

Attending: Ananthan, Beard, Geng, Hoover, Jantz, Marker, Mills(recorder), Triggs, Yu

Agenda

- 1) Announcements - all
- 2) 8.2 status, deprecation of 'rulinkrutgersedu' LDAP attributes – all
- 3) Large(quantity/size) files and ABBYY update - all
- 4) Jobs & Reports - Triggs

Announcements

- 1) VDC grant has begun; Jantz is the groups contact for the grant.
- 2) Weekly Agenda has a piece highlighting the Roman Coins projects and that projects use of high resolution image capability in RUcore.

<https://blogs.libraries.rutgers.edu/agenda/200-coins-added-to-the-roman-coins-project-over-the-summer/>

8.2 Status & Deprecation of 'rulinkrutgersedu' LDAP attributes

1) 8.2 Status

All packages will be ready for testing in the test environment by Monday, October 10th. The RUcore SSO still needs some commits from Geng.

The EAD display is partially ready. Geng and Mills will be meeting with Radick tomorrow to finalize the implementation. After that final code changes will be put in place, developer tested, and prepared for the testing group. The expectation is that will be also ready by Monday, October 10th.

Triggs and Yu will make changes to any LDAP attribute dependencies in WMS and SSO that use the rulink* prefix. Mills will create JIRA issues and tie testing methods to these issues for this release. While this is a "last minute" change if this is included now RUcore will easily meet the deadline set by OIT of April 2017 to move away from rulink* attributes. The expectation is that will be also ready by Monday, October 10th.

A placeholder link to WMS File Handling Best Practices will be added to the WMS help menu. This link will direct the users to a PDF document stored under the RUcore Reference Materials section. The expectation is that will be also ready by Monday, October 10th.

Yu has been in communication with KeyMark and ABBYY engineers regarding support for Office document to PDF creation using ABBYY Open API feature. Yu was asked by ABBYY engineers to provide a set of sample code that we are using to interact with the Open API. Yu provided that yesterday and is now waiting for a response on this issue. Using ABBYY for Office document to PDF generation will not make the 8.2 release.

Yu reported that working with a sample ingest of 8,000+ files/datastreams has encountered more issues that are mainly centered around using DOM to handle XML processing and generation. Initially, Yu found issues with creating a structural map using DOM and 8,000+ files. Switching to using XMLReader solved the issue. Now other XML processing using DOM is creating a similar issue. For now, we will not be able to support resources with thousands of datastreams in RUcore. For the Kopp data project, it was suggested that Jantz and Beard consider dividing the files up into smaller(+100GB) tar.gz files. We can then offer multiple datastreams as part of a single resource. This will not provide

the granular downloading we had hoped for, but it will make the data available sooner while we explore supporting thousands of datastreams per resource. For 8.2 structural map generation will move forward using XMLReader. Beard and Triggs will test this capability using an 800 page Syriac document that has proved troublesome in the past. This will be ready for testing by Monday, October 10th.

Jobs & Reports

- 1) THUMBJPEG Job – Hoover reported the job was run with commit mode enabled on the production system earlier in the week. He will provide the report when time permits. Mills needs to now make additional configuration changes now that the larger thumbnails are in place.
- 2) High resolution image job – Hoover discovered that 2 copies of every PTIF were created during the job. Mills looked into the script that was used to run the job and found that calling the Fedora “validateChecksum” method after initial ingest triggered a copy to be made. At this point Mills thinks there is something wrong with the job script and will look into it further. At the very least a cleanup job will need to be written; tested, and run to remove the duplicate copy.
- 3) UnTARing legacy resources – Triggs is going to start concentrating on generating technical metadata for the untarred archival datastream contents. More will be reported at the next meeting.
- 4) Mills asked that PDF presentation files be created for a handful of Motion Pictures Catalog resources that have DjVu only presentation datastreams. Once completed there shouldn't be any resources that only have DjVu presentation datastreams. This means deprecation and subsequent deletion of DjVu datastreams can begin.

<https://software.libraries.rutgers.edu/node/3576>