Software Architecture Working Group   **February 18, 2008**

Minutes of February 14, 2008 Meeting

Agenda

- R4.5 Status
- Review of R4.6 Release
- Release 5.0 and 5.1 Review
- Disk Failure actions
- OJS 2.0 export and metadata
- Getting Started with Shibboleth and xacml
- Pending items
  - Statistics reporting for next release
  - Rels-ext and xacml experiments
  - Digital docs spec
  - Process for alert, especially, external users

R4.5 Release Status

Kalaivani reported that the faculty deposit service is in good shape for the release with a few minor changes regarding department name and language that need to be made. Given remaining testing on lefty and mss2, we set the release date for February 25.

Ron will plan to use lefty for demonstration purposes while at Northwestern University on 2/25 and 26. He will send out a demo alert on Friday (2/22).

R4.6 Release

The WMS open source capability with file handling (v1.5 of WMS) will be available as an initial release to Northwestern in mid to late March. We will do testing in March, however it is also important for NUL to do their own testing. Ron will emphasize this point in the late February visit. We also made it clear that we will deliver only the WMS code and indicate where other open source modules are needed (e.g. ImageMagick). NUL will have to download and configure other open source modules that are required for their own local environment. KA will provide a more detailed feature list for the final open source release scheduled for June/July, 2008.

R5.0 and R5.1

R5.0 is an architecture upgrade only. We decided last meeting that we will move forward to Fedora 3.0 which will provide the CMDA architecture. This upgrade will have an impact on WMS (e.g. when associating a disseminator with an object) and also on our disseminator architecture. Sho will work to get Fedora 3.0 installed on lefty64 in March.

We also need to upgrade our storage architecture to accommodate large files. This is important for NJVid and also for large science datasets. With some discussion, we decided on a storage architecture proposal for large files. Note that we will review this again in the sw_arch meeting of 2/21.

The two major advantages of this proposal for handling large files are a) there is no http ingest of the archival master leading to much faster ingest times and b) we no longer have to artificially break up the video archival master into 2GB segments. The single disadvantage is that the master will no longer be under Fedora management. With proper procedures, this should not be a problem.

The basic proposal includes the following major points and at this juncture is restricted to video files although it could be easily extended to other genres:

- The archival master datastream will be marked "E" for external meaning that it will not be managed by Fedora. See Figure 1 below, specifically the arch1 datastream. Other than this change, the object architecture remains the same as previously discussed.

- Pre-ingest manual procedures for the video will require locating the archival master and the two presentation files (QuickTime and Flash) on the Fedora server (either mss3 or the NJVid server).
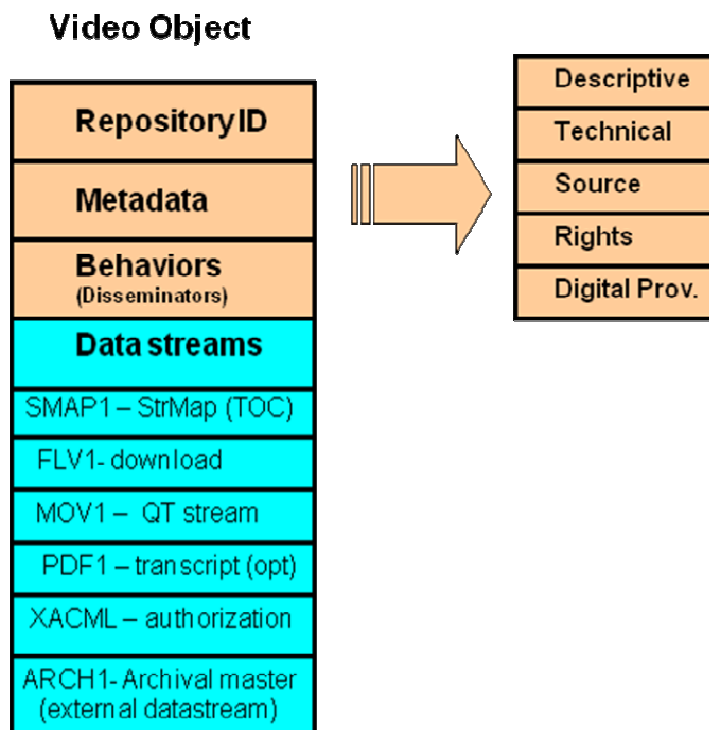
Video Object

Repository ID		Descriptive
Metadata		Technical
Behaviors (Disseminators)	⇒	Source
Data streams		Rights
SMAP1 – StrMap (TOC)		Digital Prov.
FLV1 – download		
MOV1 – QT stream		
PDF1 – transcript (opt)		
XACML – authorization		
ARCH1 – Archival master (external datastream)		

Figure 1 – Video Object Architecture

The storage architecture proposes a backend storage server with a file system labeled "external" where all the arch1 datastreams will be stored – see Figure 2 below. This configuration could also be set up transparently to house the external file system on the same server. We will set up our development

environment as two separate servers to accommodate the optimal configuration. The following points are made in reference to Figure 2 below:

- The backend storage server will house both the archival master and the checksums for the master (left side of Figure 2).

