

## Software Architecture Meeting, Thursday - October 4th 2007

(Ananthan, Beard, Ellis, Geng, Hartman, Hoover, Jantz, Langscheid, Liew, Mills, Triggs , Yu)

---

### Agenda

---

1) R4.5 Specifications deadline. Expected outcome, all specifications will be submitted and reviewed and a code freeze date for R4.5 will be established.

- Collection Hierarchy display and searching specification
- PDF Server implementation specification
- Metadata Working Group WMS metadata specification
- User Services Faculty Deposit, search interface specification

Chad has placed the specification documents at the following for your convenience. Since they are not finalized yet they are not currently linked to from the RUcore website so this is the only way to get to them.

[http://rucore.libraries.rutgers.edu/collab/ref/spc\\_sawg\\_r4\\_5\\_coll\\_hierarchy.pdf](http://rucore.libraries.rutgers.edu/collab/ref/spc_sawg_r4_5_coll_hierarchy.pdf)

[http://rucore.libraries.rutgers.edu/collab/ref/spc\\_sawg\\_r4\\_5\\_pdf\\_server.pdf](http://rucore.libraries.rutgers.edu/collab/ref/spc_sawg_r4_5_pdf_server.pdf)

[http://rucore.libraries.rutgers.edu/collab/ref/spc\\_usawg\\_r4\\_5\\_interface.pdf](http://rucore.libraries.rutgers.edu/collab/ref/spc_usawg_r4_5_interface.pdf)

[http://rucore.libraries.rutgers.edu/collab/ref/spc\\_mdwg\\_r4\\_5\\_metadata.pdf](http://rucore.libraries.rutgers.edu/collab/ref/spc_mdwg_r4_5_metadata.pdf)

2) Handle server, single handle pointing to multiple objects (NOT COVERED AT THIS MEETING)

3) Digital Docs/Administrative Space R5.1 Specification Review (NOT COVERED AT THIS MEETING)

### Discussion

---

1) High Priority items from User Services Specs

**Release 4.5 action item:** Activation of **OAI Harvesting** and **Google Indexing** will be implemented by Jeffery Triggs in the next release of RUcore.

**Future Release:** An option in the collection record (RUcore and NJDH) will be added to indicate whether resource records can be harvested with OAI-PMH.

We will NOT be adding Multiple Related Items in R4.5. We will make do with the existing functionality of **collections having multiple parents** and **dynamic collection searching**. Chad mentioned that there needs to be architectural work done to support item-level objects being included in various collections, and it will need to be addressed in future releases.

While discussing the recently acquired NJVid grant, it was suggested we investigate federated, cross-repository searching of multiple Fedora instances as it is part of the grant.

**The following specifications are high priority and considered “Easy” by the group:**

1. When a collection record is displayed, allow the user to browse all the items in that collection. What we have in mind is a button or hypertext phrase: <Browse This Collection> in the area at the top of the initial record display that currently has the word <Collection> and an icon of a file cabinet. When that button/hypertext is selected, all the resources in that collection will be listed.
2. Link from the display of the Organization Name to contact information (from the Collection Record) of the Organization
3. In the “View Search in Context” disseminator, add an “i” [Information icon] for the PDF. The content of that information bubble should read:  
If the PDF link does not take you to the correct page, install (or re-install) the latest version of PDF <<http://www.adobe.com/products/acrobat/readstep2.html>>. For best results, close your browser before the re-install of Acrobat Reader, e.g.:
  - a. Download the latest version of Acrobat Reader  
<http://www.adobe.com/products/acrobat/readstep2.html>
  - b. Close all open browsers
  - c. Install/re-install Acrobat Reader (if re-installing, choose option to remove the old version first)
  - d. Restart browser
4. On the Intellectual Property Rights page, the name “Rutgers, The State University of New Jersey” must have a capitalized “The.” This needs to be corrected in the introductory sentence and in the Work-for-Hire choice.
5. In Step 2, add this sentence and link before the “Contact us” sentence:

Alternatively, you can consult this Director of Publishers and Vendors <http://www.acqweb.org/pubr.html> to determine if they allow deposit in RUcore, or to contact the publisher directly.

