

SW Arch Meeting Minutes – May 3, 2012

Agenda

- Announcements and Updates
- Progress on Non-release projects
- Fedora 3.5 on staging
- Complex objects- user interface
- Multiple techMDs – continuation
- Background ingest and alerting
- Release summary discussion

Announcements and Updates

It appears that we have year-end funding for Wowza and related server hardware. Ron indicated that we may want to consider another dot release to update the OAI harvesting function. Ron and Chad tested the API for harvesting and Ohio Link also tested it on our development server. The API update is a minor change and would have no impact on the rest of the system. Ron mentioned that we should include high impact, simple changes in the faculty deposit interface in our next release (R7.0). Basically, these are “cosmetic” changes that can be made easily and do not require a specification. The user interface group will identify those changes that can be considered “cosmetic”. We did not discuss the recent Fedora conference at the University of Connecticut – Chad will set up a brown bag where those who attended can discuss their experiences.

Non-Release Related Updates

The status of each project is summarized below:

- JPE. It appears that the update to the outputsmil module is working (e.g. in handling direct links to JPE videos from IRIS). Dave and Chad will pursue the one remaining “refresh” bug – it may be a configuration issue. Dave mentioned that he is seeing messages in the php error log that still require some explanation.
- XML-1 Datastreams and Jpeg Thumbnails. The scripts have been updated to handle the case of an embargoed resource. The datastream and jpeg scripts have been run on staging and results verified by Rhonda. We agreed to move ahead and do the update on the production server; the scripts will be run off-hours. Jeffery will continue to address the remaining issue of Fedora objects that are text-based but do not have an xml-1 datastreams. A diagnostic procedure will be required to identify these objects and add the appropriate xml-1 datastream.

Fedora 3.5 on Staging

On Monday, R6.1.2 will have completed the post-release 2 week window. Dave will move ahead in the next week or so to install Fedora 3.5. The Software Architecture group will test RUCore R6.1.2 on Fedora 3.5. At this juncture, we will not be using any of the new features in Fedora (e.g. FESL) but we should benefit significantly from bug fixes. If testing goes well, Fedora 3.5 will be installed on our development server and we'll begin development on RUCore R7.0

Complex Objects – User Interface

Chad presented the 3rd installment of the complex objects specification that deals with the user interface. Two scenarios were discussed as download options: 1) download the entire complex object which would be packaged as a zip file or 2) give the user a display representing the directory structure of the object and allowing the user to select various folders and files. The user interface would also provide feedback on the size of the download and the estimated time for download. We agreed to pursue the use of Bagit as a file packaging standard. Several issues were raised for further investigation. We need to understand more fully how the datastream label field will be used. In the absence of meaningful file names from the user, the label field might include a designation that indicates what the file is about. It was also suggested that the download manifest might be displayed to the user before beginning the download. In response to Dave's question, Chad indicated that Google could not request a download of a complex object. There also remains a concern about the time it will take to package very large files as bagit zip files in preparation for download. In general, we considered the specification final subject to a review by the Data Group and CISC. Aletia will provide feedback from the Data group in our next sw_arch meeting (May 17).

Multiple techMDs

We continued the discussion on how to handle technical metadata for multiple archival masters. We revisited an earlier decision and decided to retain part B (digital file relationship) of Yang's specification. This relationship would be represented as an rdf block in rels-int that relates presentation files (e.g. jpeg1) to the respective archival master.

The discussion regarding how to deal with technical metadata for multiple archival masters presented some difficult technical issues. We did decide on the following: 1) we will use rels-int to represent the relationship between a techMD block and its respective archival master, 2) the rels-int will be implemented as a single managed datastream to minimize the size of inline XML for the object and 3) we will have to manage the editing of the rels-int. For example, if an archival master file is purged, we will have to reflect this change in the rels-int (Fedora provides no help for this function).

We then began discussing issues that can best be described as "messy" and ones that we have ignored in the past.

- Consider a book with 10 pages and 10 associated tiff files. Assume that the tiff files were created on different platforms (operating system, scanners, etc). Even though these files are all of the same type, do we want to have technical metadata that indicates the different platforms? Note, that we had originally thought we could have one techMD block for all 10 files, especially given that we are moving to include the size, mime type, and checksum as datastream attributes.
- How do we handle the more complicated metadata user interface? As in the book example, we do not want the user having to deal with multiple sections of technical metadata if there is no need to supply additional metadata.
- From our last meeting, we decided to move ahead with using the ExifTool and Mediainfo tools to extract technical metadata. However, these tools give us the opportunity to include much more metadata than previously. What additional technical metadata do we want to include? Rather than including all additional data we might consider creating an additional datastream for all of the extracted metadata (e.g. by use of sidecar files which have the same base name as the source file, but with a different extension). This approach would preserve all of the technical metadata but not encumber the user interface with data that may not be important to display. Isaiah will review this approach with MDWG and report at our next meeting.

Ron suggested that it might be better from an architectural perspective to have a techMD block for each archival master regardless of whether the object were simple (e.g. the book) or a complex object. For the book example, others suggested that we might have one “super” techMD block that represents unique metadata for the book and multiple techMDs for each tiff that are generic (i.e. do not require any user input or editing). Unique metadata could be retained in the sidecar datastream.

Other Items

As a final note, we decided that we would eliminate the METS ingest option from WMS – it is no longer supported in Fedora. We did not have time to review the background ingest or release summary documents (for next meeting).

Agenda Items for Next Meeting

- Feedback from the Data Group on the complex object user interface - Aletia
- Multiple techMDs - continuation (use of ExifTool and MediaInfo tools) – Yang, Isaiah
- Status of Non-release projects
- Background ingest and alerting - Kalaivani
- Release summary discussion
- Pending
 - RUetd – WMS update

- Enhanced UI for the landing page (a possible framework) – Chad and Jeffery
- WMS – creation of relationships (rels-ext) for data projects

rej – 05/14/2012