

**Minutes of March 18, 2010 Meeting**

**Agenda**

1. Updates and Announcements
2. Review of R5.1.1 progress
3. WMS/RUetd import/export
4. Batch ingest for NJEDL

**Announcements and Quick Updates**

Regarding demonstrations on lefty64, we decided that a "demo alert" email should be sent out several days before the event to the sw\_arch and steering committee lists. In addition, a reminder should also be sent out approximately 24 hours prior to the event. NJVid has indicated an interest in allowing harvesting of metadata. They were told that this feature is part of the Fedora platform and they can configure their installation to allow harvesting. Unrestricted OAI harvesting will provide metadata on all objects, some of which they probably don't want (e.g. disseminator objects). Those who harvest may have to filter these objects or export the use of the OAI "set" command as a possible filtering mechanism. NJVid is also proceeding with the LOD feature which will enable users to ingest to Fedora using the API (and presumably one of our ingest scripts). The rationale is that they want a "light-weight" ingest method with minimal metadata and do not want to use WMS for this application. It was noted that they will have to use WMS to set up collection objects. There are over 1200 ETDs in RUcore and these are also available from IRIS using the persistent ID. Rhonda noted that we may need permanent embargo for a very few ETDs. We should add "permanent embargo" as a new feature since there are likely other objects where we will need to do the same thing (e.g. OJS oral histories, certain data objects, etc.) We added another quality to initiative to look at legacy PDF files that might not have the xml datastream populated for full text searching.

**R5.1.1 Review**

The code complete state has been reached for R5.1.1 and Kalaivani will organize test teams to begin the testing. We should be able to do a public release in the mid-April time frame. The target for completing the testing is Friday, March 26. Regarding the checksum issue for videos in the rarch directory, Jeffery will edit the metadata for one of the videos, purge the old file, and verify that signature checking is working. If this works properly, we should schedule edits of the remaining video objects. Yang raised a question about how to determine if multiple files should be concatenated or handled as separate presentation files. We acknowledged that the user needs to tell us what he/she wants to happen, however the language for this can be confusing. Rhonda agreed to propose user-oriented language and then we will need to schedule the change in faculty deposit and WMS.

**WMS/RUetd Revisions**

Chad explained that there are some issues with file handling in RUetd and the interface to WMS including file naming conventions and files never being deleted from RUetd. He reviewed the three API methods (retrieve, delete, and manifest) that will be used to delete files after export has been completed. The draft specification was accepted and will be posted.

**NJEDL Ingest**

Vince reviewed the projected ingest statistics based on testing of two sets of 50 objects each on lefty64. For the approximately 5000 objects in NJEDL, we expect the size of the collection in RUCore to require about 300 GB additional storage (assuming the current collection is purged). The collection index will probably range up to 2.5GB. Vince reviewed the process for ingesting batches of 500 objects each, indicating that this will be done off-hours so as not to create load on mss3 during the prime shift. Vince will need to address and clarify three issues: a) how to handle thumbnails for maps and photographs, b) how to handle OCR for those objects that do not have already have ocr-ed text, and c) how to deal with the 2 GB index limit on mss3. Regarding the index limit, it was proposed that sub-collections may need to be created. (In a post-meeting discussion, it was thought that a sub-collection could be created and all objects go into this collection until the index approaches 2 GB. At this point, another sub-collection is created and the process is repeated. Vince will pursue this approach.) We have also set a target of completing the ingest in April.

#### **Other Items**

Completion of and verification of two important quality initiative (Google Scholar indexing and the signature script running properly), we may want to consider another team celebration.

#### **Agenda Items for the Next Meeting**

- Release R5.1.1 review
- Relationship services in R5.2
- Rights-event/document linking capability
- Review of R5.2 requirements status
- Faculty survey requirements
- Storage management on lefty64