

**Susan Thomas,
Project Manager**

PARADIGM

An overview of the project

Wellcome Library, 10 October 2006



PARADIGM

Outline

- What is Paradigm?
- Lessons so far
- Some future challenges
- Next steps

What is Paradigm?

- Funded for 2 years by the JISC, ends Feb. 2007
- Collaboration between Oxford University Library Services (lead) & John Rylands University Library, Manchester
- 1.5 fte archival, 1 fte developer plus input from Oxford Digital Library and Special Collections departments
- Explores digital preservation from 'personal' and 'collecting' perspectives in the context of a 'hybrid archive'
- Hands-on experience of soft issues by working with politicians and their materials (selection and acquisition, creator attitudes, legal issues, etc.)
- Hands-on experience of technical issues and tools (Fedora, OAIS, METS & PREMIS)

Aims

- Test an alternative approach to traditional archives collection development for hybrid archives – early intervention
- Harmonise long-standing archival standards and workflows with digital repository standards and workflows
- Develop prototype preservation repository
 - Key standards: OAIS, Fedora, METS and PREMIS
 - Focus on acquisition, ingest and preservation rather than access
- Develop expertise in management of hybrid archives at partner institutions and provide a platform for future activity
- Develop and share strategies for personal digital archives
 - That are based on experiences with politicians and their digital archives
 - Through Paradigm's Online Workbook & 'roadshow'

Archival Principles & OAIS

- Actually reasonably similar, but different terminology
 - Relationships with Donors and Readers > 'Producers' and 'Consumers'
- Concepts of Authenticity, Original Order, Provenance and Relationships and Intellectual control important to archivists
 - Expressed differently in OAIS, but included in its Information and Functional Models
- What's different then?
 - Workflows and procedures – how principles are implemented
 - Implementing hybrid solution
 - The preservation metadata – greater variety and detail required
 - The dissemination metadata – collection still needs overarching finding aid, but items in a presentation repository must be self-describing
- Combining Archival Principles and OAIS...

