

## **SW Arch Meeting Minutes – May 17, 2012**

### **Agenda**

- Announcements and Updates
- Progress on Non-release projects
- Fedora 3.5 on staging
- Feedback from Data Group on the UI
- Multiple techMDs – continuation
- Background ingest and alerting
- Release summary discussion

### **Announcements and Updates**

Our next meeting is on May 31. Because of vacations, our first June meeting will be on June 21, thereafter we will get back on the two week schedule. We have been notified that the Jazz Oral Histories interviews can be ingested and made accessible to the general public. We're waiting to get the OK to proceed with ingest; there are some 80 records in WMS that were prepared some time ago. Rhonda indicated that we are updating rights statements for various collections (e.g. Voorhees and yearbooks), per Janice's instructions. We have concluded that these types of activities are going to be routine, requiring us to run various scripts against the production collection. We are encountering different situations in which we want to repair various objects given new information or to correct something we did years ago. We will review the process for identifying and running these scripts in the next sw\_arch meeting.

### **Non-Release Related Updates**

The status of each project is summarized below:

- JPE. No update here. We're still working the "last" remaining bug.
- XML-1 Datastreams and Jpeg Thumbnails. Given vacations, we decided to postpone the running of the required scripts on production until early June.

### **Fedora 3.5 on Staging**

The target is to install Fedora 3.5 on staging next week. We are hoping to get some testing of R6.1.2 completed before we begin to encounter vacations in June.

### **Data Group Review of the User Interface for Complex Objects**

Aletia reviewed the results of the Data Group review. We reconfirmed the decision that we will always retain the filename (in the label attribute of the datastream). In the typical simple object scenario, the original filename may not be important, however for complex objects and multiple level directories, the filenames becomes important. We discussed how to determine when a structural map for a complex object is created. We concluded that the trigger for WMS to create a structural map would be the presence of a directory for upload. For example, one could conceive of a book having many tiff

images and also having an audio and video file. In this case, the user might present a directory with several subdirectories (e.g. the tiff directory, a video directory and an audio directory). In this case, WMS would create a structural map. This new capability raises questions about workflow and the user interface, both in WMS and for presentation to the user. For example, in WMS, how does one indicate that there is a directory in which tiffs should be converted to a pdf for presentation. In the end user interface, we now would see three datastreams for the book: the pdf, the audio, and the video. In addition, given the structural map, we may want to offer the user the capability to zip and download the complete object. These user interface issues were postponed until Yang and Chad get further into the implementation. It should also be noted, especially for data projects, that relationships (rels-ext) for complex objects will have to be created manually in dlr/EDIT for this release.

### **Multiple techMDs**

We had a very good discussion of the rather complex process for handling multiple techMDs. We focused on how to handle objects that will be ingested in the next release (R7.0), acknowledging that we will at some point have to address how to update previously ingested objects. The first part of our discussion focused on how to handle the wealth of metadata coming from the exif and media metadata extraction tools. The MDWG is meeting on May 21 to discuss the issue. Since it will take some time to work through the various issues of what technical metadata to extract for the different content models, we discussed how to proceed with architecture while the decision process is unfolding. For RUcore R7.0, we decided on the following approach, acknowledging that we would want to enhance the capability after the metadata decisions are approved.

- We have previously agreed that each archival master file will include mime type, checksum, and size as attributes on the datastream. Thus, these items will no longer appear in the techMD.
- We concluded that adding another datastream for each archival file that included all of the extracted metadata (the sidecar approach) would add a lot of datastreams that would be little used. We decided not to proceed with this option.
- As a result, for R7.0, we will only use the metadata extraction tools for file validation and will incorporate the MDWG decisions about which metadata to be used in the following release, R7.5.

We then proceeded to the discussion of multiple techMDs. For simple objects (e.g. a book with many tiffs), we decided that we will have one techMD in R7.0. Thus, in this release, we are temporarily ignoring the more complex problem in which the tiffs have been created on separate platforms and thus have different metadata regarding the creation process. It was pointed out that the result may be an “empty” techMD, the only data item being the content model. We concluded that this is not a problem

since users may want to come back later and add some technical metadata. For complex objects that have many different file types, WMS will create multiple techMDs. These techMDs will be linked to their respective datastreams by using the Fedora rels-int capability. The major discussion here focused on whether each techMD is a separate datastream or whether the techMDs become blocks of xml code within a single datastream. In each case there are user interface issues. For example, if one wants to locate the techMD for page 15 in a book and add to the metadata (since this page is in color), the user will have to scan through text or select multiple datastreams in order to locate the correct file. In the end, we decided to represent each techMD in a separate datastream, largely because this approach can more easily use the native Fedora rels-int without us having to develop our own custom linking mechanism.

### **Other Items**

Ron indicated that we will review the pending releases in the next meeting. Basically, we have three releases in the queue for 2012: R6.5 (RUcore R6.1.2 on Fedora 3.5), R7.0 (large objects, complex objects, and updates of faculty deposit), R7.5 (more faculty deposit, video updates). We also postponed the discussion of background ingest and alerting for next meeting.

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

- Status of Non-release projects
- Background ingest and alerting – Kalaivani
- Faculty deposit items to be included in R7.0 – Rhonda, Jie
- Discussion – process for running of various scripts on production
- Review of Pending specifications and releases - Ron
- Pending agenda items
  - RUetd – WMS update
  - Enhanced UI for the landing page (a possible framework) – Chad and Jeffery
  - WMS – creation of relationships (rels-ext) for data projects