Archive Ingest and Handling Test

Tim DiLauro
Johns Hopkins University
<timmo@jhu.edu>

DLF Fall Forum 2005
Charlottesville, VA
November 2005

JHU Motivation

- Investigate content repositories as platforms for digital preservation activity
- Evaluating the feasibility of implementing an higher level API over multiple repositories
- Learning more about Fedora and DSpace
- Validating this minimal level of “preservation”
- Continuing involvement with NDIIPP

JHU Approach

- Current or emerging metadata standards (MODS, digiprovMD, METSRights)
- No manual processing. Bulk operations only.
- Maintain relationships between objects
- Application-agnostic repository layer
- Both Fedora and DSpace repository applications
- METS as the metadata container

Phase I: Ingestion

- Develop very simple high-level data model
- Implement METS-based interface to support the model
- Map interface to repository functionality
- Map content and metadata from 9/11 Archive to our SIP format and load

Phase II: Export

- Instantiate data model in file system for DIP
- Export our version of the archive
- Import someone else’s
- Three of four participants used METS, just not the same METS profile.

Phase III: Format Migration

- Generate new TIFF version for each JPEG item, but…
- Focus on bulk mechanism; actual transform not that important to us.
- Used MIME Type for selection, but format registries might allow finer granularity
- Keep all versions of a given object

Repository as Preservation Platform

- Repository improves preservation chances …
  - Content in well-known, central location
  - Consistent data model and API eases bulk operations like ingest, export, migration, and audit
- However …
  - API different for each repository application. Application-agnostic interface can help, but integrity problems if bypassed.
  - Agreement on data models / semantics still a problem.