Spring 2016

Comparing Institutional Repository Software: Pampering Metadata Uploaders

Craighton Hippenhammer
Olivet Nazarene University, chhammer@olivet.edu

Recommended Citation

Comparing Institutional Repository Software: Pampering Metadata Uploaders

by
Craighton Hippenhammer
Digital Initiatives Librarian
Olivet Nazarene University

Presented at
Carson-Newman University
Jefferson City, Tennessee
June, 11, 2015

"Uploader": one who uploads digital material and metadata to institutional repositories.

Topics Covered

- Institutional repositories, Digital Commons and Wesleyan Holiness Digital Library strengths
- Important software structures and features
- Important support systems
- Institutional repository quality factors

Digital Commons (DC)

Digital Commons is a commercially hosted institutional repository product whose purpose is to host searchable, electronic files, usually faculty scholarship and archival material. It was originally created for the University of California with a customer base of about 400 institutions, mostly U.S. universities, all in English. It has a robust support staff that enables customers to have highly customized sites that work well for them, even though they're based on highly structured templates. Their main competitors are open source options like Dspace and Fedora, which both take intense Information Technology (IT) development time and ongoing local support. It's very common to feel you have to fight for IT's attention when they have a whole campus to attend to. With Digital Commons, those concerns go away.

Digital Commons' home pages contain a colorful circle that is a graphical representation of the disciplines that uploaded material have been divided into, It's one way content can be browsed, not only in one's own repository, but throughout the repository of all DC customers.

Digital Commons has a hierarchical internal structure. Documents are filed within scholarship and archival series which are filed within departments and communities. This structure shows up in document urls and creates an academic way of browsing. Each document type has its own metadata form that is created to relate to the discipline that it's in. So each article, ejournal, book, image gallery and event form limits its fields to those necessary to its discipline-influenced document type. A book gallery, for example, is unique as it has the capability to make the first page of its pdf into a separate thumbnail and use it to display the cover.

Searching faculty and students can also browse for topics by department, university center, or program. Subsumed under each department will be department-flavored and document-type subcategories helpful for coming across topics through guided serendipity. These are highly configurable, like an art gallery painting project by one of our art professors that shows an area of a map of California where the retreat site was held. Breadcrumbs (Home > Art Dept > Art Image Galleries > DVP Art Gallery) for these pages are also configurable.

Digital Commons' strengths include metadata upload forms that are configurable and kept simple with no distracting unused fields. Metadata forms tie document type to academic centers and disciplines which keeps the repository academically related. It has a very strong search capability – both Google and local site searches. DC technical staff stay in close touch with Google to make sure metadata field codes are up to date and the way that communicates best with Google technology. DC also has very strong support. New DC sites can be built within days. It encourages sharing between members and conducts webinars and other training sessions. Members can call and email support with one-day response times.

Wesleyan Holiness Digital Library (WHDL)

For the past three years, I have been involved with the creation of a made-from-scratch institutional repository, a theological IR that is currently being created by and supported by the Nazarene Church, the denomination with which my university is related. The repository specifications, design, and ongoing development are being led by a team of 15-20 Nazarene librarians and denominational leaders who meet in person once a year and monthly via conference calls. So an open source software developer was hired in the Fall of 2012 and they programmed it in PHP using the Drupal Content Management System, going live in June, 2013, at http://www.whdl.org, an institutional repository called Wesleyan Holiness Digital Library.

WHDL is now starting to give birth to university IRs using the same engine, the first going up at MidAmerica Nazarene University in Olathe, Kansas, in April 2015, at http://repository.mnu.edu/content/mnu-institutional-repository. The goal is to provide daughter IRs to most of the over fifty Nazarene institutions around the world (as long as they have an adequate electrical infrastructure for the IR to run on) as well as to other Protestant denominational universities who are in the same Wesleyan theological universe. The Nazarenes are now looking for denominational partners to join them in this effort.

A key specification desired by this group at the very beginning was that the IR had to be fully multilingual, a difficult hurdle to jump, since even Digital Commons cannot yet provide that. To be truly multilingual there must be three areas in the software where the language must be compatible and relate correctly to each other: the documents themselves (the easiest hurdle), the metadata that describes the documents, and the interface which includes the navigational words to get around in the program (also called the site language).   WHDL went live with five site languages – English, French, Portuguese, Spanish, and Korean – and has documents up now in 58 different languages.

Although this IR is new and still developing, it is amazing that it is so thoroughly multilingual. The IR highlights the capability with a language drop-down box at the very top of the home page. Librarian administrators and uploaders can log in and get a left navigation bar that contains a workbench, forms to upload new items, collection queues for in-process unpublished-yet documents, subject category lists and training documents.

IR Comparison

In WHDL the document type is selected first without being attached to a discipline or academic center of study. It has many more document types than Digital Commons has. But Digital Commons' forms are selected by document type tailored by the discipline community it is associated with.   This emphasizes the academic quality and nature of the material.

Digital Commons' web site layout is quite configurable within a fairly limited number of strict template formats. DC staff, however, will add fields that no one else has if you really need them and they don't conflict with the structure of the system. Some fields you have to know to ask for because they don't come with the default configuration. The peer reviewed checkbox, for example, is one they have available, but may be suitable for only certain series or galleries. So you have to tell them which ones you want them to add it to.

