

## **SW Arch Meeting Minutes – April 11, 2013**

### **Agenda**

- Announcements and Updates
- Metadata for DOIs
- R7.0.1 Issues and R7.1 (Fedora Upgrade to 3.6.2)
- Virus scan update
- Faculty Deposit Features for R7.2

### **Announcements and Updates**

Rhonda reported that FD alerting does not appear to be work. (In a follow-up email, it appears that the problem cannot be reproduced). Ron indicated that his use of openETD was successful – the user interface is very intuitive.

### **Metadata for DOIs**

In structuring metadata for DOIs, there appears to be problems using the DC profile where there is no support for repeated elements. For example, the only way to represent multiple authors is by using a single field with comma-delimited author names. The consensus was that we should move to using the DataCite profile. Rhonda and Kalaivani will discuss the issue in MDWG next Monday to get a final decision on how to proceed.

### **R7.0.1 Issues and R7.1**

We discussed two issues related to the R7.0.1 release. There appear to be problems in handling certain special character encodings which is thought to be related to different versions of the pdftotext software. The xpdf version is on staging and production, whereas the Poplar version is on development. There is also a problem in creating jpeg thumbnails for ETDs, likely related to ghostscript. Dave notes that using the binary on development for the 9.0.6 ghostscript upgrade appears to solve the problem. We decided to take the following actions: 1) As a first step, Dave will upgrade staging and production with the Poplar version of pdftotext, 2) Secondly, for the jpeg issue, Dave will recompile the code, rather than using the binaries, and we will test on development. If this works, we can then move the ghostscript upgrade to staging and production. In the meantime, Chad and Kalaivani will manually repair the ETDs that do not have thumbnails.

Following the above actions, Dave thought that he would be able to install Fedora 3.6.2 on staging within a week or so after the soak period for R7.0.1 is complete (i.e. April15). We'll then do some quick sanity tests with R7.0.1 on Fedora 3.6.2. Although we expect some issues, if all the basic functions are working, we will then move Fedora 3.6.2 to development. This upgrade will also include an upgrade of PHP and to service pack 2.0.

### **Virus Scanning Update**

Isaiah did clamAV time tests on various machines showing a range from 3 to 6 seconds for a 100 MB file. He also noted that clamAV has a 4 GB maximum limit even on 64 bit architectures. Given these tests and minor limitations, we decided to move forward with Isaiah's recommendations: 1) scanning three file systems once or twice yearly. The mountpoints for these scans are /workarea, /rarch, and /repository and 2) we will implement upload scans in R7.4 for WMS, FD, dlr/EDIT. We should also follow through with upload scans for openETD. For implementation, we will need to devise a notification strategy when a virus is detected. This notification becomes somewhat more complex since WMS uploads are now in the background.

### **Faculty Deposit Features for R7.2**

As discussed in the previous meeting, we will add the capability to upload html files in R7.2. In addition, Rhonda requested that we provide for the ordering of author names. We will target this feature for R7.2. If it doesn't make the code freeze date (May 1), we'll include this feature in R7.3.

### **Other Items**

Given our recent experience with differing versions of pdftotext and ghostscript, we concluded that a review and update of our configuration document is necessary. We'll look to Jeffery (as the editor of the document) to bring this forward in a subsequent meeting.

### **Agenda Items for the Next Meeting (April 25)**

- Content models – next steps - WMS configuration, updating FD, stats, techMD ( Rhonda)
- Open discussion on how to move forward with rels-ext. Where are we currently using rels-ext (data projects, analytics)? Proposals for new uses (e.g. administrative documents). User friendly labeling of ontologies and filtering for private use.
- Review and update of the software configuration document
- Further discussion on when to schedule sha-256 updates for legacy objects
- Begin discussion for R7.3 – User accounts