- As part of the preparation process, the original checksum will be generated off-line and will thus save further ingest time.

- Periodic verification of checksums will be done on the backend server using remote procedure calls from the Fedora server.

- WMS will need to be modified to handle the special procedures for video although these should be quite straightforward.
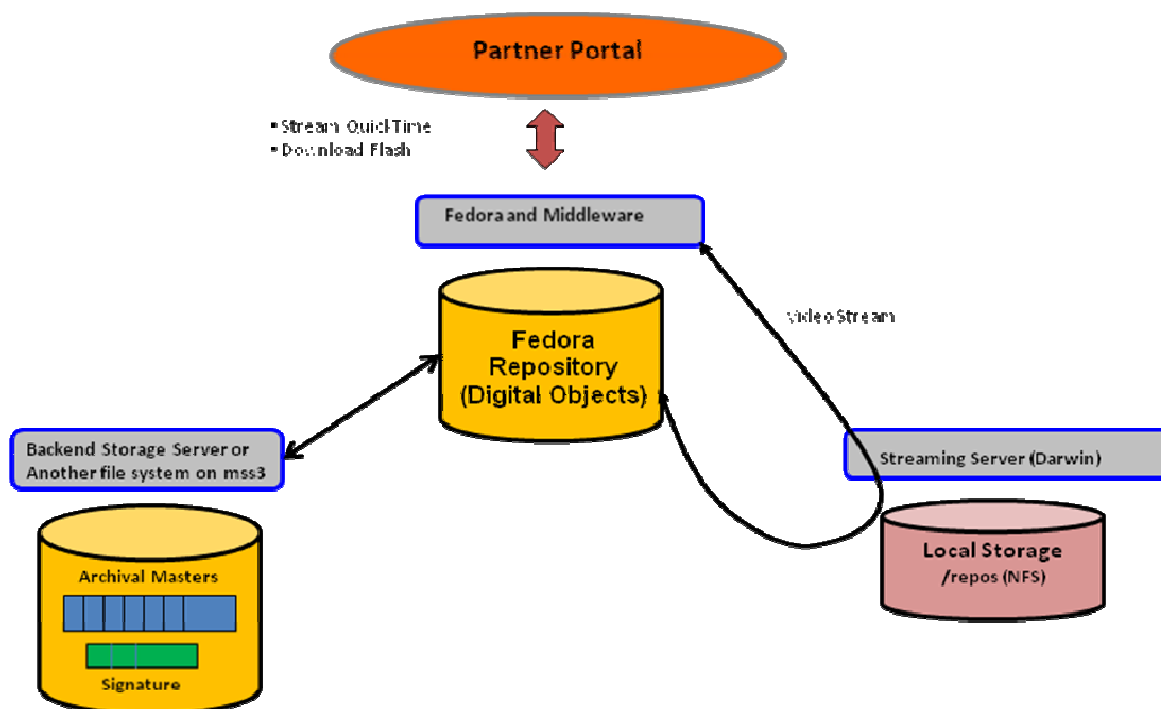

Figure 2 – Storage Architecture for Video

Other Items

For the shibboleth item, Ron indicated that he will forward an invitation to sw_arch for those who want to join an ad hoc shibboleth working group. Shibboleth and xacml implementation may be the most complicated part of the NJVid project. In our first meeting, we will try to get someone from Chuck Hedrick's organization to provide an update/tutorial on what they have done so far. (MAMS project)

We briefly discussed the results of a conference call regarding MetaArchive and a LOCKSS network. Isaiah, Jeffery, and Ron participated in this conference call with Katherine Skinner at MetaArchive. The approach to get more storage redundancy looks promising. Ron subsequently sent out notes and proposed that this be reviewed in the next C-I meeting.

We briefly discussed the request to submit NIH sponsored grant papers to PubMed Central. Rhonda and Ron are working this issue with Grace. Given the priority of this work, it may affect our release schedule.

The OJS and disk discussion were postponed until the next meeting scheduled for February 21 at 9:30 in the SCC Heyer room.