OAIS Functional Model

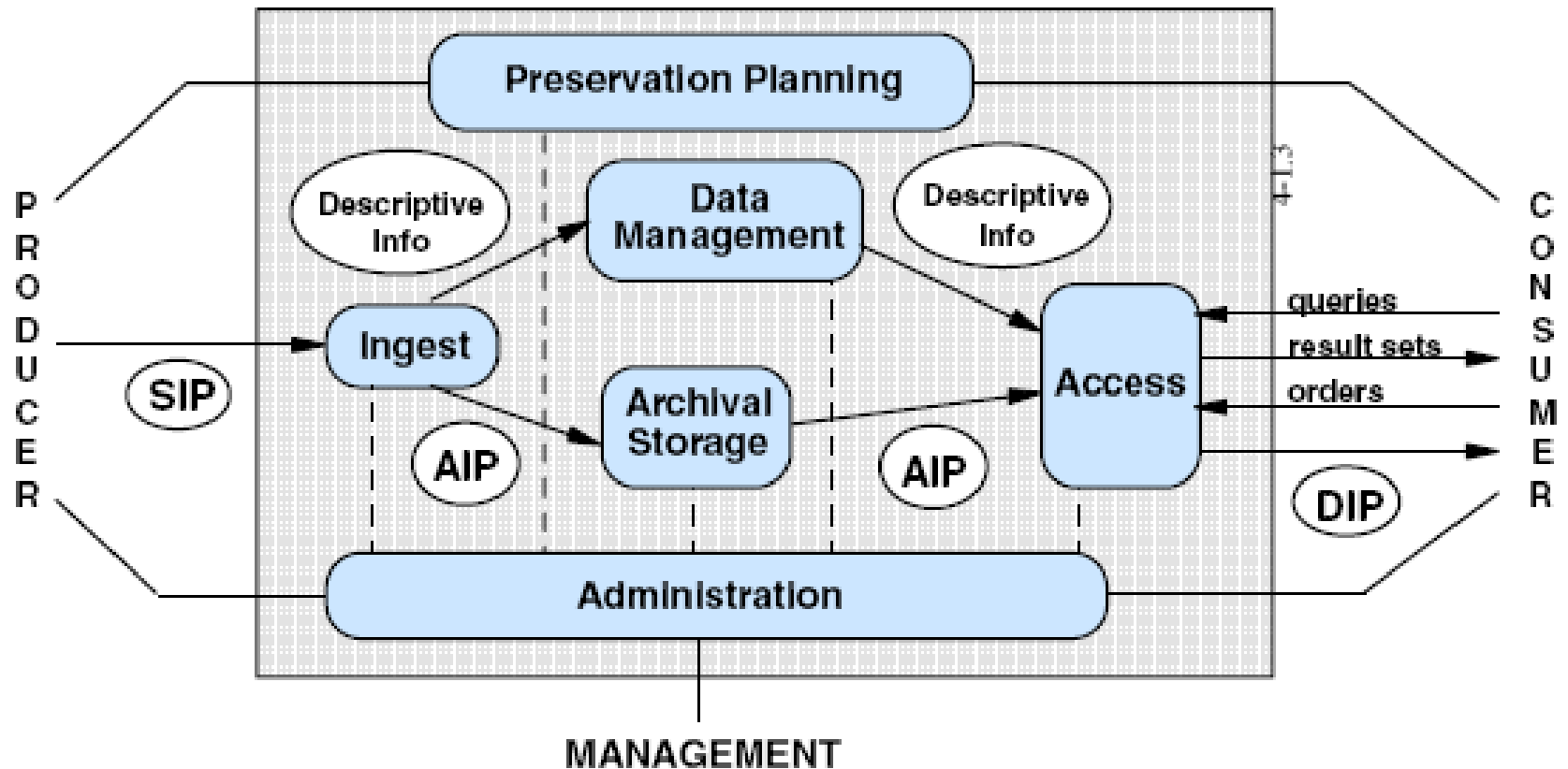
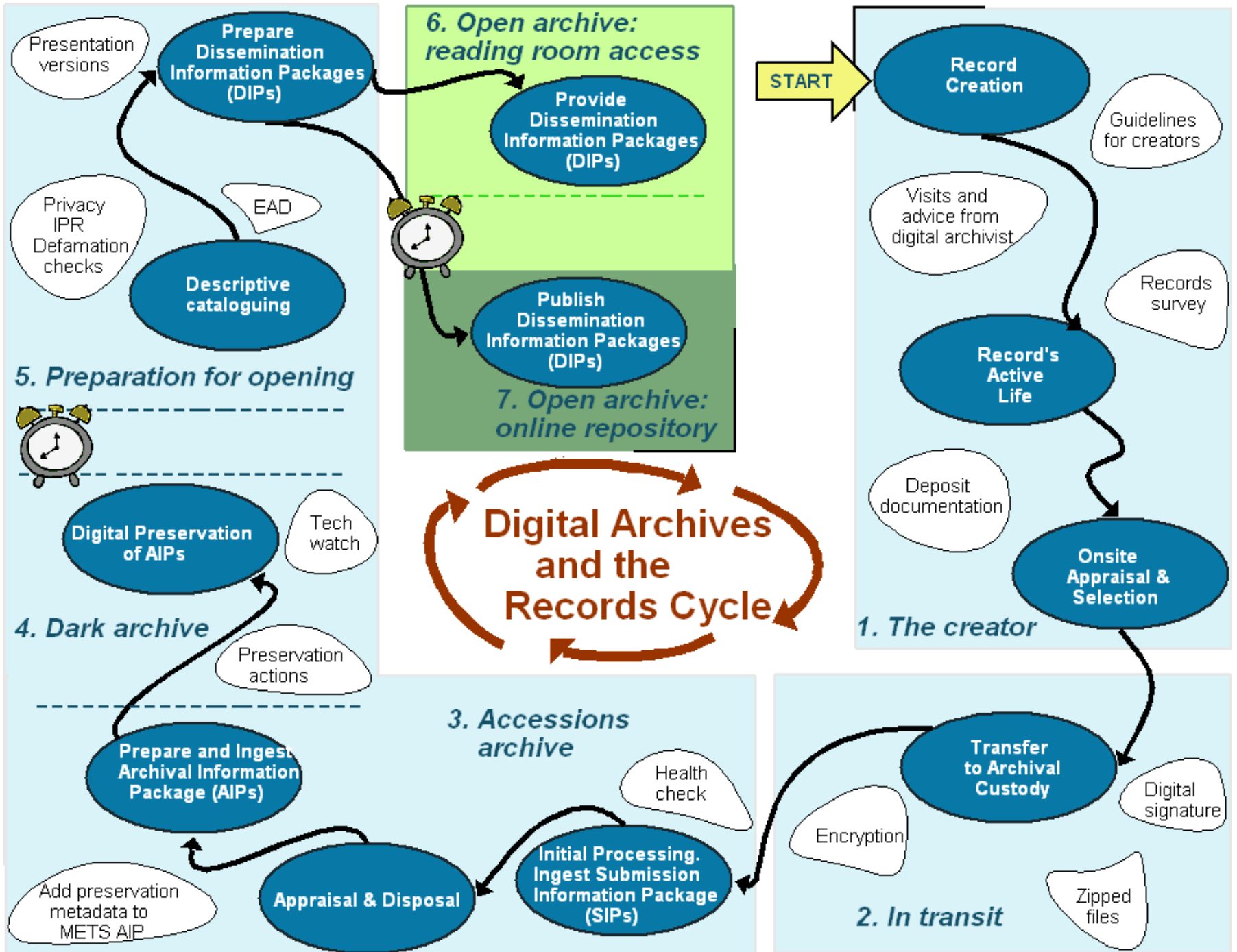


