Minutes of January 7, 2010 Meeting

Agenda

1. Updates and Announcements
2.   Configuration document update
3. Release summary and R5.2
4. Major initiatives for 2010/2011

Announcements and Quick Updates

Rhonda and Linda will offering a one-day WMS workshop in March. Kalaivani reported that there is increasing interest in openWMS and we are looking to a first release in the March timeframe. We reiterated that openWMS is also code base for our ongoing RUcore releases, i.e. we will not have to code bases for WMS. Ron and Rhonda met with Tom Frusciano shortly before the holidays to discuss a photograph collection and portal for University Archives. We decided that UA collections would be positioned under RUL and at the same level as SPCOL. Tom and a student are working on a photograph collection dealing with the history of Rutgers. After the photograph collection is available, we will work with UA to establish a portal.   On another topic, there is considerable related work ongoing in the Data Working Group, specifically regarding compound objects and content models. This work will need to be brought into sw_arch in the near future. The DWG has targeted Feb. 10 to report to the steering committee. We have yet to be scheduled in CISC to report on the statistics package for R5.2. Ron checked download statistics for calendar year 2009, reporting that there were just over 50,000 downloads for the year. Grace suggested that we should publish more broadly this increased traffic and the attention ETDs are getting. In addition to our regular members, Sujay and Vineet from the NJVid project were able to join us for the discussion of R5.2 content.

Configuration Updates for R5.1

Jeffery has circulated the draft configuration document for R5.1 reflecting the changes from R.5.0. He has also circulated the "diff" document with a brief summary indicating that the primary differences included: 1) no longer requiring java bridge for WMS editing, 2) a new Perl module for handling Unicode texts, and 3) a new partner portal API. SW_ARCH members are urged to report any other changes to Jeffery that they are aware of. It was noted that we do need to "un-configure" java bridge so we won't be tempted to use it in the future. This diff procedure will become a standard part of our release process. Dave noted that the timing of this document should be earlier, i.e. prior to release. It was recommended that we target the release of the updated document with the code complete date, i.e. so this information is available for testing and installation on mss2, mss3 and for NJVid. As part of this discussion, several issues were raised about where documentation (other than readme files) should be placed and how do we identify and use common variables (e.g. a server variable). These questions are probably best addressed in the software methodology group.

Release Summary and R5.2

The group arrived at a consensus for R5.2 release content. The next steps include review with CISC and finalizing target dates. In this meeting, we set the preliminary target date for completion of requirements at January 31, 2010. The proposed target for releasing R5.2 is May, 2010, however it should be noted that we need to review in CISC and in sw_arch again before we can set more firm dates.

As part of this discussion, we addressed the question of updating NJVid to R5.1 or possibly skipping a release and going to R5.2. Ron noted that the important features in R5.1 for NJVid including the new portal API, a stats update, WMS support of MARC, bug fixes and significant performance enhancements. Sujay felt that these were important features and Dave indicated that the release on mss3 went quite smoothly. There is also an advanatage for software support to have both installations on the same release. So, Sujay and Vineet will proceed to move to R5.1. Jeffery will be his primary contact for release questions and will involve others as necessary.

The details of the release R5.2 are reported in the release summary. Major points and discussion are highlighted here.

- Rhonda reported that there is no apparent reason to export to UMI from Fedora. The current export capability from openETD meets all the UMI requirements and is an appropriate function for the openETD package. Ron had some concern regarding updates to metadata that would not be reflected in the ETD export. UMI does not need these updates, therefore we have eliminated the Fedora-based UMI export.

- XACML Utilities. We currently have two xacml utilities. One utility is used by NJVid to insert xacml license policies into the video object. This utility is not used by RUcore and is used only by one person (the Super Collection Manager at NJVid). As a result, it is probably not necessary to integrate this utility. For ETD embargoes, we need a quick fix by Chad to post an embargo email to Rhonda from the ETD application. Secondly, for R5.2, the embargo request must be posted to WMS, either as embedded xml or as file that can be ingested. For the Jazz Oral History Project (JOHP), we need what we called an "infinite embargo" since there are some oral histories that we will likely never get rights for. Regarding embargo capabilities, it was noted that we need to consider a common architecture and service for embargoing any object where the embargo date may be specific or intended to be forever.

