PREMIS Implementations at the British Library
&
PREMIS and the Planets Project

Angela Dappert
The British Library
PREMIS roundtable, February 2009
Overview

- PREMIS Implementations at the British Library
- PREMIS and the Planets Project
- Example: Creating Metadata Profiles for eJournals
PREMIS Implementations at the British Library

- Used for eJournals and newspapers
- Will be used for further content types
- Blue-print for capturing preservation metadata for archival information packages
- Not prescriptive: flexibility means making decisions
PREMIS Implementations at the British Library

- Parts need to be adjusted for each content type: e.g. preservation object data structures
- Parts are re-used across content types: e.g. agent and event model, handling metadata updates
- Embedded within METS
- Various descriptive metadata: e.g. MODS for eJournals, means capturing content-type specific relationships differently
Parts of Planets work …

- … reuse PREMIS ideas
- … could influence future PREMIS versions
- … are orthogonal to PREMIS
Reuse of PREMIS ideas:

- agent
- event
- intellectualEntities, representations, files (bytestreams)
- maybe bitstreams
  - logical components = physical bitstreams or logical bitstreams (?)
Parts could influence future PREMIS versions:

- Characteristics, significant properties
- Environments: PREMIS identifiers have no identifiers – no specification about which of multiple environments applies to which event
- Requirements / business rules
  - (as key concept supporting preservation processes)
  - as provenance metadata
- push towards machine-actionable representations
- ...
Parts are orthogonal: Supporting preservation processes rather than metadata capture and exchange

- Characterisation:  
  XCEL: language for specifying how to extract properties from a given file format

- Preservation actions:  
  e.g. ingest and transformation file sets, transformation pathways

- Preservation plans – machine-actionable
Creating Metadata Profiles for eJournals

Goal:

For eJournals:

■ Create metadata profiles for eJournal archival information packages (AIP)

■ Store metadata in the repository with the content to create complete, self-descriptive units
3 questions:

1. Which objects do we describe?
   a. Which?
   b. How many?

2. Which metadata do we need?
   a. Which do we need?
   b. Which do we get?

3. Which standard do we use for which metadata?
Creating Metadata Profiles for eJournals

Answers are based on analysis of the

- Concepts in the domain
- Technical architecture
- Use Cases
  - Workflow
  - Functions supported
- Sources of objects and metadata
Question 1a: Which objects do we describe?

Objects include intellectual entities, representations, files, bitstreams

For eJournals:

- Journal, Issue, Article
- Representation: usually several representations (provider specific, XML, HTML, PDF, not identical content)
- Submission: packages contain all the content files, metadata, manifests; for convenience, records provenance information (events) that are shared by many files
Question 1b : How many objects do we implement?

For eJournals:

Because of the write-once architecture of the Digital Library System, we split objects into chunks which are updated together. This avoids, for example, creating new generations of journal objects with every submission of a new issue.
Question 2a: Which metadata do we need?

- Which functions are supported by the system and what information do they need? (Preservation, technical requirements, resource discovery, management information, reading room, …)
  - This is hard: little experience with digital objects, uncertain technical possibilities, uncertain future legal framework in which we will operate
    - best guess on the future

- Which relationships exist between objects?

- Which events, agents, rights do we describe?
Question 2b: Which metadata can we get?

For eJournals:

- Many suppliers of eJournals to one repository
- Formats of metadata and content are out of the control of the repository
- Translators to the internal metadata format need to be written
- To guide the writing of translators, the metadata profiles need to be very precise so that the translators will produce high-quality, uniform metadata
Question 3: Which standard do we use for which metadata?

For eJournals:

- **METS:**
  - Structural relationships between files
  - File location
  - Digital library system identifiers
  - Basic technical metadata
  - Bundling up remaining metadata
Question 3: Which standard do we use for which metadata?

For ejournals:

- **MODS:**
  - Descriptive metadata
  - Non-actionable, descriptive rights
  - Hierarchical, structural relationships between intellectual entities
  - Identifiers of intellectual entities
Creating Metadata Profiles for eJournals

Question 3: Which standard do we use for which metadata?

For eJournals:

- **PREMIS:**
  - Events (provenance)
  - Agents
  - Basic technical metadata
  - Specific technical metadata
  - Identifiers for AIP generations