

Software Architecture Working Group

July 18, 2008

Minutes of June 17, 2008 Meeting

Agenda

1. Brief Report on GDFR
2. Shibboleth update
3. FOXML ingest into Fedora 3.0
4. Darwin streaming server status
5. Ingest of videos into NJVid common – progress
6. Configuration/Installation at NJEDge
7. Review of R5.0 Schedule

Brief Report on Global Data Format Registry (GDFR)

Ron and Isaiah reported on the GDFR meeting in DC. The bottom line is that GDFR represents an essential technology for RUcore in order to be able to know what formats we have and to be able to migrate to new formats in the future. GDFR has a flexible architecture and data model and presumably will be available sometime in the Fall, 2008. We should consider a prototype/pilot in which we validate formats on ingest. Harvard University Libraries is managing the Mellon grant and will propose a pilot project in August, 2008.

Shibboleth Update

OIRT (Eric Marshall and Joe Lizzi) have set up a Shibboleth server (1.3) with LDAP. They are also working on setting up version 2.0. The current version will allow only authentication, i.e. no attribute passing. Sho will need to recompile Apache which will probably happen late July or early August although we need to get migration with content models in Fedora 3.0b2 working first.

FOXML Ingest

Jeffery has tested a foxml ingest with multiple sets of metadata using the rels-int for pointers between the various sections of metadata and file datastreams. Ingest works fine so the next step is to get WMS exporting this format. Yang has provided an initial test file but he has also indicated that there is more work to be done before the export is ready to go. We decided the following:

- For Open MIC WMS, pointers to reference multiple sections of metadata will use the METS ID mechanism.
- Our primary ingest format for RUcore will be foxml and multiple sections of metadata will be referenced using rels-int.
- WMS will also export METS however we will do not plan to use this format for ingest in R5.0.

Jeffery and Kalaivani will be testing the foxml export by using WMS to create the multiple sections of metadata and then testing to see if we can ingest the object.

Darwin Server Status

Isaiah and Sho are working to put the Darwin server on the public network. This will be done using a work-around while we await the arrival of another network switch (which has been ordered). Ron indicated that the NYU contact for shibbolized Darwin has not responded to several emails. The Shibboleth website reports the software as “under review” for release as open source. A contact at the GDFR meeting reported that shibbolized Darwin was actually working for NYU.

NJVid Commons

In Progress

- Jeffery and Isaiah tested streaming with a MOV-1 (a SMIL file) datastream on lefty and streaming on odin. This test worked and should serve as the model for R4.5 and R5.0 with Darwin.
- The NJVid Commons collection object is in mss3 (ingested by Mary Beth). The general consensus was that we will also need a sub-collection before we can begin ingest.
- Darwin on the network (being worked by Sho and Isaiah)

Video Object Architecture. The video object architecture will include datastreams as follows:

- The Flash FLV-1 (M-managed) datastream (no change here).
- MOV-1 (Managed) will be a file which contains the SMIL script.
- ARCH1 (Managed) will be a dummy file in R4.5 to avoid ingest of large archival masters.
- RARCH1 will be an “R” file which is a tar of the archival master – an AVI file. The RARCH1 will be added manually in R4.5 since this type is not yet supported by WMS.

In addition to the above, we discussed the possibility of including the video QT stream as a derivative in the tar file. The motivation for this is to preserve the QT since this stream will not be backed up on the Darwin server. We ultimately decided that the QT stream should not be tar-ed but rather will be added into the object as another managed stream. In R4.5, this will be done manually and WMS will need to handle the ingest in R5.0. (In a subsequent discussion, it was suggested that the datastream ID be “SMOV-1” for stream for MOV-1).

File System and Directories on mss for /rarch. As discussed previously for the R5.0 release, we will need a directory tree on mss3 to accommodate the archival masters. From previous discussion, we have decided that all video archival masters will use the “R” mode and will be located in a separate directory, the motivation is to avoid having to segment large files and to improve ingest times. It is proposed that this web-accessible directory be set up as /rarch with an estimated size of 1TB. It was

noted that the 1TB is not a fixed limit and the videos may consume more than the 1TB of space. There was much discussion about the naming conventions to be used for the /rarch subdirectories. This discussion will be continued in the next sw_arch meeting (July 31). A draft specification has been distributed to sw_arch.

R4.6 and R5.0 Schedule

We discussed the R4.6 schedule (Open WMS) and the need to finish up the work on Open-MIC WMS. Yang and Kalaivani felt that the Open-MIC changes were well defined and that the work could be finished by the time Yang goes on vacation in early August. However, there will not be much time in this interval to get Open WMS (R4.6) in shape for testing in August. In particular it was expected that foxml ingest, WMS edit, and related functions in the Faculty Deposit module still need work. We decided on the following steps:

- Open-MIC will take precedence and we will try to finish the work in July.
- Jeffery will work with Rhonda to create a WMS exported foxml file that has multiple sections of metadata. They will review the file and attempt to ingest it, reporting problems to Yang.
- Kalaivani and the test team will be testing Open WMS in August, however fixes will have to wait until September when Yang returns.

At this point, we believe that Open WMS is a critical path item for the release of R5.0 in September. The other critical path item is the release of Fedora 3.0. The last note from the Fedora architect indicated 3.0 would be ready for general availability by mid-July.

Read-only System

Dave reported progress on setting up a read-only system on mss2. Although there is work to be done (e.g. to point handles at mss2 objects), this work is showing great promise for providing service during planned or unintentional down times.

Pending Items for Next Meeting

- Review of file and directory naming conventions
- Portal Specification
- Configuration/Installation at NJEDge
- Shibboleth/Darwin/Fedora 3.0 status
- OJS/ETD specifications
- Filtering of DS IDs
- Faculty Deposit object architecture
- Annotation specification
- NJVid objects – license and report documents to be ingested with associated rights statements