DPC
OAIS Briefing Day
4 April, York
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’
• 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
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
UK Needs Assessment Exercise:

- Map of DPC Members
- Scenarios of Data Loss

Surveys:
- Interviews
- Other Data

MLA Funded Regional Survey (2005)
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
Edinburgh University

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

SIP

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

1. Assign Unique ID
   ed:eucscda: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
Archival Information Package

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
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
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
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
Dissemination Information Package

**DIP**

1. Digital Object
2. Software to view it
3. PDI - some fields
The University of Edinburgh Calendar 2001-2002: Faculty of Arts: archived...

000 01157camaa2200253 a 450
001 977508
005 20020509161023.0
006 m d
007 cr cn ---m|||
008 020509m20012002stka s eng d
040 __ |a StEdU |c StEdU
084 __ |a ed.eucsda.lib:webdoc:01
245 04 |a The University of Edinburgh Calendar 2001-2002: Faculty of Arts: archived version |h [electronic resource]
246 3_ |a Archived version of University of Edinburgh Calendar
256 __ |a Electronic text data.
260 __ |a Edinburgh : |b University of Edinburgh, |c [2001-2002]
500 __ |a Title from title display (viewed on 09/05/02).
506 __ |a Restricted access $c available via FTP account to digital vault
538 __ |a Systems requirements: Ability to print and/or view PDF files, Browser
538 __ |a Mode of access: World Wide Web.
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
<significantProperties> The intellectual content of the text in the HTML files only
<sizeOfFile>15.3 MB</sizeOfFile>
<technicalInfrastructure> 51 HTML pages, 3 Gifs etc
<quirks> Javascript files converted to text
<uafDescription> File tree (HTML), JPG
<formatType> Web (HTML) See AIP....
<uafTransformer> Untar the byte stream
<platform> Windows 98, See AIP.....
<parameters> Start at index.htm
<renderAnalyseEngines> Browser/server
<raoObject> Browser (any)
PDI: Reference

AIP

<referenceInformation>
<resourceDescription>
<Dctitle>
<Dcssubject>
<DCdescription>
<DCcontributor>
<DCrights>
<DCdate>

Preservation Description Info.

Reference Context Provenance Fixity
AIP

<contextInformation>
<relatedinformationobject>

<relationship> University Calendar- current analogue version

<reference> Available at:
DS 198.1

Preservation Description Info.

Reference Context Provenance Fixity
<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
<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
<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
<rightsManagement>

<negotiationHistory> None needed

<copyrightStatement> Copyright University Edinburgh

<nameOfPublisher> University of Edinburgh

<dateOfPublication> 2000

<placeOfPublication> University of Edinburgh Library

<permittedByStatute> For open access
PDI: Context

AIP

<Fixity>
<checksumString> 2hht57583920uii.....
<dateUndertaken> 12.08.01

Preservation Description Info.

Reference Context Provenance Fixity