Cornell University
ILR School

Cornell University ILR School
DigitalCommons@ILR

Articles and Chapters

ILR Collection

3-1-2007

Creating a Multipurpose Digital Institutional Repository

Suzanne A. Cohen
Cornell University, sac23@cornell.edu

Deborah J. Schmidle
Cornell University, dj13@cornell.edu

Follow this and additional works at: http://digitalcommons.ilr.cornell.edu/articles

This Article is brought to you for free and open access by the ILR Collection at DigitalCommons@ILR. It has been accepted for inclusion in Articles and Chapters by an authorized administrator of DigitalCommons@ILR. For more information, please contact jdd10@cornell.edu.

Creating a Multipurpose Digital Institutional Repository

Abstract
DigitalCommons@ILR is a multipurpose institutional repository (IR) for scholarship produced by faculty at the School of Industrial and Labor Relations at Cornell University. Unlike most IRs, it also functions as a subject-based repository for workplace-related information. This paper will discuss the issues involved in the implementation of DigitalCommons@ILR, including the choice of software, collection scope and policies, organization, and staffing. Keys to success in developing repository content, including building administrative support and developing partnerships, will be noted.

Keywords
repository, scholarly publications, digital, commons, ILR, digitalcommons@ILR, subject

Comments
Suggested Citation:
Cohen, S. A. & Schmidle, D. J. (2007). *Creating a multipurpose digital institutional repository*. Retrieved [insert date], from Cornell University, School of Industrial and Labor Relations site:
http://digitalcommons.ilr.cornell.edu/articles/118

Required Publisher Statement:
Copyright by Emerald Group Publishing Limited. Final paper published as Cohen, S. & Schmidle, D. (2007). Creating a multipurpose digital institutional repository. *OCLC Systems & Services, 23*(3),287-296.

This article is available at DigitalCommons@ILR: http://digitalcommons.ilr.cornell.edu/articles/118

CREATING A MULTIPURPOSE DIGITAL INSTITUTIONAL REPOSITORY

Suzanne A. Cohen
Collection Development Librarian
Martin P. Catherwood Library
521F Ives Hall
ILR School
Cornell University
Ithaca, NY 14853-3901
voice:(607)255-7105
fax:(607)255-2741

Deborah J. Schmidle
Social Sciences Coordinator
Collections, Reference, Instruction and Outreach
106 Olin Library
Cornell University
Ithaca, N.Y. 14853-3901
Ph: (607)-255-3680
Fax: (607)-255-3609
Email: dj13@cornell.edu

March 1, 2007

Copyright by Emerald Group Publishing Limited. Final paper published as Cohen, S. & Schmidle, D. (2007). Creating a multipurpose digital institutional repository. OCLC Systems & Services, 23(3),287-296.

Creating a multipurpose digital institutional repository

Abstract

DigitalCommons@ILR is a multipurpose institutional repository (IR) for scholarship produced by faculty at the School of Industrial and Labor Relations at Cornell University. Unlike most IRs, it also functions as a subject-based repository for workplace-related information. This paper will discuss the issues involved in the implementation of DigitalCommons@ILR, including the choice of software, collection scope and policies, organization, and staffing. Keys to success in developing repository content, including building administrative support and developing partnerships, will be noted.

Introduction

The Catherwood Library, a unit within the Cornell University Library system, physically resides in and is funded by Cornell's School of Industrial and Labor Relations (ILR). In addition to its four-year undergraduate program and graduate programs focusing on workplace issues, ILR offers practioner-based programming at five Extension sites throughout New York State. The ILR School currently consists of 110 resident and extension faculty, and library staff have many opportunities to work closely with these faculty through reference, instruction, and outreach services. The Catherwood Library collections, including the special collections housed in its Kheel Center for Labor-Management Documentation and Archives, serve as a comprehensive national resource for workplace-related information.