- Although we have the jpeg2000/aware for maps as a unit goal, everyone felt that this capability should be lower in priority. As a result, we are recommending an investigation that would explore the benefits of moving to jpeg for different formats and for use as a presentation format, an archival master or both.

- Regarding Flash streaming, Isaiah reported that Wowza does not really support the streaming of Flash – it looks more like progressive download. There are also two remaining requirements issues: 1) determining if annotation can work with QT streaming and 2) understanding the NJVid need for Flash streaming, in addition to QT streaming. In any event, we concluded that Flash streaming should not be an R5.2 feature and will be identified as an investigation.

- The ongoing collection architecture working group under Rhonda's direction will likely produce requirements for RUcore. We haven't yet been able to assess the impact of these requirements.

- It was noted that significant improvements were made in the handling of Unicode characters in R5.1. However, entering characters can still create some problems. Jeffery, Yang, and Sho will continue to investigate.
- For ongoing quality initiatives, we need to understand and utilize Google scholar.

Major RUcore Architecture Initiatives for 2010/2011

Ron compiled a list of major architecture initiatives (see attachment A) for 2010/2011. We decided that compound objects will be addressed as part of the Data Working Group's efforts. Rhonda indicated that ETDs could survive without having this capability in R5.2, in part because Mason Gross is not sure how they want to handle supplemental files (e.g. they may not ingest supplemental files into the repository). Relationship services will be used to support compound objects for ETDs, datasets, and presumably for annotations (a question for Chad). Ron suggested that we consider introducing in R5.2 a capability to create, edit and delete relationships (Fedora provides API capability to do this). The major issue that is not already addressed as an investigation is covered in the "virtualization" item in the attachment. The most important feature is probably the ability to share collections across repositories, e.g NJDH will likely want to share videos that are in NJVid. There was considerable discussion on this issue regarding both requirements and multiple ways in which it could be implemented. Since the primary need at this point is with NJDH, Ron will ask Linda to lead a working group that should include some members from sw_arch to develop requirements for this capability.

Other Items

The objective is to review the release summary in the next CISC meeting scheduled for Jan. 13. Ron will touch base with Chad and Sho, who were not able to attend this sw_arch meeting. Our next meeting is scheduled for January 21, 2010.

Agenda Items for the Next Meeting

- Review of annotation specification
- (Other requirements that are available for review)
- R5.2 review and update per CISC comments
- Batch ingest for NJEDL – statistics and process
- Storage management on lefty64

rcj – 01/15/2010

Attachment A

Major RUcore Initiatives for 2010/2011

1. Compound objects. In general, we need to support compound objects. A compound object is a network of simple objects with semantic relationships for connections between objects. Compound objects are important for ETDs with supplementary files and for data objects that have a variety of simple objects.

2. Supporting relationships between objects as a general capability (i.e. creating, editing, deleting relationships)

3. Content models. We need more flexibility in our content model approach. For example, in R5.1, we discussed the issue of having WMS provide pipeline conversion of a smaller .wav files or, alternatively, doing an upload of a larger mp3 file. Fedora 3.3 supports enhanced content models (maybe as a plug-in?) – a capability that has relevance for both content models and compound objects.

4. Virtualization. This term is being used very loosely and refers to issues related to support of multiple repositories. As such, we need to consider federated searching, one WMS supporting multiple repositories, collection management across multiple repositories, sharing collections across multiple repositories, etc, etc. For example, NJDH would like to include videos from NJVid in its collections.

5. Search engines. Evaluating SOLR/Lucene and other search engines and replacement of Amberfish.

6. Use of annotation/analytics for objects other than video.

7. Storage architecture and purchase, configuration, installation of a new high capacity storage system.

8. Use of wowza in lieu of Darwin for video streaming of both flash and QT.

9. Jpeg2000/Aware server for maps (one of our 2010 unit goals).

10. Upgrade to Fedora 3.x

11. File validation using JHOVE2 and GDFR.

    Supporting multiple installations, i.e. NJVid and possibly others.