Digital Commons has no subject fields, but uses keyword fields and disciplines field only. The number of disciplines and sub-disciplines is quite narrow which balances quite nicely with the infinite number of key words and key word phrases that can be used. WHDL has been toying with the idea of using subject headings, like Library of Congress, which is unwise. Who is going to do the translations of the subject headings list into exotic languages to keep up the multilingual idea? It's undoable. Translating one to two thousand discipline terms per language, though, is quite attainable.

Having a scholarship domain field in Digital Commons based on the Boyer model is unique to Olivet. Ernest Boyer, president of the Carnegie Foundation for the Advancement of Teaching, published *Scholarship Reconsidered: Priorities of the Professoriate* (https://www.csusm.edu/community/facultyengagement/scholarshipreconsidered.pdf) in 1990 that sought to widen the definition of scholarship beyond just publishing by creating four overlapping dimensions or domains: the scholarship of discovery, the scholarship of integration, the scholarship of application, and the scholarship of teaching. To the Boyer model of four scholarship domains, Olivet added a fifth: the scholarship of faith integration. We just asked Digital Commons Support to add that to our metadata forms and they did it.

An embargo period for publisher-published articles means that they insist the article can't go open access for x-number of months after they publish them. Digital Commons can institute that field for any series or gallery you want; you just have to ask and designate where. Just type in a date twelve months down the road and a year later the document you hid with this function will become available to the Web automatically without further effort. Some of these more exotic features may be developed in WHDL, but were not a priority in its early development stages.

Digital Commons automatically creates an OpenURL for previously published articles. You can also attach additional files of any type to the record. These additional files can be designated to be visible or hidden. It is common to attach copyright permission documents in this function and make them invisible but also nearby, associated with the document it belongs to. DC also has a Creative Commons license field that most universities use, which lets researchers know how and whether they can use the material you post.

WHDL Metadata forms are divided into tabs: essential, supplemental, administrative settings, revision information, flags. Less used fields are on other tabs or tabs that only administrators access.

WHDL divides responsibilities into Librarian 1, Librarian 2, and Librarian 3 levels for access privileges. In Digital Commons, access can be restricted to particular series or other document types.

IR Software Quality Factors

The design of the software is the most important quality, especially in designing it for ease of use for those who upload. That is created by giving the uploader the ability to hide metadata fields that are not needed, the ability to add fields that other universities may not need, and the ability to tailor field options. Why not allow the option to pick from dropdown lists if the options are known? For example, if the metadata field is "Department," and all college departments are known, then the software should have all of the department names in a drop-down list readily available for the uploader to select rather than having to type them as free text. If a field entry is the most-often-chosen option, then that entry should be made the default option at the top of the list. If one department publishes a lot more than any other department, then that department option should be selected as the default. That way it doesn't have to be manually selected, which will save keystrokes and speed up the metadata entering process.

Different document types need different metadata templates. For example, electronic theses and dissertations (ETD's) need mentor name fields; other document types don't. Books don't need volume and issue fields like journal articles do. If your metadata form expects the uploader to know that and to therefore skip over such fields, then sites where untrained personnel or volunteers are used are likely to run into trouble because of the unnecessary software structural design problems. No IR should use IT jargon like "node" or direct uploaders to do illogical things or retrieve data from "out on the web." Newly developed IR software is more likely to have these sorts of problems.

Also, book galleries profit from having the first page of a pdf made into a thumbnail graphic to highlight the book's cover. The IR software should be able to process pdf's created by word-processing software and to display the thumbnails properly as well as those produced by Adobe products. Digital Commons has had a problem with this.

Search Functions

The search function in institutional repositories is extremely important. Documents they display must be well searchable by Google and Google Scholar. Metadata field tags must be Google friendly. Most documents in the IR must be scholarly or Google Scholar will not index it. IRs must stay current with Google field and algorithm changes. IR metadata and IR documents must both be searchable and have the capability of using search limits by metadata field and language.

Excellent Support

Every institutional repository should have expert support people who can change and tailor the IR software to suit your college/university needs; who can guide you to use the software effectively; who offer webinars and instruction; and who can readily be reached via email & phone. Support should also supply online discussion groups among members, and regularly maintain and upgrade the software, including Google search engine optimization (SEO).

Sources

Digital Commons. Berkeley Electronic Press. Berkeley, California. http://digitalcommons.bepress.com/.

Digital Commons @ Olivet. Olivet Nazarene University (Bourbonnais, Illinois).
http://digitalcommons.olivet.edu.

Wesleyan Holiness Digital Library (WHDL). Church of the Nazarene. Kansas City, Missouri.
http://whdl.org.

MNU Institutional Repository. MidAmerica Nazarene University. Olathe, Kansas.
http://repository.mnu.edu/content/mnu-institutional-repository.

Boyer, E.L. Scholarship Reconsidered: Priorities of the Professoriate. The Carnegie Foundation for the
Advancement of Teaching. 1990.
https://www.csusm.edu/community/facultyengagement/scholarshipreconsidered.pdf.