Figure 4-1: OAIS Functional Entities



Early-intervention: why?

- Archives traditionally reach a repository once an individual has retired or passed away – potentially a long time after creation
 - Physical survival of paper and parchment straightforward, but bit-level survival uncertain for digital objects of this age
 - If objects survive at bit level, digital archaeology may be required to liberate them
- Individuals have limited IT support
- Usage of third party storage solutions growing, so likelihood of capturing entire archive without active engagement reduces
- Reduce risk of loss and uncertainty of digital archaeology by bringing digital archives into a managed environment and/or providing advice while records still active

Early-intervention: lessons

- Lessons from working with politicians' current records
 - Digital increasingly used as 'master', but poorly managed
 - Poor understanding of archiving for historical purposes
 - Privacy and security concerns – own and third party – increased by recent date of material. Reluctance to deposit material now, or at all
 - Repository must manage material with legal protections for longer
 - Finding time for history in the present
 - Authority to act
 - Variety: individual concerns; technical set-up; organisational set-up; IT literacy or support
 - Frequency and scope of accessions; dealing with duplication
 - What about the paper, audio, video, photographs, etc.?
 - Opportunity to acquire valuable contextual information
 - Contemporary formats are easier to access and normalise

Early-intervention: outcomes

We have developed

- Good relationships with creators
- Surveying techniques
 - Context and content
 - Technical
 - Legal
- Deposit documentation
- Guidance for record creators
- Secure and authentic archive extraction, transfer & storage
- Procedures for extracting data from popular desktop software/web-based services
- Some familiarity with forensic data extraction and analysis tools, especially for older material

Conclusions

- A worthwhile approach
 - Individuals have lost material!
 - Can obtain excellent context
- But relies on
 - Headhunting individuals
 - good will and trust of individuals
 - Sustaining relationships over long periods of time
 - May produce **different** collections
- Digital archaeology inescapable
- Need to repeat with other groups
- Not the only way. See 'Approaches to Collection Development' section in Paradigm Workbook



