

**Minutes of June 25, 2009 Meeting**

**Agenda**

1. Quick updates and Announcements
2. Migration and testing status
3. PDF text specification
4. Handling licensed documents
5. Review of R5.1 release

**Announcements and Quick Updates**

Per the NJVid conference call, we will soon need to begin the training and installation process for NJEDge.

**R5.0 Migration and Testing Status**

Kalaivani will need to do one more test with a fully compatible video object (including the archival master). Isaiah will set up the object for access. Kalaivani also outlined the remaining problems that need attention including 1) for one collection the type of resource is not recognized. This is a metadata problem and will be fixed after installation of R5.0, 2) we are encountering a bizarre problem in which the content model is not properly recognized in WMS edit. Some tests have been successful and others have failed. More testing is required, 3) the IE problem in which the url is too long will be solved in the next release, and 4) Jeffery will make changes in the libxslt script and test – to make sure we do not encounter problems with new versions, e.g. greater than 1.1.18. Regarding the migration progress, Sho will ingest all 15,000 plus objects from the box to mss3 starting on Thursday, June 25. Dave will continue to update mss3 (i.e. php, apache, mysql) over the weekend. At this point, the above problems do not seem to be blocking and we are still on target for public release of R5.0 sometime during the week of July 6. Dave also brought up an issue about what old objects should be kept. He currently has copies from our Fedora 1.2 and Fedora 2.1 installations. This issue is best brought up within the context of our digital preservation policy and should probably be considered by CISC.

**PDF Text Specification**

The problem is that not all text documents are being created with the xml1 datastream or alternatively the xml1 datastream doesn't contain the page markers that enable useful full text searching. The speculation at this meeting was that there are 3 cases that are not working properly as follows: 1) faculty submission of a pdf without a text layer – likely that no xml1 is created, 2) faculty submission of a Word document – text layer probably created however no page markers created and 3) ETDs in which the xml1 is created however the page markers are not created. Given the difference of opinion as to what has been implemented and what works, it was decided to hand this problem off to Jeffery and Yang for

further off-line investigation. Jeffery and Yang will report at the next sw\_arch meeting with specific actions for the 3 cases or note that they are in fact working in R5.0. The draft specification will be updated to reflect the actions to be taken.

### **Handling of License Documents**

Rhonda had earlier distributed notes on the actions she has taken to handle license documents for the Jazz Oral History collection. In reviewing these notes, we decided to define more generic requirements. There are two scenarios (at least) that we need to deal with as follows: 1) a unique license document for each resource and 2) a license document that covers many resources. In addition, there are at least four methods for one object to reference another one including: a) using the rutgers-lib identifier in metadata, b) using a datastream “E” reference, c) using rels with a specific semantic relationship such as “hasLicense” and d) using the PID. (An editorial comment: from an architectural point of view, we should use rels-ext to point to licensed documents however this approach requires further discussion). A related architectural issue is how relationships are established and managed in other contexts such as for annotations and ETDs. Finally, we need to be clear about who (what users) will want to view the licensed document. For limiting access, there are several options including creating a sub-collection that is not active or restricting access through xacml. Depending on the requirements for viewing, it is likely the presentation software will also need to change. Ron and Rhonda will collaborate further to pull together requirements for review at the next meeting.

### **Review of R5.1 Release**

Ron reviewed the R5.1 release summary and several changes were noted as follows:

- A “changeable content model” feature will be added to WMS, initially to support MARC ingest for videos. A requirements document is not needed for this feature.
- Two major features in WMS/FD have been committed and requirements are complete, i.e. 3<sup>rd</sup> party deposit and multiple authors (including multiple relatedItem xml segments). There was some concern that we need to check with MDWG to see if there are other critical enhancements that need to be in R5.1. Note that we are not soliciting enhancements at this point. Kalaivani will check with MDWG.
- The xacml utility will be used to create xacml for both licensed videos and ETD embargo. WMS should be updated to ingest this xacml, however a user interface for creating the xacml will not be implemented in R5.1.
- Two studies are required for possible implementation in R5.2: a) the “mss3” to “rucore” transition and b) can we simplify our configuration/installation process by moving perl code to PHP and not using java bridge? What are the implications?

It appears that the changes to R5.1 are not major and that we can still hold to the target “code complete” date of August 1, 2009.

**Pending for Next Meeting**

- Migration status
- Finalize PDF text specification with actions required in R5.1
- Review of license documents requirements
- Check on whether there are any new requirements coming in from MDWG
- Launch studies of “Perl to PHP”, java bridge and mss3 to rucore transition
- Discussion of how to handle training for NJEDge

rcj - -7/07/2009