

Software Architecture Working Group

February 2, 2009

Minutes of January 29, 2009 Meeting**Agenda**

1. Quick updates and Announcements
2. R5.0 code complete status
3. Review of annotation spec
4. Review of NJEDL migration requirements
5. Compile/Config options (brief discussion).

Announcements and Quick Updates

- There is a meeting next week with FMG to better understand how we might annotate commercial videos.
- Sujay will be setting up meetings with specific members of sw_arch to review and understand our software. Areas include: shibboleth/Darwin, management (dlr/EDIT), custom portals, and configuration/compile options for R5.0
- Ron mentioned that our archive of development documents is becoming quite useful and urged everyone to make sure that documents are transferred to Chad for posting (i.e. after the documents have been baselined).

R5.0 Code Complete

Yang reported significant progress, indicating that java bridge was working and that he had been able to ingest foxml on lefty64. METS ingest is still not working; Yang will work with Jeffery to try a few quick experiments. If this capability is not working soon, we should plan to implement it in R5.1, i.e. it should not become a blockage for getting R5.0 out the door. The last major feature yet to get working is WMS edit. Yang will post to sw_arch the progress over the next few days. If WMS edit becomes a blocking issue, we may also want to push it out into R5.1

It was also noted that additional technical metadata changes dealing with format have surfaced in the last few days. After the changes have been identified, it will probably take Yang 3 to 4 days to implement. Grace has indicated that these changes are required and must go into openMIC and openWMS. Ron will email Grace to suggest an alternate strategy in which 1) the metadata changes do not go into R5.0 (openWMS) and 2) the changes are made to openMIC after the R5.0 release, e.g. in February, and 3) the metadata changes first appear in RUcore in release R5.1 (Spring/summer 2009).

Annotation Spec Review

Chad reviewed the updated assumptions for the annotation tool. The major issues relate to assumptions 1 and 9, specifically that the annotation tool will only annotate Flash videos. A meeting in the following

week is scheduled with FMG (commercial licensor or videos) to understand what they are doing with annotations. There is some opinion that they have been able to annotate a streaming video. We also discussed the current model of having “in progress” annotations and annotations with a state of either publish or unpublished. The discussion centered around the concern that a published annotation cannot be changed. The underlying concept is that a published annotation may be cited and cited material must not change in order to preserve citation integrity. Two possibilities were discussed: 1) if a published annotation must be modified, take a copy and create a new annotation or 2) if minor changes are required, make the changes to the annotation and make visible to the user (via display of the audit trail) that changes have occurred.

NJEDL Review

From early work by the NJEDL group (Langschieid, Colonna, Beard, Geng, Jantz) it appears that there will be no software impact on exporting NJEDL objects and ingesting into Fedora. Vince reviewed the specification for this work with the following noted from the discussion:

- A separate collection with a new collection name will be set up for ingest on NJEDL objects.
- As clarification, after the appropriate collection object has been established, all ingest of existing objects will occur through batch ingest.
- After all objects have been ingested, any new objects will be ingested via WMS.
- Batch ingest will follow all of the ingest (file naming) conventions used by WMS. (There is a potential clash with the unique WMS integer namespace. Jie pre-pended an integer with a “b” for batch).
- The proposed “delayed digital object ingest” may create some problems. The sub-group will re-examine this approach and suggest a process change for NJEDL.
- Because of the many pdf documents with text layers, NJEDL represents the largest index of any of our collections. A test on lefty64 will be run to make sure we do not have any problems with index build or index files size.
- To make sure all proper conventions have been followed, a test with WMS edit will be run on lefty64 to make sure that WMS can edit objects that have been batch ingested.
- For relationships, there is likely an already existing Fedora ontology that will work.
- The NJEDL portal website will be hosted in the SCC. This will need a bit more discussion. The existing map portal is hosted on the Windows server, in part because the Data Center website is also hosted on Windows. Future portals should probably be hosted on reaper64.
- WMS will need a controlled vocabulary for NJEDL keywords.

- After the CF site is decommissioned, we should archive the code and database. This should probably be archived with other CF sites on Windows. We may want to consider a more generic archiving process that creates an archive object in Fedora.

Other Items

- Release 5.0 code complete date was moved to January 23, 2009.
- Jeffery and Sho will provide to Dave the version numbers and compile options that are different from the previous release.
- Jeffery will work with Yang and Chad to produce a specification for how PDFs with a text layer are to be handled in R5.1. We uncovered during the course of this discussion that ETDs are apparently one large text file with no page markers. This lack of page markers renders full text searching useless.

Pending for Next Meeting

- Finalize ETD spec
- Review of the updated Annotation spec
- Compile/configuration options
- R5.1 pdf text layer with page markup
- Review of R5.1 schedule
- Global registry discussion
- Use of API/A
- Relationship services
- Shibboleth configuration for mss3

rcj – 02/02/2009