

SW Arch Meeting Minutes – February 28, 2013

Agenda

- Announcements and Updates
- Release summary and upcoming dot releases
- Status of R7.0 testing
- Review of Content Models
- Specification for faculty deposit of html files
- SOLR dynamic indexing
- RUetd-WMS update

Announcements and Updates

Ron raised several issues that will require our attention. First, there are still issues with some PDF files being created without a text layer (e.g. in NJEDL). We're not sure if this is a software or a user error and will require more investigation. Linda will enter a task in software libraries assigned to Jeffery. There was a consensus that we need to begin virus checking, especially for faculty deposit and ETDs. The most expeditious approach would be to initiate directory/file level scanning independent of any particular software capability. We may as well want to provide further capability in WMS. CLAMav is an open source antivirus that can detect Trojans, malware, etc that we could use. Isaiah will bring a proposal for virus scanning to the next sw_arch meeting. The third item involved our pipeline OCR process which has had limits based on file size and number of pages. With the introduction of WMS background processes in R7.0, we should have more time to do OCR of large documents. Jeffery will investigate with Yang to determine what needs to be done. On other issues, Isaiah reported that he has acceptable work-arounds to ingest the Syriac materials. Dave indicated that we need to discuss how we want to handle service pack upgrades, Iselon firmware, and version upgrades (on the agenda for the next meeting).

Release Summary and Dot Releases

We revised release numbering, identified target dates, and added another dot release to the release summary (the revised summary will be sent out separately). Briefly, R7.1 is the upgrade to Fedora 3.6.2. Dave will install Fedora 3.6.2 on staging after two weeks of soak time for R7.0. After sanity testing of R7.0 with 3.6.2, we will move Fedora 3.6.2 to development. The proposed release date of R7.1 (i.e. 7.0 with Fedora 3.6.2) is April 22. R7.2 focuses primarily on upgrading to the use of DOIs and we decided to add a few faculty deposit features (TBD). The code freeze date for R7.2 is May 1. A new dot release (R7.3) will include the jpage2000 page turner, Wowza with MP4, the metadata extraction tools (exif and media), and the implementation of facets. The proposed code freeze date for R7.2 is August 2. R7.5 includes major new capabilities including authentication/authorization, versioning and user accounts.

Status of R7.0 Testing

Kalaivani distributed a test document for testing of R7.0 production on the morning of Monday, March 4. Testers should be at TSB at 8 am on Monday with certain configuration work to be completed before 8 am. Our objective is to have the system ready to go by 10 am on Monday.

Review of Content Models

Rhonda reviewed the proposed changes to our content models, notably the addition of a 3D object, EAD, and Analytic models and the changing of manuscript to document, video to moving image, pamphlet to book, and photograph to still image. In addition to WMS configuration changes for the new models, there will need to be code changes in faculty deposit and statistics. Also, we will need to run a script to update content models in previously ingested objects. Rhonda will bring in a proposal in the next meeting as to how we should proceed with these items.

Specification for Faculty Deposit of HTML Files

Jie reviewed the specification for providing an HTML upload capability in faculty deposit. We agreed to provide the capability to upload one page which will be referenced via a URL (only one URL will be permitted). The one page will be encapsulated in a PDF as the archival master. We will plan to implement this capability in R7.2

SOLR Dynamic Indexing

Jeffery reviewed the specification for dynamic SOLR indexing. This capability will allow us to easily add new fields to be indexed without restarting SOLR and delivering code changes as part of a major RUcore release. By using the “*_txt” capability, we can avoid making changes to the schema.xml file and also avoid restarting the SOLR server.

A capability will also be provided in dlr/EDIT that will allow authorized users to create or delete dynamic SOLR fields. Jeffery will explore the portal framework to implement this capability. We will also convert as many of the existing fields to dynamic as possible. When these changes are made, we should be able to avoid code changes when a new field is added and thereby enable administrators to add fields as needed without waiting for a major RUcore upgrade. Jeffery will also check to see if there are any performance related issues with these changes. This capability will be implemented as part of our R7.2 release.

RUetd-WMS Update

Chad briefly reviewed the specification for improving the import/export interaction between RUetd and WMS. A series of RESTful web services will be implemented to remove the need for RPC code and delete ETDs after export (see the specification for more detail). This capability will be included as part of R7.5.

Other Items

Regarding our specification review process, we agreed to try to distribute draft documents to sw_arch at least 24 hours before the meeting in which the document will be discussed. This will help us review the

document more quickly, given that time is limited in the meeting. Ron will forward the proposed release summary to Grace for comments and possible review in CISC.

Agenda Items for the Next Meeting

- Service pack, firmware, and version upgrades (Dave)
- Virus scanning proposal (Isaiah)
- DOIs at the file level and versioning (Ron)
- Content models – next steps - WMS configuration, updating FD, stats, and techMD (Rhonda)
- FD features to be added to R7.2 (Rhonda)

rcj – 03/11/2013