

SW Arch Meeting Minutes – July 11, 2013

Agenda

- Announcements and Updates
- Status of R7.2
- Update on unpacking archival tars and sha-256
- Yearbooks – using djvu-to-jpeg
- Content Models – next steps

Announcements and Updates

Ron reported on topics covered in the CISC meeting. He has developed an RUcore statistics report by collection and by total RUcore. We expect to see well over one million downloads in RUcore for the calendar year 2013. We have preliminary approval to proceed with digitizing historic dissertations (i.e. prior to 2007) and should be able to start this work in the Fall, 2013. Several sw_arch members will participate next week in the webscale evaluation which provides a single search box for searching our OPAC, electronic materials and RUcore. It was noted that we will soon have to deal with ETDs in which some, or possibly all, of the ETD supplementary files will need to be made public, possibly after a suitable embargo period.

Status of R7.2

Dave is making progress on installing R7.2 on staging although it appears that there are still some configuration issues. Assuming we can resolve these issues quickly, the roll-out schedule is as follows: 1) finish testing on staging during the week of July 15, 2) install R7.2 on production on July 24, and 3) send out a public announcement of the new release and possible downtime on July 16 or 17, depending in part on the status of testing on staging.

Unpacking tar files and sha-256 Update of Legacy Objects

Jeffery reviewed the progress with unpacking tar files and assigning sha-256 checksums to legacy objects. He has done significant testing on the development server using several different collections. For each object, the end result is the creation of separate archival masters for each file encapsulated in the tar file, the creation of appropriate technical metadata for each archival master, creation of the SMAP for the structure of archival master, and the creation of the sha-256 checksum for each master. Note that Jeffery is temporarily retaining the original archival master as a precaution. For these legacy objects, we have to retain the inline representation of the technical metadata (as opposed to the managed “M” representation). We concluded that our layer of software that accesses the technical metadata will work fine with both representations as long as we use the Fedora REST API. At some later date, we may want to explore the possibility of moving all technical metadata to managed format.

Jeffery will bring a mini-spec into the next meeting that outlines the procedure on production for doing the update by content model. We will start with the “Photograph” model and proceed to work through all of the content models. It was noted that the update process will have to deal with videos that have both rarch files (with a dummy ARCH1) and with videos that have legitimate ARCH1 files. We need to also verify that all of our display software deals appropriately with handling many blocks of technical metadata (i.e. one for each archival master).

Using the djvu-to-jpeg approach for Yearbooks

We discussed using the djvu-to-jpeg software for the Yearbook project. This approach is much more functional and reliable than our original proposal to use the java applet (e.g. the djvu-to-jpeg works on iPads). We need to have the capability working in August and have proposed a dot release (R7.2.1) for late August.

Rhonda has suggested a demo in CISC on August 14 and we will need to do a bit of preparatory work. Jeffery's software appears to be working and he will move several, good quality yearbooks from production to the development server. Chad will need to setup a Yearbook test portal on development and reference the djvu-to-jpeg capability.

Content Models – Next Steps

Rhonda reviewed the steps we need to take in R7.3 to add three new content models (3D object, Analytic, and EAD) and discontinue four content models (manuscript, record, transcript, and pamphlet). The list of Fedora content models will also need to be updated appropriately. There will need to be changes in WMS, dlr/EDIT, and the statistics software. It should be noted that our software will need to continue to work with the discontinued models since these models are all represented in the metadata of legacy objects. The plan is to map these discontinued models to the appropriate model in R7.5. See Rhonda's specification for the detailed actions to be taken in R7.3 and R7.5.

User Accounts

Rhonda handed out a preliminary specification and asked everyone to review and provide comments. We will review the user accounts specification in the July 25 meeting.

Agenda Items for the Next Meeting (July 25)

- User account specification - Rhonda
- Mini-spec for sha-256 update for photographs
- OCR investigation (Isaiah)
- Review of release content for R7.3, 7.4