Due to a desire within ILR to better promote the work of its faculty and researchers in a cohesive manner, along with the Catherwood Library's interest in obtaining and preserving scholarly information produced by ILR, a school-wide committee was formed to discuss the possibilities. This committee, which was endorsed by the ILR Dean, had representation from the library, resident and extension faculty, the ILR Web Team, and a technology specialist from the

Cornell University Library. This committee laid the foundation for Catherwood's IR and developed early support throughout ILR. However, the project received the focused attention that was needed only after Catherwood hired a full-time Web and Digital Projects Manager to assume full responsibility for the implementation of the IR.

Choosing a Platform

Much of the discussion that took place during the early committee meetings focused on the platform for storing materials in the IR. The options included using existing open source software, contracting with a commercial service to host the repository, or building a repository platform specifically to meet Catherwood Library and ILR needs (since Catherwood already had experienced technologists on its committee). By the time the Web and Digital Projects Manager was hired, the development of IR technology platforms had rapidly evolved, and there were many more choices within each category to consider. While the idea of a customized system to meet Catherwood and ILR needs was attractive, the requisite financial and staffing investment was prohibitive. Also, by this time, Cornell University Library had volunteered to participate in the Mellon-funded project to promote adoption of the DSpace platform at major research universities, and there were also more commercial possibilities. After much collaborative research with staff in the ILR Web Team, a decision was made to contract with Proquest/bepress to use the Digital Commons[1] platform, starting in December 2004.

This was not an easy decision, considering the local availability of DSpace at Cornell. However, the following factors proved instrumental. Digital Commons

- was ready to use and came with Proquest-provided training, enabling production to move quickly;

[1] http://umi.com/products_umi/digitalcommons/

- allowed branding of the site to match the branding of the ILR School (this branding was viewed as a means of getting greater participation from faculty and other partners);
- had a proven track record with its adoption at the University of California and other academic institutions;
- accommodated OAI-compliant metadata and allowed controlled vocabulary pick lists;
- offered multiple backup and safety protocols;
- allowed web-crawling;
- provided usage statistics, including hits on both document abstract pages and full text downloads, and included statistics on referrer pages;
- hosted data off site (no need to find local server space); and
- provided all of the technical support as part of a written contract.

Most importantly, the partnership with Proquest/bepress included a 12-month contract, which allowed an exit strategy if another platform proved superior in meeting the needs of Catherwood. The contract allowed for the retention of ownership of the data and the domain name if the relationship with Proquest/bepress was terminated.

Although the rapid evolution of IR software and increasing institutional investments into IR development addressed some of Catherwood's original concerns about DSpace and other platforms (e.g., DSpace is now crawled by Google), Catherwood identified both technical and organizational reasons to renew its contract with Proquest/bepress after the first year. The technical reasons ranged from specific items (such as automatic conversion of Microsoft Word uploads into PDF) to the more general issues of technical support (including the ability to take advantage of Proquest's marketing and business resources). Proquest/bepress staff have been responsive to Catherwood's requests for technical improvements and they established an organized online forum for the Digital Commons user community to share questions and make suggestions. Catherwood was offered opportunities to participate in beta-testing of tools such as "automatic population" of the IR using

Proquest data, and has benefited from development of search engines such as Research*Now*[2], which includes the peer reviewed journals and other content in all of the IRs hosted by Proquest/bepress. The organizational reasons for Catherwood's decision about a platform hinged on a need for the greatest level of flexibility in managing and customizing the repository to satisfy the needs of different partners, and that need has been addressed thus far.

Developing Content: Selection Issues and Criteria

An IR is traditionally defined as "a digital archive of the intellectual product created by the faculty, research staff, and students of an institution and accessible to end users both within and outside of the institution, with few if any barriers to access" (Crow, 2002). IR's are usually contrasted with discipline-specific repositories and subject-oriented or thematic digital libraries, which are organized around the discipline/subject, rather than the institutional origins of the contents (Crow, 2002; Johnson, 2002; Lynch, 2003). DigitalCommons@ILR is a unique combination of two types of repositories. First, it is an IR for the ILR School. Second, it is a subject-based repository, which captures born-digital and grey literature on the subject of industrial and labor relations. The only criterion for inclusion is that the documents must relate to workplace issues.

