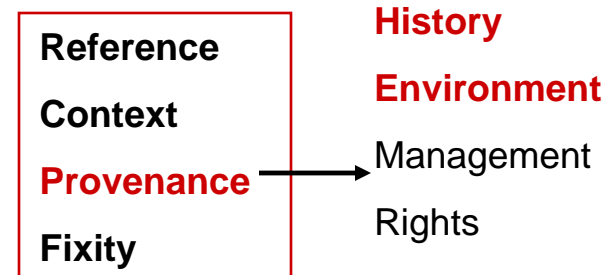


<historyOfOrigin>

<reasonForCreation> University of Edinburgh corporate record

<custodyHistory> Managed and maintained by University Calendar

<changeHistoryBeforeArchiving> Originally created on Quark Express. Available in HTML and PDF



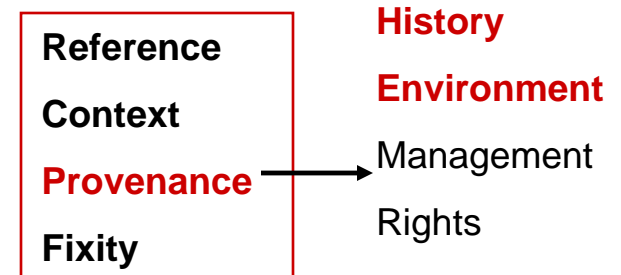
PDI: Provenance: History & Environment



<originalTechnicalEnvironments> Website run on SOLARIS 8 operating environment. APACHE 1.3.12 webserver Server side includes. Robot text file.

<documentation> URL for Apache 1.3: <http://apache.org/docs/>. See AIP.....

<sizeOfFile> 15.3 MB before compression



PDI: Provenance: Management



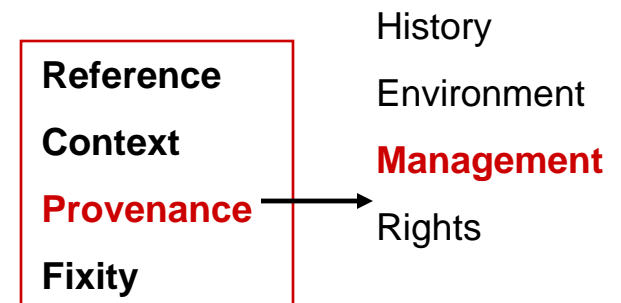
<managementHistory>

<ingestProcessHistory> Files received via FTP from CPA, Dreamweaver files transferred to plain text files. Used tar for Windows to create byte-stream

<administrationHistory> (Occurrences after ingest:) (Metadata modifications....)

<actionHistory> Files received via FTP from CPA, files stored on archivist's hard drive,

<retentionPeriod> Permanent



PDI: Provenance: Rights



<rightsManagement>

<negotiationHistory> None needed

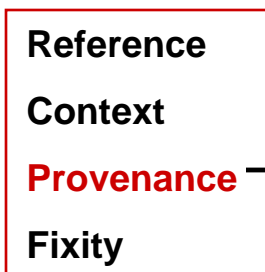
<copyrightStatement> Copyright University Edinburgh

<nameOfPublisher> University of Edinburgh

<dateOfPublication> 2000

<placeOfPublication> University of Edinburgh Library

<permittedByStatute> For open access



History
Environment
Management
Rights

PDI: Context



AIP

<Fixity>

<checksumString> 2hht57583920uii.....

<dateUndertaken> 12.08.01

**Preservation
Description Info.**

**Reference
Context
Provenance
Fixity**