SCHEDULE OF TRANSFERRED MATERIAL



Media ref. no. <i>[Ref. no of CD-R or USB stick, e.g. CD-R-1]</i>	USB 1
MD5 checksum(s) <i>[record values of MD5 checksum(s)]</i>	e072dad61e152c96889d3e117038e5741mydocs.zip
Extent <i>[In bytes]</i>	I 32.7
Technical description <i>[description of file formats, passwords]</i>	Word files and Excel spreadsheets created by Joe Bloggs using Microsoft Office 2002. No pass words required
Content Description <i>[Covering dates, subjects, record types, etc.]</i>	Folders include: Press releases Constituency party information General Election 2005 General Committee reports Annual reports

Pre-Ingest workflow

Accessions server – initial appraisal and packaging for storage

- Virus check and quarantine
- Assign object-level hash values
- Identify archival files and maintain original order
- Identify and validate formats (DROID/PRONOM, misc registries, and JHOVE)
- Ensure that incidental copies of archives are securely deleted
- Delete duplicate files, system and software files
- Add information on new formats encountered to PRONOM, etc.
- Build METS submission for Ingest

METS & Information Packages

- METS acts as a wrapper for digital object and metadata (descriptive, structural, administrative and preservation) to form Information Package
- Ideal for OAIS Information Packages because it keeps all the metadata about an object together in one place in XML
 - SIP: Submission Information Package (user-supplied)
 - AIP: Archival Information Package (for preservation)
 - DIP: Dissemination Information Package (for access)
- For a simple overview of OAIS see
 - <http://www.paradigm.ac.uk/workbook/oais/index.html>
- We will be adding an introduction to METS soon

