

Session title: Developing a Content Management System for a Federated Digital Library Using Open Source Tools

and

Disruptive Components: Re-imagining Finding Aid Delivery on the Web

and

New Strategies for Managing Manuscripts in Digital Libraries

Presenter(s): Dr. Michael Slone, Digital Library Services, University of Kentucky

Eric Weig, University of Kentucky

and

Shaun Ellis, Princeton University Library

Regine Heberlein, Princeton University Library

Jon Stroop, Princeton University Library

and

Bradley Daigle, University of Virginia

Mike Durbin, University of Virginia

Time of session: Monday, November 5, 9am-noon

Room: Continental A

Notes on session: Developing a Content Management System for a Federated Digital Library Using Open Source Tools

Challenges and goals:

Search / load everything together

- experimented with Dublin Core in Blacklight
- METS with DC for 3 loading scenarios
 - METS freeform (simple compound objects)
 - EAD + “gossamer tree” (archival collections)
 - associated media files
 - METS + [Oral History Metadata Synchronizer \(OHMS\)](#) XML

Introduce greater degree of automation

- post archival capture - such as image capture & basic metadata
- all other aspects of load are automated
 - sip -> aip -> dip (using FCLA/FLVC’s DAITSS)
 - using bagit packages for SIP structure
 - loader creates AIP and DIP
 - METS production
 - generation of derivatives
 - indexing in SOLR

Handle NDNP data

- 600k historic newspaper pages so far

- devised method for reforming NDNP dataset but not identical with METS/ALTO
- extract bounding box data from born-digital PDFs from current newspapers in ALTO

Use OHMS oral history viewer

- oral history metadata system to gather/edit metadata and synchronize with subject index or transcript
- developed over 4 year period; current development under IMLS NLG; will be open source

UK has lots of experience with mass dig of newspaper/serial formats

Archival description is expressed within EAD

Past process for preparing collections for online delivery has been manual

Issues with digitization of archival collections has been lack of integration presenting description and digitized objects

UK Model for delivery

- Collection targeted for digitization after processing
- Description in Archivists' Toolkit
- Gossamer is a tool that builds a directory tree for digitized objects
- Digitization of collections at folder or item level depending on size/significance - decision made by archivist, not digital library group
- Resulting media files placed in Gossamer directory tree, which is used to generate tree
- DIP indexed in blacklight
 - Integration of EAD with METS allows for toggle between views: description & page turner
- <http://eris.uky.edu/>

Notes on session: Disruptive Components: Re-imagining Finding Aid Delivery on the Web

- Princeton redeveloped finding aid system - new site: <http://findingaids.princeton.edu/>
- Wide variety of physical document structures of archival descriptions
- EAD originates in document-centric tradition; now maintained by SAA's Technical Subcommittee for EAD
- Information that is common across levels of description should be described at the highest level where they apply
- Linear nature of finding aids gets translated to web-based presentation: static HTML or PDF
- Boilerplate or repeated data: administrative information/contact information, descriptive standards, context/biographical information
- using ID attributes instead of hardcoding data
- Search returns components at every level; not just collection level. Better than KWIC for the entirety of finding aids.

- Princeton's component-level description has stable URIs and provides at least minimum-level DACS description for ANY LEVEL.

Architecture:

URI templates were developed early on in the project and allowed for consistency and clean code
topics infrastructure using SKOS closeMatch and exactMatch to tie back to vocabs like LCSH

User experience

Developed goals statement

Path to usability: reduce visual complexity and increase prototypicality

Discount usability testing: identify personas and recruit; create tasks for testing scripts

Use an ethnographic approach

"Stubs": help system (wordpress); google analytics using evnts

Prototyping: see Shaun Ellis and Maureen Callahan, "Prototyping as Process for Improved User Experience with Library and Archives Websites," *Code4lib Journal* 18 (October 3, 2012):

<http://journal.code4lib.org/articles/7394>

Notes on session: New Strategies for Managing Manuscripts in Digital Libraries

- This is a discussion about how to leverage Fedora / Solr / Blacklight to manage and deliver manuscript data
- no clear method to apply finding aid templates