Fedora 3.0 and METS: A Partnership for the Organization, Presentation and Preservation of Digital Objects

Open Repositories 2009
@ Georgia Tech, Atlanta, GA
Thursday, May 21, 2009
Nancy J. Hoebelheinrich, Stanford University Libraries
Patrick Yott, Brown University

Topic Outline

- Task = How to integrate existing practice (METS) with Fedora content models

- METS at a gallop

- Context for Fedora use at Stanford

- Fedora / METS mechanisms to facilitate digital object registry / digital library stacks / preservation repository functions

Task = Integrating existing practice with Fedora

- METS as standard, adaptable mechanism for organizing, linking MD & content
  - Few differentials among content encodings, so relatively automated creation
  - Capabilities for:
    - Using identifiers to facilitate linking / locating among object / component files, and tracking among systems / repositories
    - Collecting, typing both assigned and generated MD
    - Describing physical & logical structure of D.O. at desired level of granularity

METS at a gallop…

- Metadata Encoding & Transmission Standard
- XML specification used in many digital libraries as means to package content & MD of all types
- Often used for OAIS SIPs, DIPs, AIPs

- Flexible
- Limited requirements to validate schema
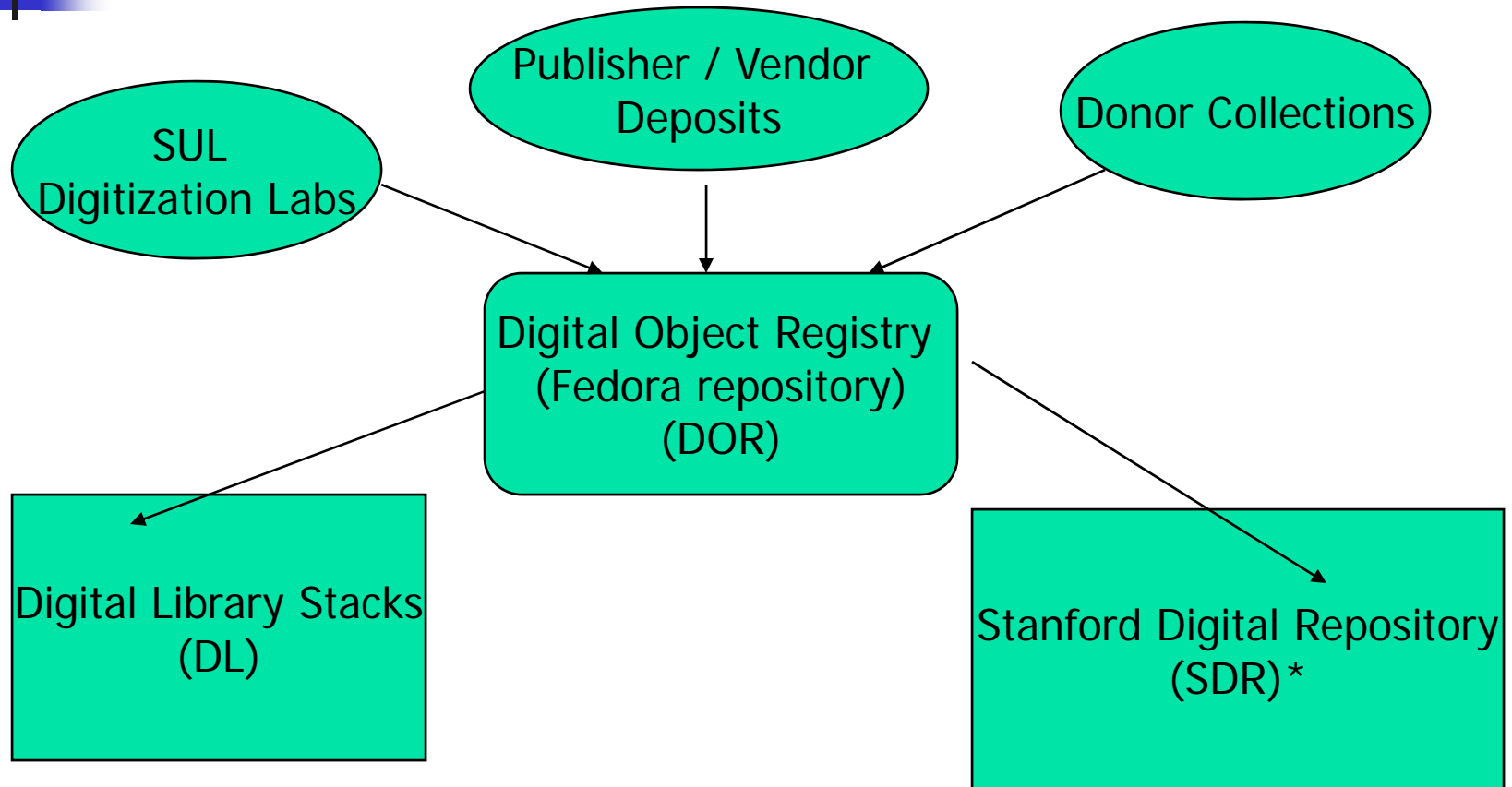- Has some built-in capability for use with Fedora

METS at a gallop...

