

SW Arch Meeting Minutes – October 11, 2012

Agenda

- Announcements and Updates
- Status of R6.5 – Fedora 3.5
- Checksums and applying sha256 to legacy objects
- Review of R7.0 code complete date
- Process for reviewing bugs
- Object definitions - continued discussion

Announcements and Updates

Ron voiced a concern about our ingest volume which averaged about 330 objects/month for the third quarter with a peak of 540 objects ingested in September. We have a total of approximately 27,000 objects in RUcore whereas some other research libraries have around 100,000 objects. Since the total number of objects represents an important marketing indicator, Ron suggested two projects that other libraries are initiating: 1) scanning and ingesting print ETDs and 2) ingesting our faculty articles that have been previously published in commercial journals. It was noted that the NJEDL review may add another 1000 objects. It was also noted that our work in software architecture has been on the design and implementation of RUcore and we may have to rely on others to initiate these types of production projects. Ron will query Grace to see what her view is.

We don't have a specification for producing the personal faculty bibliography. Rhonda will follow up with the user services group to see if a specification can be developed in the R7.0 timeframe. This capability would have to be targeted for a post 7.0 release.

Ron mentioned that we have tested the API for harvesting records from RUcore using OAI-PMH. We decided this feature did not require a quick fix so will target it for R7.0. We did decide to do a quick fix for portal browsing in a dot release – R6.3.1. Chad will follow up on this item.

Kalaivani raised an issue regarding workflow for large files, noting that it took about 8 minutes in WMS to upload/copy a 10 GB file. Kalaivani will work with Isaiah to bring in process description for next meeting so we can decide how to address this problem. Warning notices to users in WMS might be one approach.

Status of R6.5

Dave will work with Chad to introduce the quick fix noted above and then install Fedora 3.5 on staging. He will try to take care of both of these items next week.

SHA-256 and Unpacking tars for Legacy Objects

We continued the discussion on how to update legacy objects with sha-256 and to unpack tars into individual archival master datastreams. Jeffery noted that there is a lot of variability in legacy object masters – some have only a d-master and there were many non-standard naming conventions. We clarified that we need to preserve the identification of masters and d-masters as we move to individual datastreams. We decided not to use the object ID to make this distinction, rather we will use the label field and the structural map for this purpose. Therefore we will continue with the naming convention of ARCH1, ARCH2, etc and will not use the ID to embed any further intelligence regarding the object.

It was noted that WMS is doing all of the object processing necessary for updating the legacy objects, i.e. creating individual datastreams with appropriate IDs and constructing the structural map. Yang will share this code with Jeffery; we should be able to use the same code in processing the legacy objects. Regarding the non-standard object configurations, we agreed to take whatever is represented in the legacy tar file, i.e. there will be no attempt to correct for any past violations of our current standards.

Code Complete Date for R7.0

To allow developers to address the complexity in handling the ingest of directory type objects and to accommodate vacation schedules, we decided to move the code complete date to Nov. 16, 2012. As part of this discussion, it was noted that the user interface must be able to present many archival masters (e.g. 100s of ARCHs) in a coherent way. Also, we will need to update all legacy objects to replace the title in the label attribute with the ID (e.g. PDF1). The batch script for this capability can be executed as soon as it is ready. Jeffery will bring a mini-spec for these features into the next sw_arch meeting.

Process for Reviewing Bugs

Ron proposed that we should review outstanding bugs to determine which ones can be fixed in R7.0. We have about 50 bugs that are targeted for R7.0 or did not make R6.2. We agreed that sw_arch members who have outstanding bugs will report in the next meeting as to which bugs can be fixed in R7.0. Everyone should review their pending bugs prior to the meeting and bring hard copies of the list if there are more than five or so outstanding bugs.

Object Definitions - Vocabulary

Ron reviewed a summary of our discussion on object naming vocabulary. We have basic agreement on simple, linked, and context objects. However, there are two classes of networks (examples include the data project and EADs) in which we do not have good naming conventions. Ron will forward the draft to Grace for comment and next steps.

Other Items

Per Kalaivani's request, we will add Marty Barnett to the software architecture list.

Agenda Items for the Next Meeting

- Process for uploading and handling large objects in WMS
- Status of R6.5
- Mini-spec for handling the updating of labels – script and dlr/EDIT feature
- Mini-spec for the dlr/EDIT UI to manage large numbers of datastreams (e.g. ARCHs).
- R7.0 bug review
- Specification for applying SHA-256 to legacy objects
- 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

rcj – 10/22/2012