

Software Architecture Working Group

June 27, 2008

Minutes of June 26, 2008 Meeting

Agenda

1. Finalize content model specification
2. Continue discussion of shibboleth/apache for R5.0
3. Discuss actions required for ingest of 30 NJVid videos
4. Use of NFS for video streams
5. Discussion of R5.0 schedule
6. OCR issues related to faculty deposit
7. Configuration/installation of RUCore for NJVid

Content Model Specification

The main points were reiterated and additions from the previous review were highlighted:

- All objects will have content models
- For the content model object ID, we will standardize on the following form: rutgers-lib:[objectArchitecture]-cmodel, e.g. rutgers-lib:Book-cmodel or rutgers-lib:Photograph-cmodel.
- Initially for R5.0, we will create content models for each existing object architecture. These architectures are listed in the specification on the RUCore developers' site.
- The specific FOXML rels-ext to be inserted in the object by WMS is listed in the specification.

Outstanding issues relate to Fedora 3.0, specifically the issue of how to validate against content models and the resource index does not build properly with Fedora 3.0b1. Ron will email Dan Davis regarding these issues. Once validation is working, we will need to standardize on what should be validated.

Related to validation, we agreed to explore a filter/upgrade of objects, many of which we know will not meet standards. In particular, some time ago we moved to functional IDs for each object datastream rather than use the generic DS1, DS2, etc. As a result, we have many datastreams that still use the DS convention. Jeffery and Dave will explore the possibility of doing an automatic update of these IDs. Jeffery and Kalaivani will work to put in place the required content models for R5.0. The content model specification with minor updates from this meeting is ready for posting on RUCore.

Shibboleth/Apache

From discussions with Chuck Hedrick and Eric Marshall, we have a commitment from them to put in place a shibboleth server by mid-July. We will also have to work with them to make sure that we have an LDAP connection that will allow us to do proper testing, i.e. we should be able to access attributes (e.g. roles) that are passed through shibboleth. Although we do not need shibboleth in R5.0, this work will provide the base for doing the required development in R5.1.

We will need to shibbolize Apache which will require a recompile. Sho indicated that he could have this done by the end of July. Ron indicated that he has finally established a contact for the Darwin-shibbolizing work that was done at NYU.

Actions required for ingest of 30 NJVid videos

A summary of these actions is included in Attachment 1 and was reviewed in the meeting. The key decisions are as follows:

- We concluded that using R4.5 (the current production version) would be the best alternative given some of the schedule risk in R5.0.
- The required workarounds are itemized in item 2 of the attachment.
- The primary issue was the use of NFS to enable the Darwin server to access video streams. We concluded that we will not use NFS because of some security issues and perhaps somewhat more complexity in the storage network. As a result the video stream will be located on the Darwin local store. This configuration may not work for NJVid however it was pointed out that we do not yet have a definition of the storage and server architecture for NJVid (a topic for the next meeting).
- To avoid software changes in R4.5, we will need to do a bit of manual editing of some of the video datastreams. For example, we would ingest a pseudo-archival master since all video masters will not be ingested. When we converge on R5.0, the ARCH1 datastream would have to be re-pointed to the directory on mss3 using the “R” mode.

Given limited storage space on mss3 and R5.0 implications of cpu requirements for checksum auditing, it is recommended that we limit the NJVid ingest to RUcore to the 30 objects and that these 30 be moved to the NJVid/NJEdge installation at the appropriate time.

Schedule of R5.0

There is concern about the WMS open source schedule. Our current schedule calls for delivery of WMS in July. Given continuing effort for the MIC version, there are concerns that we will not make this date. Although the major development is complete there remain some lingering issues with how to handle multiple sets of metadata in the FOXML format. Jeffery will use a WMS export of FOXML and add the rels-int segments and try an ingest into Fedora. If this works, the FOXML can serve as a specification for Yang to complete the export function. Our objective is to have a test-able version of WMS by August 2 (when Yang leaves on vacation).

Agenda for Next Meeting

Because Ron and Isaiah will be at a GDFR meeting in Washington on July 10, the next sw_arch meeting will be at 9:30 on Thursday, July 17. Agenda items include:

- Shibboleth. Server progress and possible presentation by Sujay.
- FOXML ingest. Proof of concept (Jeffery, Yang).
- Darwin streaming server (Isaiah, Jeffery)

- Faculty deposit content model and associated ocr issues.
- Configuration and installation for NJEdge.
- Review of R5.0 schedule
- Update on Fedora 3.0 issues and release schedule.

Attachment 1 – Ingest of NJVid Videos

Objective

- “Go live” in September, 2008 with at least 30 videos in NJVid Commons
- These videos will need to be in RUCore and tested several weeks before the annual NJEdge conference later in the Fall, 2008.

Summary/proposal of RUCore Actions

1. RUCore will host this initial collection. The collection may be replicated or moved to the NJEdge installation at some future date.
2. There will be a single NJVid collection object created at the same level as NJDH and RUCore.
3. This collection and search engine will be attached to the NJVid website for search and browse capability (similar to what we did for Virginia Tech)
4. The videos will all be publicly available so for this first collection, shibboleth and xacml will not be needed.
5. The videos will need to be streamed, i.e. no download.

Use of RUCore R4.5

The recommendation is to use R4.5 to ingest these videos however we will need to implement some “work arounds” given that some of the capability we need is scheduled in R5.0. The following is proposed:

1. Use WMS as is in R4.5 (i.e. no changes required). We need to verify this.
2. We can use the video object architecture as described for R4.5 with the following modifications/issues:
 - a. No changes to the FLV-1 (Flash) datastream, i.e. the Flash presentation file will be ingested under the “M” option.
 - b. For the archival masters, we will set up the external directory as planned for R5.0. Therefore we will not ingest large archival masters.
 - c. We will ingest a pseudo-object for the archival master. We will add, via dlr/EDIT, the RARCH datastream – a pointer to the archival master – as planned for R5.0. Note that this datastream will have no effect in R4.5.
 - d. We will ingest a pseudo-object for the MOV-1 (QuickTime) datastream. The file to be streamed will be located in another file system accessible to the Darwin server via NFS, per the R5.0 plan.
 - e. We will need to edit the MOV-1 datastream to point to or contain the script for playing the QT stream. This step requires some further investigation.
3. The steps outlined in item 2 above should be largely compatible with the R5.0 architecture.