SW Arch Meeting Minutes - September 1, 2011

Agenda

- Announcements and Updates
- Progress Updates
  - o Performance
  - o New Platform
  - o SOLR/Lucene
- Release R5.2.1 Update
- Faculty Deposit Genre – update to the specification
- AV objects with transcripts

Announcements and Updates

Kalaivani reviewed bug fixes that are pending including the LDAP "building" fix, technical metadata regarding the "creating application", the faculty deposit fix – creating collections, and the pending performance fixes. We decided to batch all of these bug fixes into a dot release to be called R5.2.1. Sho mentioned that the work area on lefty64 is low on space, having used about 400 GBs in the last testing effort. The space is spread around in different collections; Sho will run a script to show the details so we can delete some of the files.

New Platform

Dave reported that the server he is setting up has the RUcore website, the directory structure, Fedora 3.2.1, and indexing has been run on the 160 objects from mss2. Searching appears to be working across these 160 objects. One object was added to the repository using WMS. There are still some remaining problems with LDAP. Dave will send out the IP address for the sanity test within the next few days. We will all test next week and hopefully declare the new platform "sane." Permissions still need to be worked out for the NFS shares. In general, at this point there doesn't appear to be any blocking problems. Regarding Fedora 3.4.2, Sho indicated that he will work during the next week to get this version of Fedora installed with the R5.2 software base. He expected that it will take about one week, assuming this work is his highest priority. Ron will verify priorities with Tibor. Sho will also send out a summary of what software development changes are needed in order to accommodate Fedora version 3.4.2. We also decided not to move to Fedora 3.5 – in this environment it is not a good idea to be an early adopter.

On a related matter, the PDF server has been unreliable and Dave has pointed to the development machine as a temporary fix. The recommendation is to reinstall the OS and upgrade the software on the production PDF server.

Release R5.2.2

What we previously designated at R5.2.1 now becomes R5.2.2 and includes WMS fixes and moving all objects to the /rarch directory. Kalaivani will clarify the bugs that should be included in this release and Yang will focus on bug fixing. Isaiah and Kalaivani will update the /rarch directory. Note that we also need to work the AV object specification (see below) which will also impact /rarch.

SOLR/Lucene Update

Chad has set up a two portals – one each for SOLR/Lucene and Amberfish. This will enable the user services community to do comparisons between these two search options. Rhonda indicated that there is sufficient representation in the group to carry out the testing and to come up with the xml schema required by SOLR/Lucene. The objective is to define the schema by the end of September.

Faculty Deposit Genre Specification

Everyone acknowledged that we should forward the updated FD genre specification to Cyber-Infrastructure for final approval.

Structure of AV Objects with Transcripts

We reviewed notes prepared by Ron on an approach for ingesting AV objects with transcripts. We established consensus on the following:

- Transcripts represent the same content as the media and should not be separate objects. This approach will also minimize user search confusion and simplify management of these objects.

- In the case of the audio project example, there would be three files that form the archival master: 1) the uncompressed audio – a .wav file, 2) the transcript in pdf/a format, and 3) the originally submitted transcript file – an MSWord document.

- Given the above, we agreed that that there should be multiple techMD sections to describe the different formats that represent the archival master. This capability already exists in WMS, however the linking mechanism between techMD and the associated datastream would have to be developed. (Note that others in the Fedora community are using rels-int for this purpose.

The discussion on this topic will continue in the next meeting. We will need to resolve how the archival master is structured (e.g. multiple archs) and how this will be represented in /rarch. In addition, we should discuss the linking mechanism to published articles that are deposited in a faculty collection.

Agenda Items for Next Meeting

- Report from Sho on software development changes for Fedora 3.4.2
- Continuation - Video/audio objects with transcripts
- Update of /rarch specification
- Getting started on the large file (GT 2GB) specification

- Review of the url migration implementation specification

rcj – 09/12/2011