SW Arch Meeting August 18, 2011

Agenda

- Announcements and Updates
- Specification for /rarch directory
- Project registration system
- SOLR/Lucene indexing specification
- Faculty Deposit genre specification update (postponed until next meeting)

Announcements and Updates

Reaper64 will be moved to TSB at 3 pm on Friday. This move will position reaper64 in a production environment where there will ultimately be a backup generator to cover for power outages. This move also facilitate the move of CF sites to the PHP/MySql platform. Eagleton Polls is currently on reaper64 and just needs a bit more testing before we move and decommission the CF site. Jeffery reported that SOLR/Lucene v3.3 is now running on lefty64. He will report on indexing times and the size of the index as compared to Amberfish. Chad reported that he needs another week (by Aug 26) to get the portal code working with SOLR/Lucene. Dave reported on the work to move to the new hardware/OS platform. Isilon storage is available with 51 TBs of space. Dave is running SUSE 11 and most of the software seems to have compiled and is running (e.g. apache, amberfish, php, Java v1.6, etc). Fedora 3.2 has been installed and NFS is working. Sho is working on installing Shibboleth. Setting up the various permissions in this somewhat more complex environment will take a little more time. Dave expects that we can start testing with the existing code base (R5.2) sometime next week. The first step will be to pull the objects from mss2. If that goes well, then we'll pull the some 15,000 objects from lefty64. Kalaivani noted that the China boom project is progressing. The collection consists of videos about economics in China and ten videos are already in RUcore. Isaiah reported that we will start using MP4 files on the new China project. (In a post-meeting comment, Sam McDonald noted that there are about 100 videos, each with a transcript and possibly a biography. Work on the portal and website is progressing; it is expected that the Asia Society will want to comment on the portal before final launch.)

Specification for the /rarch directory

Isaiah presented the specification for using the /rarch directory for all objects with masters less than 2GB. During the discussion, we decided that the extFile xml file is not needed. Hence it will not be created for masters less than 2GB and this file will be deleted by WMS after initial processing for masters greater than 2GB. Isaiah will update the spec to reflect this change. Kalaivani and Yang will discuss offline the move/copy issue with the temp file directory. We still need to address the large file (GT 2GB) including the issue of how to deal with presentation files for those files that only need to be downloaded. Do we have multiple copies or somehow provide access to the /rarch directory?

Project Registration System

Peter provided an overview of the proposed project administration system. The system is intended to capture data regarding all RUcore projects. We noted that it is important to distinguish between "content" related projects and software projects. Software projects are currently documented in the software libraries website. It was suggested that Peter submit his requirements to the ad hoc group that is looking into project management systems. It

is important to have a system that can handle classes of projects (e.g. RUcore projects, non-RUcore projects (e.g. Eagleton), Web projects, etc). Peter is working on a database structure that will provide a good set of requirements for selecting a project management system. There was some discussion about who will use the system. Certainly project managers who are trying to assess priorities and allocate work to various projects will use the system. However, given the rich set of data, it is conceivable that there will be many different users.

SOLR/Lucene Indexing Specification

Jeffery and Chad reviewed the SOLR/Lucene indexing specification. Jeffery has the initial code running to create the XML object from metadata in the Fedora object and post this XML to the SOLR/Lucene index. The process took about four hours for an initial run across all 15,000 objects on lefty64. We decided that the best approach was not to retain portal IDs in the metadata of the object but rather store them in a separate MySql database. This approach avoids much editing of the object metadata when someone creates a new portal or wants to change portal assignments. As a result of this approach, a second pass is needed to update the index with the portal IDs. It was generally thought that this approach was a good one and Chad will create an initial prototype, at which time we will re-visit the specification. A major issue is the ability to have real-time updates rather than wait for the overnight indexing process. We concluded for standard operations of adding an object to a collection or editing the object that the real-time update can be preserved. More discussion will be required after initial prototyping.

Performance Updates and Other Quick Fixes

Chad reported that the performance updates appear to give us a 100% improvement. Assuming the test on mss2 is good, we will plan to do a quick fix of R5.2. It was also agreed that other fixes would be batched with the performance fix including the LDAP (special character) fix, the iFrame bug, and timeout values for flash playing of large videos. One other possible fix is related to a WMS bug where technical metadata is lost in an edit.

Agenda Items for Next Meeting

- Review of R5.2.1 schedule (WMS updates, bug fixes, and /rarch)
- Video/audio objects with transcripts
- Review of the FD genre specification
- Getting started on the large file (GT 2GB) specification
- Review of the url migration implementation specification

rcj – 08/31/2011