Software Architecture Working Group
Minutes, August 13, 2015

Attending: Rhonda Marker, Jie Geng, Kalaivani Ananthan, Chad Mills, Yang Yu, Aletia Morgan, Jeffery Triggs (recorder)

1) Announcements - All
* CISC yesterday. Isaiah presented on supported file formats. He noted that Microsoft would have diminishing support for old MS Office file types. We need to consider migrating legacy datastreams from doc to docx, etc. We'll begin by running a report on the old formats in the repository. This will mainly turn up candidates for .xsl to .xslx, though also for all older MS Office files. We won't be keeping old versions of the docs and other old formats.
* There will be a report of sites that need to be converted to use CAS. It would affect RUETD and other sites. This does not need to be tied to a release schedule. The WMS (like other apps that need non-Rutgers users) may be an exception we could ask for down the line. We need clarification from OIT.
* OOI was discussed at CISC.
* We hope to test on staging on Tuesday.
* There has been a delay in negotiating getting LDAP information about graduate students from OIT, so we will hold back on authenticating graduate students for 7.7. or a possible 7.6.2 dot release.

2) Update on Reports & Jobs - Triggs
* Migration tests for inline to managed content went well on rep-dev. The only datastreams we need to migrate at first will be TECHNICAL1, though all datastreams except the audit trail were shown to work in Fedora as managed content. (The audit trail is a special datastream in Fedora.) We can't move descriptive metadata datastreams with the current RUcore presentation landing pages. We will look into whether we can run the script without removing all POLICY restrictions.
* Dave ran the getprepdf script and added the new REF-1 datastreams on the production server.
* He also added the new MODS datastreams with dates for the Tombstone collection. All that remains for this is to run the rundoidopublic.sh script on these objects.
* We will review our issues in software.libraries to update as needed.

3) Update on 7.7 release - Mills
Chad showed the latest 7.7 list to CISC. The file counting report has been moved to a later date. ABBYY OCR/PDF and Google Video sitemaps have been added to 7.7. We need to include Nick in the ABBYY process. Among other things he will have to document how the server is set up.

4) Additional SOAR items in 7.7 - Marker

Rhonda distributed and described two specs that she and Jane Otto wrote on how to handle articles and book chapters in the SOAR submission form. The SOAR group conducted a usability study that found that the scholarly users were confused by the current vocabulary. This led to the first two forms for different types of work. The initial choice of "type" leads to different sets of questions. These have been made clearer in certain instances based on user suggestions. There was some discussion about whether or not the list of identifier types (for things like PMCID) should be configurable or not for this release. Once the type has been selected, all the options for a type should be displayed at once. Book chapters will include certain additional material, such as separate Book and Book Chapter titles and book editor information. What to do with place of publication for a book chapter that has not yet been published need to be considered. She will be presenting similar papers on the other types of work. Each type will need its own form. After this there should not be any more major feature changes for SOAR.

5) Workflow review of ingest process in WMS - Yu
Yang created workflow charts for the ingest processes. These include pre-ingest, ingest, and post-ingest processing. We may want to prevent purging an object simply because it failed the initial Solr index. This would allow us to examine what went wrong and perhaps fix it. The Solr indexer might do some tests that would allow a warning to be sent back to WMS and still allow indexing. We might also change the order of some WMS processes so that ingested objects do not need to be purged so often. If ingest works and WMS is able to update its status, perhaps no purge needs to take place. This would make failures of indexing or updating the stats table non-fatal rather than fatal errors requiting a purge of the object. The update WMS process should be moved up to right after ingest and collection table perhaps. Alert author should be moved to the end so that authors will not receive messages about ingested objects that have to be purged. Decisions: We will no longer purge for post-ingest processes. We will make external warnings more robust. Yang will update his document and send it to the group.

6) Review update of 7.7 JPEG2k specification - Beard, Mills
JRA Publish will no longer be a JPEG2000 creation option. Chad and Isaiah will look into three new open source JPEG2000 server options. Chad will post the new version of the spec as an update to the previous version.

7) Ocean Observatories Initiative/RUresearch project integration - Mills
(postponed)