Using An Open Source Institution Repository for Preserved Email

Preserving Historical Correspondence:
Email Preservation Progress and Future Directions

November 10, 2008

Ricc Ferrante & Lawry Persaud

Requirements or Business Needs

- Reliable storage of archive packages (AIPs)
  - AIPs consisting of multiple files as a single entity
- Support/enforceable taxonomy
- Security
- Manageability
- Accessibility
- Affordability

Citation
Ferrante, Riccardo, and Lawry Persaud. (2008). "Using An Open Source Institution Repository for Preserved Email." Sleepy Hollow, NY and Washington DC: The Collaborative Electronic Records Project.

Selecting DSpace

- Open source with strong community and install base
- Lots of out-of-the-box functionality
- Possible dual purpose
  - Archival storage
  - Dissemination
- Robust security configuration
- Multi-items Titles
- Dublin Core metadata support

Gaps / Opportunities

- Hierarchical file relationship
- Package (multi-item title) import

Gap/Opportunity #1

- **<u>Hierarchical relationships</u>**
  - Schema solves the hierarchical relationships internal to an email account
  - Still have issues where other records must maintain a hierarchical problem, e.g., attachments.
    - Addressed by 'zipping' those sets for DSpace ingest – consequence: zipped content is not accessible through the DSpace web UI.

Gap/Opportunity #2

- **Package import**
  - Email AIP consists of multiple files, even after certain sets are zipped up.
  - Need for a different interface, enhanced
    - Title and description
    - Title's items identified, technically described
    - File structure relationship

Solution = METS Import Module for DSpace

Using METS

- DmdSec – Descriptive Metadata
  - Dublin Core tagging
  - Email Account (aka Title) level information
- FileGrp – Files included
  - Filename, MIMEType, size, creation date
- StructMap
  - Files arranged into groups

METS Import Results

- Multi-item Title is created for the account
- Dublin Core fields are used for title description
- Selected DC fields populate search indices
- AIP files are attached as separate items hyperlinked from the Title display page.

Overall Results

- Archival Storage
  - Package or archival object paradigm is supported
  - Some limitations on how
    - ZIPs of directory trees

- Search & Retrieval
  - Finding Aid equivalent content displayed
  - Individual AIP files are somewhat accessible
    - Browser plugins / helper apps
    - Download and open option

Lawry Persaud
Consultant
Lawry.persaud@mail.rockefeller.edu

Ricc Ferrante
Smithsonian Institution Archives
FerranteR@si.edu
202-633-5906