

## **SW Arch Meeting Minutes – September 26, 2013**

### **Agenda**

- Announcements and Updates
- A/A specification
- Multiple versions of articles (continued)
- Ontologies and Linked Data
- EAD Finding Aids

### **Announcements and Updates**

Ron attended the summer DataCite meeting and will follow up with a summary of the relevant discussions. Briefly, we should revisit the DataCite metadata profile to make sure we include the abstract and resource type. If both of these items are in the metadata submitted with the DOI, then Thomsen-Reuters will start indexing our data that is in the DataCite database. In talking with the Thomsen-Reuters representative, he indicated that all items will be indexed except publications (i.e. they are just indexing data). We'll need to understand their filter more but photographs and maps will definitely be indexed along with all of our datasets. We should also begin looking at persistent identifiers for people. The ORCID approach appears to have established quite a bit of momentum. Regarding data repositories, Purdue University Library appears to have done an excellent job of defining a deposit process and marketing to faculty.

### **Single Sign-on Authentication/Authorization Specification**

Yang reviewed this specification and described the functions that must be developed. For example, the ticket system will associate authorization privileges with the various roles (manager, faculty, student, etc). When a user signs on, she/he will receive a ticket with privileges. The ticket will remain through the session. Initial authentication will be through LDAP, Shibboleth, or a local account. The other main functions include the A/A Management System and the Community Authentication API. Development of the various modules for these functions will be distributed among the RUcore developers.

Given complexity and size and some deficiencies, we do not plan to use Grouper open source software and will develop our own local system. We will continue to track what OIT and the university are doing and there may be an opportunity to use Grouper at some point in the future. For release R7.3, we plan to bring the Analytic, faculty deposit, and WMS under this new A/A system.

### **Multiple Versions of Articles**

Jane led the continuing discussion of how we should handle multiple versions of articles. The discussion focused primarily on issues related to embargo and the cover sheet. For the initial roll out of this capability, we decided on the following:

- We will display only the latest version of an article. If a user requests an earlier article, manual intervention will be required by an RUcore administrator.
- We will use the standard Fedora versioning capability, i.e. an xacml policy will not be required to block access to articles that have been versioned.
- Descriptive events will be required for versions and should point to the datastream ID of the versioned article. Note, that initially, DOIs will not be available at the file level.
- For the cover sheet, we agreed to the following:
  - The cover sheet will be inserted at the beginning of the article.
  - As a result, the thumbnail should be generated from the second page.
  - The cover sheet should not be generated in real time because of performance issues.
  - As a result, the cover sheet should become a permanent part of the preserved object and will not include the date of access.
  - In order to include the DOI in the cover sheet, WMS will need to request the DOI and include it in the generation of the PDF.
  - Software architecture will investigate how to retroactively add cover sheets to articles already deposited.

The version capability is currently targeted for release R7.5 which will likely be a summer/fall 2014 release. Jane will update the specification and it will be reviewed again in sw\_arch to determine what capabilities might be implemented earlier.

### **Ontologies and Linked Data**

Grace reviewed how we might get started with ontologies and linked data in RUcore, noting that some of our controlled vocabularies are becoming very large and potentially awkward for users. Taxonomies build on the controlled vocabulary and can be used to establish hierarchical relationships. Ontologies incorporate these relationships and also provide semantic definitions for the conceptual domain. Grace outlined some “next steps” for MDWG and the technical development team (these steps are included in the slides and are not repeated here). Grace noted that this is a major piece of work and we need to get started, possibly beginning with the Oral History project.

### **EADs in RUcore**

Ron pointed out that we have an initial EAD architecture that was reviewed in CISC and have done some early prototyping to demonstrate that EAD XML generated in Oxygen can be ingested into

RUcore. Grace outlined an approach for implementing EADs in RUcore in which WMS can be used to describe the minimal metadata for digital resources to be referenced by the EAD. A meeting will be set up soon to discuss the architecture with the stakeholders in order to establish the final architecture (the meeting was subsequently set up for October 15).

**Agenda Items for the Next Meeting (October 10)**

- 
- Review of version specification to identify components that can be delivered prior to R7.5
- OCR, JRA Publish proposal (Isaiah)
- Integrity Report/Validate Object (Jeffery)
- Initial VMs (virtual machines) that will be needed
- Update on A/A – development plan

rcj – 10/07/2013