SW Arch Meeting Minutes – October 25, 2012

Agenda

- Announcements and Updates

- Status of R6.5 – Fedora 3.5

- Process for uploading and handling large objects in WMS

- Mini-spec for updating labels

- R7.0 bug review

Announcements and Updates

We welcomed guests Ashwin and EK (MLS field study study) to this meeting of the software architecture group.

Several members of software architecture attended the Fedora 3.6 webinar. There are some very nice capabilities that will be available in 3.6 including all SOAP methods are now supported in the REST API. A batch command is available allowing a programmer to make one GET call to retrieve multiple object records. The Akubra storage system allows for metadata enabling the storage system to make decisions about whether to store parts of the object in fast or slow storage. There are a number of other important capabilities including improved integration with Shibboleth and better interfaces for allowing users to add their own modules to Fedora. Fedora 3.6.1 was just released in August and still has some minor problems, as noted from comments on the listserv. As a result we will stick with our plan to move first to 3.5 and then hopefully within 6 months or so move to 3.6.x.

Ron mentioned that a new category for SOLR indexing and searching has been added to software libraries. He also added a task to investigate how we might configure SOLR so we don't have to make code changes when a new metadata field is to be indexed (assigned to Jeffery). In addition, we will establish a practice to update software libraries with tasks or bugs that are discussed in the sw_arch meeting.

It was noted, from Dave's earlier email, that the thumbnail update process has been completed but there are still some lingering anomalies to be investigated (e.g. some dissertations do not have thumbnails).

Status of R6.5

Dave was not able to attend the meeting so we don't have an update. Ron asked Ashwin to contact Dave and provide an update if possible.

Process for Uploading and Handling Large Files in WMS

Although we have addressed the ingest process for large files, there are still issues, specifically long copy times, related to the process of preparing large files and uploading to WMS. Kalaivani and Isaiah presented the process as two scenarios – video and non-video. Given the special processing for

video, we concluded that the non-video is where we can make the most improvements. During the discussion, it was noted that a large percentage of non-video uploads (~ 70%) involve copying files to the local server as opposed to the user managing the upload from the client. Dissertations and faculty deposit do not require manual intervention. Ron commented that the large percentage requiring manual intervention is not sustainable as we incur higher ingest traffic (at present we are averaging about 300 object ingests per month). It was generally thought that the manual processing approach was, at this juncture, the best one and we should work to improve it. We decided to take action in two areas: 1) Yang will provide an alert in R7.0 to WMS users, indicating that file copy time could be lengthy (up to 10 minutes for a 10GB file and 2) in R7.x, Yang will provide a background copy function. This capability will allow the WMS user to continue to use WMS will waiting on the copy to complete. Yang noted that the code is basically complete for this background function.

Specification for Updating Labels

From our previous meeting, we had concluded that we need two capabilities to update labels on datastreams: 1) a dlr/EDIT capability that allows an authorized user to change the label on a datastream and 2) a script to update all legacy objects to convert the existing label, typically a title, to the datastream ID. Jeffery reviewed the specification, noting that the dlr/EDIT feature is now working on development and an option is available to not version the datastream (i.e. the version datastream is purged). For the batch script, we concluded that labels for presentation and archival masters should be updated and the script should be run before introducing R7.0. Jeffery noted that we will need to find a way to identify the datastreams requiring labels in order to prepare a list of Fedora PIDs. Jeffery will update the specification and proceed with development.

R7.0 Bug Review Process

We reviewed outstanding bugs for R7.0 and R6.2 to determine if most bugs could be completed in R7.0 and whether there was any impact on the code complete date of November 16. The three areas reviewed were WMS, dlr/EDIT, and portal search. Both WMS and dlr/EDIT had around 17 to 18 outstanding bugs. Yang and Jeffery indicated that they would address all of these for R7.0. It was not clear whether some of the more complex bugs could be completed in R7.0. Jeffery and Yang will provide a final report in the next meeting. Chad felt that most of the portal bugs could be fixed in R7.0. Ron encouraged all to make sure the status is updated in software libraries, i.e. if developers have submitted a fix the status should be changed to "test". We'll review bugs in the next meeting to determine if there are any major bugs that can't be completed in R7.0.

Other Items

It was noted that we will soon have to make a decision about what Fedora release we will use for R7.0. Our original plan was to go with Fedora 3.5 (our R6.5 release). However, we are running out of time to move Fedora 3.5 to development and stabilize it in preparation for the R7.0 release. We will need to consider the possibility of staying with Fedora 3.2 for R7.0

Agenda Items for the Next Meeting

- Code complete status and final bug review for R7.0

- Status of R6.5 and what Fedora release for R7.0

- Pending agenda items

    o   Specification for applying SHA-256 to legacy objects

    o   RUetd – WMS update

    o   DOI specification

    o   Enhanced UI for the landing page (a possible framework)

    o   WMS – creation of relationships (rels-ext) for data projects

rcj – 11/06/2012