

## **SW Arch Meeting Minutes – September 25, 2014**

### **Agenda**

- Announcements and Updates
- R7.5 Status
- More discussion
  - file uploading for MP4s - masters and presentation
  - parent-child relations for faculty collections (the extra level)
  - WMS 1:1 rule
  - passwords on PDF and cover sheet editing
- Virus scan memory investigation (Jeffery)
- NJDH spec for correcting collection structure (Kalaivani)
- Service pack update (Dave)
- Updating DOIs for legacy objects (Jeffery, Dave)
- Pending
  - unpacking tars for legacy objects
  - IP restrictions - scenarios

### **Announcements and Updates**

Chad mentioned that he conducted a class on use of the Analytic in one of the IHLs and multiple students were streaming video with no problem. This was a great stress test and we might want to consider doing other types of stress tests.

### **R7.5 Status**

We will need one more week to complete development for R7.5 so we reset the code complete date to October 7, 2014. We will retain the target for public release at November 17 and will plan to begin testing on October 17.

Regarding the cover sheet for PDFs, Jeffery will provide a capability in dlr/EDIT to add a cover sheet when a PDF is replaced (for R7.5). In addition, there was discussion about automatically populating metadata from an earlier article when a new version is submitted. We decided to place this feature in R7.6, in part because there remain subtleties regarding the reuse of the metadata (e.g. it is possible the title might have changed).

Kalaivani requested that everyone review their bugs and tasks in software libraries for R7.5. We will discuss any pending items in the next sw\_arch meeting.

### **PDF Cover Sheet Editing and PDF Restrictions**

In order to handle editing of PDF cover sheets, we agreed to do the following: 1) create a new datastream (REF-1) that would be used exclusively as a clean copy in order to regenerate the cover sheet and 2) WMS will create this new datastream upon creating the cover sheet for the first time. Note that replacement of a PDF in dlr/EDIT will also require the replacement of the REF-1 datastream. This WMS feature will be added to R7.5. Jeffery will examine the possibility of automatically adding the REF-1 datastream to the ~300 existing articles in RUcore.

We have observed that it is possible to block the modification of a PDF in various ways – an issue for adding a cover sheet and possibly for future migration. Isaiah presented several scenarios on how a PDF might be restricted. We agreed to add a capability in R7.6 that would detect this situation and send a note to the depositor indicating that the restriction has to be removed for deposit into RUcore (e.g. for preservation reasons). Before deciding about how to do this, we need to examine existing ETDs and articles to determine the extent and how the restrictions have been implemented (e.g. via encrypting the document). Finally, we will review the PDF restriction issue in CISC for final determination of policy and priority.

### **File Uploading and WMS 1:1 policy**

Isaiah reviewed the various complex objects that can create uploading confusion and don't adhere to the WMS 1:1 relationship between archival master and the presentation file. He proposed different courses of action based on the content model. For example, we can likely continue with the 1:1 policy for maps and still images. For books, ETDs, research datasets and others, there should be no expectation of a 1:1 relationship. Two approaches were discussed: 1) WMS can maintain the 1:1 relationship when there is a dmaster available and 2) WMS should allow more flexibility when there are multiple masters by allowing the user to identify one or more presentation files, however the default will remain 1:1. This capability will be targeted for R7.6 and will require a specification.

### **NJDH Old and New Collections**

Kalaivani reviewed the specification for how to correct the collection identifiers for NJDH. The NJDH collection has an old style collection ID which leads to confusion when creating new sub-collections. In particular, one has to edit the new style collection ID in order for the sub-collection to be properly recognized as part of NJDH. With an appropriate script, we will create a new NJDH collection and edit the some 28 existing sub-collections to use NJDH rather than the new style ID "rucore00...1015". This work can be done immediately and is not dependent on a software release.

### **Virus Scan Memory Investigation**

Jeffery explained the memory utilization issue in how WMS and dlr/EDIT were handling virus scanning. Basically, file access via a URL socket requires more memory. As a result, the Apache

memory increase to 1 GB will remain as a permanent fix. Jeffery will investigate accessing files from tmp upload in addition to the current file upload and url socket.

### **Service Pack Update**

Dave will start the upgrade to service pack 2.0 on staging next week – October 3 or 4.

### **DOI Update**

We will proceed with the DOI update process as soon as the service pack upgrade is finished.

### **Other Items**

Yang raised the issue of not having identifiers for certain schools. This issue will not likely be resolved quickly given the University is still integrating with UMDNJ. As a result, we will move ahead with a temporary ID scheme by assigning IDs as 99A, 99B, etc.

Chad mentioned that we must be aware (for testing) of the fact that legacy digiprovs do not have IDs and as a result email addresses may not be valid.

### **Agenda for Next Meeting (October 9)**

- Review of items in software libraries for R7.5
- Unpacking archival master tar files
- Adding the REF-1 datastream to existing objects (script or manual?)
- Process and policy for handling crons.
- Move NJDH to one main collection
- Pending
  - Scenarios for IP restrictions
  - Content model for dark archive

rcj- 10/06/2014