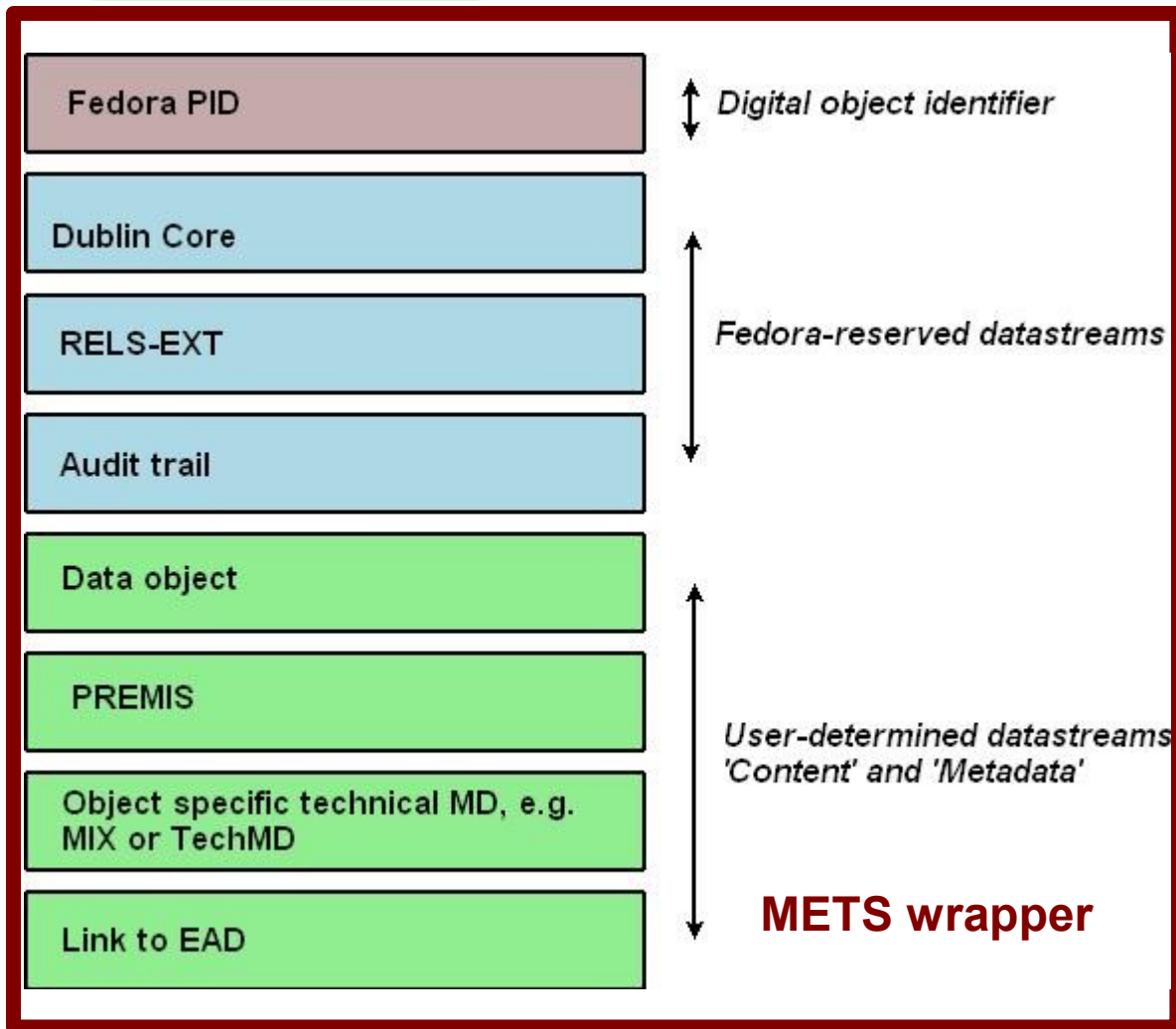


Brief diversion to Cairo...

- Putting METS AIP together for digital archivists is difficult, so..
- Cairo will develop a user-friendly ingest tool for archivists to ingest
 - Complex collections of born-digital materials
 - Consist of several object types, some complex themselves
 - Where relationships are important
 - For a preservation context
- By involving archivists (end users) in specification and testing
- Using existing metadata extract tools
- Map output to preservation metadata standards (PREMIS and object-specific) in METS package for Fedora submission via DirlIngest according to content models



Cairo METS AIP



- For preservation, *not* dissemination
- Atomistic objects + RELS-EXT + container objects
- Enable batch preservation actions on 'at risk' objects
- Base model + object-specific technical metadata stream



Preservation Strategy (1)

- Digital preservation = monitoring/planning + intervention
- Monitoring and planning
 - The basics for bit-level preservation
 - Knowing what you have (good preservation metadata)
 - Knowing what you can do with it
 - Knowing what others are doing
 - Planning interventions ahead of time
- Common intervention strategies include:
 - Normalisation
 - Migration
 - Emulation

Preservation Strategy (2)

- Recommend that preservation strategies be developed
 - In-line with community practice
 - Need for shared knowledge base
 - Dependence on community for some tools
 - Metadata should support multiple strategies (PREMIS)
 - Don't know what tools will be available in future
 - Strategies may change
 - Technology Watch should be:
 - Local (knowledge of collection profile)
 - Distributed (sum of parts greater than the whole)
- Timing of preservation interventions
 - [Always retain original version]
 - Normalisation on ingest for older and obscure formats
 - Delay intervention for contemporary formats until 'at risk'

Opening up – DIPs and EAD

- Sections of the archive can be opened according to expiry of donor concerns and legal protections for third parties
- Will create one EAD catalogue to unite analogue and digital material
- EAD catalogue will link to Dissemination Information Packages: the 'contemporary version' of the object plus some item-level metadata
- Initial access likely to be in reading rooms until IPRs expire and material can be mounted in online repositories

Some of the Challenges Ahead

- Simplify ingest for archivists - Cairo!
- Bring preservation monitoring/actions to the repository
- Develop formal content models and related XACML policies for our objects
- Work with other kinds of creator and their archives
- Integrate digital archives into existing policies for archives
- Provide controlled reading room access
- Create and enhance directories of conversion tools, etc.

Questions?

- Ask me now

- Or later:

Susan Thomas (Project Manager, Paradigm & Cairo)

Oxford University Library Services

Osney One Building, Osney Mead

OXFORD OX2 0EW

Web : <http://www.paradigm.ac.uk>

Email : susan.thomas@ouls.ox.ac.uk

- Tel: 01865 283821