

## **SW Arch Meeting Minutes – September 13, 2012**

### **Agenda**

- Announcements and Updates
- Status of Non-release projects
- Testing and Release of R6.2 and R6.3
- R7.0 specifications and bug fixes
- Tar or zip for packaging data
- Checksums and applying sha256 to legacy objects
- Object definitions - continued discussion

### **Announcements and Updates**

As an encouraging sign, we received a query from an RU researcher who has an NSF grant and was interested in archiving data in RUcore. Rhonda and Ryan will respond. Hopefully we see more of these types of queries.

Ron noted that the DOI agreement with CDL has been signed and should take effect on November 1. Also, we had an encouraging meeting with Professor Corey Brennan regarding the Roman Coins project. Corey will draft a grant application, due November 1, to apply to the Loeb Foundation (maximum grant is \$35,000).

### **Status of Non-release Projects**

Due to time limitations, we didn't get to this item. The JPE videos and jpeg thumbnail projects are still pending.

### **Testing and release of R6.2 and R6.3**

Dave will install the recent update for R6.2 on staging and we will do a quick test. The release should go to production on Friday (Sept. 14). Chad will follow up with Dave to push out R6.3 (Analytic) during the week of Sept. 17.

### **Tar and Zip Files for R7.0**

We had a good discussion of the issues which are too lengthy to report here. The basic decisions we made are the following:

- There is no need to compress files on upload since we speculate that many of the files will already have gone through some compression and we would probably only gain with predominantly text files. Also, a compression step would incur more delay and use cpu cycles. So, we will not compress files.
- For packaging the download, we acknowledged that the tar and zip formats both have disadvantages. The tar format is likely to be unfamiliar to Windows users and the zip software has problems handling files larger than 2GB, resulting in file corruption on XP. We decided to go with

the tar format with the rationale that 1) it is reliable and handles large files, 2) scientists that deal with large files will likely be familiar with the tar format, 3) tools are available on Windows (e.g. Winzip) to unpack tar files and 4) we will have liaisons who can help users if need be

- Regarding maximum limits for downloading, we decided to set the limit at 1000 files and 100 GB. Ryan felt that this limit would easily cover the primate tooth project. It is also noted that the limit can be easily reset in the configuration file.

There was one architectural issue regarding the delivery of medium and small jpegs and the recording of statistics. If we move to an improved architecture where the getAPI reports all download statistics, we will lose the ability to record downloads for medium/small jpeg disseminations. It should be noted that there are no specific file datastreams for these jpegs. These statistics may be valuable for portals that focus primarily on images such as the map portal. Ron will check with Linda on this issue.

### **R7.0 Specifications**

Although we have reviewed all R7.0 specifications, there are several documents that require changes starting with Yang's workflow/METS metadata and file policy documents.

For the workflow configuration, we need to add "directory" as a file upload option. When a file is uploaded, there will be a choice of directory or file. If directory is chosen, WMS must assume that there could be more than one file type. Since we are not using exif or the media tool in this release, Yang will use file extensions as a best approximation of file type (there are still issues with extensions such as xls and xlsx). We will have multiple techMD sections in one Fedora TECHNICAL datastream. Given that we can have many technical metadata blocks in R7.0, we decided that the technical metadata datastream should be managed content instead of inline. This will reduce the size of the in-memory object and should lead to improved performance in certain areas. For a book (as an example) that will have multiple techMDs – one for each page, we will need to create only one source record rather than a source record for each techMD. For techMD. Yang will insert the file name in "OWNERID" in METS and this will be translated to ALT\_ID in FOXML. For the shared XML file, Yang has added a policyGrp elements for ContentModel and for TypeOfResource.

Regarding file policies, there will be a file processing rule for every archive file type. After much discussion, we decided that we will need some sort of default rule when we have an unknown file type. It was decided that the default policy could be a part of the next release when we integrate the exif and media tools.

Yang indicated that the proposed changes would probably add a week to his development schedule. Jeffery will look at the issue of what we need to do to handle both inline and managed technical metadata and ultimately convert all technical metadata to managed.

**Checksums and SHA-256**

We did not have time to address this item and it is on the agenda for the next meeting.

**Object Types Discussion**

We did not have time to address this item and it is on the agenda for the next meeting.

**Other Items**

Ron requested that all developers report on their proposed bugfix activity for R7.0. Although we didn't have a chance to discuss what should be reported, the suggestion is that each developer indicate that number of bugs outstanding and specifically, if the high priority bugs can be fixed in R7.0. (Thanks to Jeffery and Isaiah for their notes on the details discussed in this meeting and which were extensively used in this document).

**Agenda Items for the Next Meeting**

- Status of Non-release projects
- Status of R6.2, R6.3 and R6.5
- Review of bugfixes for R7.0
- Checksums and applying SHA-256 to legacy objects
- Jpeg2000 specification
- Retention policy
- Object definitions – continued discussion
- Pending agenda items
  - RUetd – WMS update
  - DOI specification
  - Enhanced UI for the landing page (a possible framework)
  - WMS – creation of relationships (rels-ext) for data projects