**The following specifications are high priority and considered “Hard” by the group:**

1. In Quick Search (and also in Advanced Search in a single search line), allow a phrase search to be combined with another term. For example, <“September 11” and children> should first search on the phrase <September ADJ 11> then search for the term <children> within those results.
2. We need these metadata elements indexed and searchable:
  - Journal or book title (host) in which the article/chapter/conference paper appears
  - Volume and issue of journal/book
  - Date (of issue) of journal/book
3. Metadata change requests (i.e., add):
  - Journal or book title (host) in which the article/chapter/conference paper appears
  - Volume and issue of journal/book
  - Date (of issue) of journal/book
  - Inclusive page numbers of article/chapter/conference paper in [host] journal or book
  - URL for article/chapter/conference paper on the publisher’s web site
  -
4. Supply a Required Publisher Statement for the genres *article*, *postprint*, and *preprint*. This usually requires a link to the article on the publisher’s web site. An example:  
REQUIRED PUBLISHER STATEMENT:

The published version of this article is available at:

<http://www.elsevier.com/locate/inca/505736/>

**R4.5 Action Item:** The following specification was decided top priority and would be in R4.5 regardless of difficulty:

1. Allow upload in any of the following: Word, PowerPoint, Excel, image (TIFF), HTML, PDF, PDF/A. Do not require any additional files (e.g., Word and PDF). Create PDF/A when the uploaded file is in Word, PowerPoint, Excel, HTML, image (TIFF), and PDF. Put the PDF/A plus and uploaded format in the tar file.
- 

## 2) Collection Hierarchy Widget

Chad presented a specification for user interface display of the Collection Hierarchy to be used for Advanced Search purposes. He presented the following possibilities: Slide, Scour, and Tree.

**R4.5 Action Item:** It was decided that Chad would prototype an “Enhanced Tree Widget”, as that seemed to be the most popular path to take.

Linda mentioned that NJDH does not have the “proper structure” to be pulled into RUcore. It was noted that the “Name of Organization” should be displayed at the highest level. It was decided that the solution was in fixing the metadata and the collection information in the dlr/EDIT collections database.

---

## 3) PDF Server

The main task of the PDF server is to **convert office documents** for Faculty Submissions and to **validate PDF/A** for RUetd.

The following additional services were suggested:

- Convert office documents for RUetd, requiring the student to review the converted PDF before submission. **This will be added to the next release of RUetd.**
- Apply OCR to PDFs which do not have a text layer (image-based). It was suggested this be handled with existing tools in WMS, though we should still experiment with the OCR bundled with the Adlib software.
- Concatenation or merger of PDF documents into a single file. This is lower priority, but could possibly be integrated into the interface for document conversion.

The question came up as to where the Faculty Submission gets converted to PDF? Another question is will any verification review will be done by faculty users or collection managers?

**R4.5 Action Item:** A reusable PHP class needs to be written for web application interaction with the PDF Server. Shaun and Chad will collaborate on this utility, with possible help from others. Development will take place on Lefty, but will also need to be compatible with the PHP environment on Harvest.

Clean-up of files on PDF Server will be handled by the Windows system admin. The PDF Server is not a storage device, but a conversion device so a regular file purge schedule should not cause problems.

---

#### 4) Open Source WMS Metadata Specs

We need more clarification on the Metadata specs and Kalaivani will work with MDWG to flesh it out.

---

#### 5) Future Meetings and Code Freeze Dates:

Next Software Architecture Meetings:

- Thursday, Oct 18
- Wednesday, Oct 31

R4.5 Release Dates:

- Code Freeze: Jan 4 (Partners Test throughout January!)
- Public Release: Feb 18