

SW Arch Meeting Minutes – April 5, 2012

Agenda

- Announcements and Updates
- Progress on R6.1.2
- Progress on Non-release projects
- File policy and multiple techMDs
- Complex objects
- Data user scenarios
- A/A scenarios

Announcements and Updates

We briefly discussed the possibility of switching to sha-256 for checksums. This change would entail updating previous objects and making minor changes to provide sha-256 for new objects. We could see no downside to doing this. Jeffery will do more testing and we will re-visit this issue in a final review of the checksum specification in our next meeting.

Ron mentioned that the EAD architecture group had their first meeting. The group will be laying out the object and collection structure for EADs and, as a first step, will explore ingesting MARC xml (from Oxygen) via WMS. The group's initial target is to review an architecture specification in August at a CISC meeting.

Release R6.1.2

We will finish testing of R6.1.2 on staging next week and Dave will install it on production when he gets back from vacation. After R6.1.2 has soaked on production for 2 weeks, we will install and begin testing Fedora 3.5.

Non-Release Related Updates

The status of each project is summarized below:

- JPE. We are still having problems with directly linking to JPE videos. Jeffery and Isaiah will do some more testing to see if the issue is somehow related to Darwin or access permissions.
- XML-1 Datastreams and Jpeg Thumbnails. We are still encountering access issues with embargoed ETDs. There is apparently a remaining permissions issue on the new server platform. Dave will need to investigate further. We decided to move ahead with testing the tools for updating xml-1 datastreams and jpeg thumbnails on all non-embargoed resources. We will still need to pursue the embargo issue.

File Policy and techMDs

Regarding file policy, we decided that WMS will manage the basic file and display policies. We also decided that WMS would push out a file to be used by other applications rather than use a web

service. This eliminated an issue where we might be running a read-only system and WMS would not be available.

We also clarified several issues related to technical metadata. Going forward, we will represent basic technical metadata (mime type, size, checksum) in the attributes of the datastream. This data will not be repeated in techMD. This approach will also simplify our archival approach for resources that have many tiff files. For example, with a book, we would have only one techMD and the size, checksum, etc would all be recorded in the attributes of the datastreams for each tiff file. This eliminates the need to have many techMDs for books and similar resources. However, in more complicated situations, we will still need multiple techMDs to record unique technical characteristics. When it is necessary to have multiple techMDs, we will need to relate the technical metadata to the proper datastream. It appears that the best way to do this is to use Fedora's rels-int.

We also decided to pursue the use of file format tools that will enable us to more easily populate technical metadata for the more complex file types. Isaiah will bring a specification to our next meeting that will cover two tools – exif for more conventional document resources and a media tool for video and audio. Isaiah will also do a short demo.

Structural Map for Complex Objects

Chad continued his discussion of the use of the structural map to model directories of complex objects. He presented how a typical multi-level directory with multiple files could be represented in FOXML and how the original directory structure could be described in a METS structMap. We concluded that the ALT_ID attribute could be used to store the original file names and possible DOIs or handles. We need to address the issue when file names have blanks in them. The new structMap for archival files would require a new ID such as “ARCHSMAP1”. The structMap along with the datastreams and appropriate file names could be used to reconstruct the original directory in its entirety. Chad will continue the discussion in our next meeting and focus on presentation which might possibly require another structMap.

Data User Scenarios

Aletia reviewed the work of the user interface group with a focus on complex data. She reviewed four scenarios which are briefly summarized here.

- **Project Search.** How do we deal with the question of when a search hit is buried in one of the documents? This creates confusion for the user since they do not know why the hit occurred. We agreed that we might be able to provide browse features by “related publications” which might help in this area. However, it is not likely that we can readily implement a highlighting feature in documents although doing this for metadata might be a good first step.

- Author search. We do not have any good solutions for uniquely identifying an author's name although this will probably not become a problem for quite some time (i.e. when we have many 1000s of authors). We can in fact search by author name now but we might want to improve the UI for this capability. We can also look at using netid within Rutgers which might be a way to handle unique author's names within RU (more discussion needed here).
- Download entire project. The basic idea here is to allow a user to request the entire project, have it zipped, and be able to download it. This should be fairly straight forward and should probably be one of our basic features.
- Shopping cart. The idea in this scenario is to allow a user to select various files from a complex directory. The UI group will need to look at this more closely to present more specific scenarios and what approaches would be most useful. Can a user select a single file type from many objects? Can a user select a subset of files from an object that has many different datastreams. There are almost an infinite number of possibilities and we will need to identify those that are most useful.

Authentication and Authorization

Ron indicated that the A/A scenarios were reviewed in CISC and everyone agreed that the five scenarios covered the basic requirements for implementation. Ron will work offline to see who might be able to take on the development of the implementation specification.

Agenda Items for Next Meeting

- Finalize the checksum specification and use of sha-1 or sha-256
- Complex objects/structural map specification - presentation (continuation) – Chad
- Specification for file format tools - Isaiah
- Multiple techMDs – Yang
- Fedora 3.5 - Dave
- A/A Scenarios – how to proceed with implementation
- Status of Non-release projects
- Pending
 - Enhanced UI for the landing page (a possible framework) – Chad and Jeffery
 - File Support and background ingest – Kalaivani and MDWG