- Requirements to validate:
  - 1 layer within a structure map
- Capabilities:
  - File inventory & grouping
  - Describing structures
  - Organizing metadata types
    - Descriptive
    - Administrative
  - Associating behaviors
  - Linking between structural maps

- Declaring specializations via profiles
- Typical content types encoded:
  - Still images, text, audio, video
- Typical Institutions using:
  - National libraries, research libraries archives, museums, govt agencies (internationally)

Context for Fedora at Stanford

Publisher / Vendor Deposits

Donor Collections

SUL Digitization Labs

Digital Object Registry
(Fedora repository)
(DOR)

Digital Library Stacks
(DL)

Stanford Digital Repository
(SDR)*

* probable Fedora repository for MD management

DOR Functions (from MD POV)

- ***Accession*** into digital management environment
- ***Registration*** of D.O. with IDs for linking / consumption by other services, repositories
- Facilitation of ***D.O. management***, versioning of MD for D.O. and for components
- ***Source of record*** for:
  - non-preservation, administrative MD
  - link to sources of descriptive MD (e.g., online catalog(s), EADs)
  - D.O. resolvable location

DL Functions (from MD POV)

- ***Link*** digital format of D.O. to analog, if any
- ***Identify*** formats, copies of D.O.
- For descriptive MD, ***enable connections*** among:
    - search & discovery systems
    - D.O. / component delivery /rendering

SDR Functions (from MD POV)

- ***Maintain provenance*** of D.O. (trail of ownership, creation, modification)
- Keep ***record of significant preservation related events***
- Links to ***storage location (s)***

Fedora Content Models for [Google] scanned books [1]

- Fedora content models= [Google] scannedBook
  - MD Datastreams =
    - RELS-EXT (has Model)
    - dorAdmin (has list of key IDs)
    - descMD (DC & MODS)
    - ContentMD (fileSec & structMap)
      - File listing
      - File grouping by format, use
    - AdminMD (including techMD & premisObject MD)
    - RightsMD
    - ProvenanceMD

Fedora Content Models for [Google] scanned books [2]

- Fedora content models= Page (scanned book page)
  - MD Datastreams =
    - RELS-EXT (hasModel)
    - descMD = DC
  - Content datastream = file / bitstream

DOR functions enabled:

- **METS**
  - Identifies D.O. as managed unit
  - Inventories component files/bitstreams of D.O.
  - Organizes components into groups for various purposes
  - Outlines physical and/or logical structure of D.O.
  - Types & links metadata with D.O. & components

- **Fedora:**
- **xmlContent datastream includes:**
  - mets:fileSec
  - Parsable mets:fileGrp
  - Consumable mets:structMap & attributes (e.g., LABEL)
  - ID/IDREF links
- **RightsMD datastream**
- **ProvenanceMD datastream**

DL functions enabled

METS

- Allows inclusion or reference to "source" MD, e.g., analog catalog record
- Groups and sequences component files into groups for presentation
- Provides labeling information to facilitate presentation of D.O. (e.g., page number)

Fedora

- descriptiveMD datastream includes MODS, other content standard descriptive MD
- Actionable content datastream via mets:fileGrp USE
- Consumable content datastream via mets:structMap & attributes (e.g., SEQ, ORDER, TYPE)
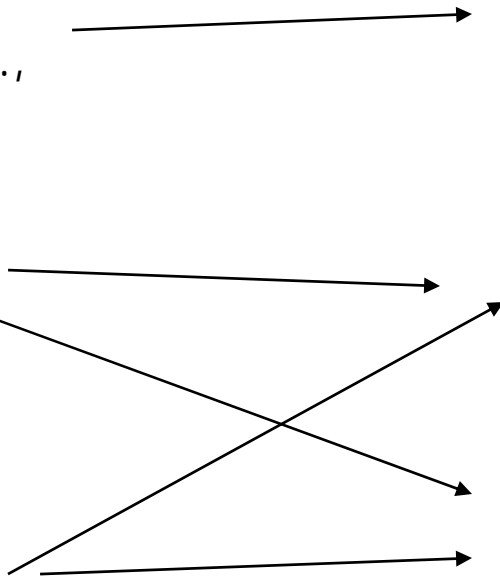
SDR functions enabled

METS

- Allows inclusion of "preservation" metadata such as:
  - Fixity, validation info
  - ID generation / retention throughout workflow, if desired
  - References to storage location, logging MD
  - Changes to filenames
  - Descriptions, recording of significant events, dates & rights agreements

Fedora

Admin datastream includes:

- mets:techMD for file/format specific MD
- premisObject to record original filenames, references to location info, etc.

ProvenanceMD datastream includes:

- mets:digiprov / premisEvent to record significant preservation events (validation, format migration, etc.)

Rights datastream links to Service Level Agreement

14

Key MD Management functions enabled for DOR:

- Takes advantage of Fedora's versioning for both MD streams and DO/component files during creation, use

- Enables creation & tracking of locally and publicly persistent IDs

- Allows SIP, DIP creation at different points in DO's lifecycle (i.e. to SDR, DL)

Key Management functions enabled for SDR:

- Retention of SIP as received from data provider while also allowing

- Derived "lightweight" AIP containing only MD necessary for reconstruction of package or component files within

- Referenced preservation MD that can be versioned separately from AIP

The view from Brown University …