

SW Arch Meeting Minutes – April 25, 2013

Agenda

- Announcements and Updates
- Content models (postponed to next meeting)
- Progress on system updates - pdftotext, ghostscript, etc
- Discussion on the use of Fedora relationships
- Updating the RUcore software version document

Announcements and Updates

The MDWG will issue a document shortly on metadata issues related to the DOI implementation, specifically recommendations on what profile to use and how to handle special cases (e.g. when one of the required fields has no values). Aletia noted that 11 videos have been ingested for the NSF grant on teaching physics, officially titled as "Content Knowledge for Teaching Energy Project (CKT-E)". The videos are not publicly accessible and video access is authenticated through the Analytic.

Progress on System Updates

Dave reported progress on several system updates. The poppler version 12.3 of pdftotext has been installed on all servers. This will solve some of the anomalies were seeing with special characters. A new version (9.0.7) of Ghostscript will be installed on the development server. We'll need to do a quick test to verify that the thumbnail creation problem with ETDs is fixed with this upgrade. Fedora 3.6.2 has been installed on the staging server. It appears that most of the presentation functions are working however WMS ingest and dlr/EDIT functions for pulling datastreams from the object appear not to be working. Developers will do some more sanity testing and then Dave will move Fedora 3.6.2 to development (targeted for May 1). Dave noted that we should specifically test restricted datastreams. We also decided not to install the restrictions on API/A that Sho had imposed, not being able to discern any good reason for doing this. We'll also move to a new version of PHP at this time. It was noted that we will have to suspend development for some time until all the compatibility issues with Fedora 3.6.2 are fixed. Some of the compatibility issues may be related to the SOAP interface being deprecated. After we have the development server running on Fedora 3.6.2, we will need to reset the code freeze date for R7.2. Dave will also run the script this week for the country code database. The script to support the statistical profiles will be run next week.

Open Discussion on the use of Fedora Relationships

Ron introduced this topic with a brief document that provided an overview of Fedora's rels-ext and rels-int. The discussion then moved to the different types of relationship models that we expect to use going forward. The three types of models with examples are described below:

1. The ETD model. We concluded that ETD supplementary files should not be separate objects and therefore would not use relationships. Supplementary files are closely bound to the ETD and should therefore be represented as datastreams within the ETD object.
2. The Data project. We have modeled data projects using the collection object as an aggregator and developed various relationship ontologies for different types of data (datasets, documents, instrumentation, etc). These various resources are part of the data project but can frequently stand by themselves (and therefore are different than the ETD example). The relationships are represented as tabs in the data project user interface. As a side note, it was observed that there is still some confusion in the data portal search interface about when we should list projects in the results versus individual resources. We decided to leave the interface as is for now.
3. Various two-way relationships. We noted that there are many situations in which we want to implement a two-way relationship between resources. Examples include the relationship between a data project and the published article or the relationship between a resource and the license document. Rhonda cited a more specific NJDH example where two-way relationships should be supported between a portrait of an individual, an interview, and a biography. We concluded that we need to move forward on a general capability to support these two-way relationships with the rels-ext capability. It was decided that these relationships should be described using events with appropriate labeling. The ontology is frequently not an appropriate label for the user. The appearance of the event in the metadata may act as a trigger for WMS to create the two-way relationship and the resulting RDF XML. This more general capability is planned for RUcore R7.5 (code freeze February 15, 2014). In the meantime, we proposed to improve labeling by defaulting the label to the content model. MDWG will examine other options (e.g. type) before we implement this capability in R7.2.

With the above framework, we should be able to proceed with a specification for more generally handling of relationships in RUcore. This capability is a significant effort affecting WMS, dlr/EDIT, and portal code.

Discussion on the RUcore Software Versions Document

Ron introduced the discussion of how we want to update our software version document.

Although this process can be fairly tedious, everyone agreed that the document is a valuable tool to insure the correct versions of software are installed on all three servers. We also decided to use a format similar to what was used in the previous version in which valuable installation notes are provided. To get started, Ron asked everyone to review the document and suggest what needs to be added and what should be deleted (see http://rucore.libraries.rutgers.edu/collab/ref/doc_sawg_r5_0_software_notes.pdf). For

example, we will need to add a section on DOIs and the EZID version we are using. Although we have multiple layers of software, we will try an organizing principle in which the highest level software packages (e.g. WMS) are identified along with their dependencies. Ron agreed to be co-editor with Jeffery to begin the work on the document. Our plan would be to update this document for every major release and for dot releases if major version changes have been implemented.

Other Items

Yang raised a question regarding the labeling of archival masters. In a scenario in which there are multiple videos and documents (pdfs), we decided that these all would be sequentially numbered (i.e. ARCH1, ARCH2, etc) rather than restarting the ARCH numbering for each different file type.

Agenda Items for the Next Meeting (May 9)

- MDWG recommendation on the DOI profile
- Content models – next steps - WMS configuration, updating FD, stats, techMD (Rhonda)
- Status of Fedora 3.6.2 on development
- Google search results for RUcore records
- Further discussion on when to schedule sha-256 updates for legacy objects
- Begin discussion for R7.3 – User accounts

rcj – 05/06/2013