SW Arch Meeting August 4, 2011

Agenda

- Announcements and Updates
- Release progress
- URL Migration Spec
- Google Analytics

Announcements and Updates

Several updates on projects were offered. Ron noted that the Yearbook project should be able to proceed with current capabilities that are available in R5.2. At this point, it looks like Djvu will be the best presentation format. PDF files for yearbooks appear to be very large. An initial examination of OCR-ed text indicates that some of the surnames in bold italics are not being picked up. Since surname searching will be popular, considerable manual effort will have to be expended to clean up the ocr-ed text. Kalaivani reported that the last batch on ETDs (252) ingested quite smoothly with very few problems. There are now over 2300 ETDs in the repository. We are poised to bring the large Genome gzip files online. The temporary work-around is to ingest a PDF that points to an external website where the user can download the four large files. There is one outstanding mime-type issue that has to be resolved in WMS before proceeding. Chad reported that he has made some chances that appear to improve performance on lefty64. We agreed that performance is not what it should be and that these fixes should be introduced into the current release.

Release 5.2.1 and 6.x Updates

Progress continues with R6.0 (new hardware platform). Dave reported that SUSE was now running on the HP server with the new storage system. There is work to do with storage redundancy and NFS before we can move to the next step of installing Fedora 3.2 and the R5.2 application code base. Regarding R6.1, we hope to have SOLR/Lucene (v. 3.x) running on lefty64 next week. For R5.2.1, we will need a brief specification for using the /rarch directory for all archival masters under 2GB. Ron will talk with Isaiah about authoring this specification since it will have considerable commonality with how we are currently handling large video files (Note, Isaiah couldn't attend since he is on the AUL search committee which was hosting a candidate).

Google Analytics

Chad reviewed a proposal for using Google Analytics in RUcore. There are probably several ways we can use Google Analytics, e.g by examining search trends over time. We agreed to proceed with Google Analytics and Chad will take the proposal to CISC.

URL Migration Specification

      Chad reviewed the URL migration specification which outlines the syntax to be used for the GET API. Suggestions were offered as to how the syntax for embedded files (e.g. flash video) might be handled. Chad will update the document and add an implementation specification. After discussion the effort for implementation, we should be able to place this capability in a release, possibly R6.2 or R6.3.

Agenda Items for Next Meeting

- Review of release summary
- Video/audio objects with transcripts
- Spec for /rarch

rcj – 08/15/2011