In the fall of 2005, the DigitalCommons@ILR Working Group developed a formal Collection Development Policy[3]. This policy, which was essential in order to guide decision-making and dedication of resources, specifies collection priorities as follows:

Priority 1 – Any material (scholarly articles, congressional testimony, etc.) emanating from ILR faculty (resident and extension);

Priority 2 – Scholarly material published or produced by the School;

Priority 3 – Non-ILR-produced documents.

Related to Priority 3 is a Digital Repository Submission Agreement that was developed by Catherwood Library staff and vetted through the Cornell Office of

[2] http://researchnow.bepress.com/
[3] http://digitalcommons.ilr.cornell.edu/dc_information/4/

University counsel. Modified versions (depending on the situation) have been signed by depositors outside of the ILR School.

In addition to content criteria, another important issue was the organization of the repository content. In order to accommodate an interest from the Dean and Associate Dean in presenting the ILR School as a single entity, documents are not categorized by distinct sources (e.g., Academic Departments, Centers and Institutes). Instead, there is a single category for "ILR Collection," which is organized by document type: Articles & Chapters; Book Reviews; Policy & Issue Briefs; Conference Proceedings, Presentations and Speeches; Manuals and User guides; Monographs; Newsletters; Research Studies and Reports; Student Works; and Working Papers. For those interested in browsing by specific units on campus, the Digital Commons platform allows Catherwood to cross-list the documents into unit-specific series (see Figure 1). Institutions considering the creation of a new institutional repository should think carefully about how to organize the collections within a repository, since it is not always easy (or desirable, if there is an interest in permanent access) to move documents around.

Figure 1. DigitalCommons@ILR Browse Screen

Administrative Support

Support from the ILR Dean and Associate Dean from the beginning of this project has been crucial to getting participation from faculty. In addition to the library director's communications about the availability of DigitalCommons@ILR, the deans sent school-wide e-mail communications encouraging participation, and they also welcomed library staff making a presentation at a full faculty meeting. In preparation for this full faculty meeting, the DigitalCommons@ILR Working Group and project staff did the following:

- worked to establish a "representative sample" of documents in the repository, enabling more successful demonstration of possibilities;
- organized an educational session for the faculty administrative assistants, anticipating the need for their assistance in working with faculty;
- developed a DigitalCommons@ILR brochure[4] to aid in marketing; and
- named library staff liaisons to serve as a single point of contact for various ILR groups (e.g., resident faculty, extension faculty).

At the faculty meeting, the presentation included general education on scholarly communication issues and copyright regulations, an overview of DigitalCommons@ILR, and specifics about how to participate. The meeting, one year after the launch of the IR, was a turning point in terms of faculty participation, as many faculty afterwards expressed interest in getting their works into DigitalCommons@ILR.

If we build it, will they come?

A theme throughout the IR literature is the challenge of growing content in repositories (Allard, 2005; Buehler and Boateng, 2005; Foster and Gibbons, 2005). Recognizing this challenge, the Catherwood Library made a commitment that once a faculty member expressed interest in participation, DigitalCommons@ILR project staff would do most of the work related to getting the content into the repository. This work involves checking vitas for eligible documents, checking and seeking copyright permissions, and uploading the documents with accompanying metadata. Available tools for checking copyright permissions such as SHERPA's Romeo[5] make this process a bit easier, but not all publications are listed on this site, and the negotiation with publishers can be very time consuming. As soon as faculty learn that there is no work required of

[4] http://digitalcommons.ilr.cornell.edu/dc_information/5/
[5] http://www.sherpa.ac.uk/romeo.php

them (aside from the possible provision of a final manuscript version of an article, if required by the publisher), they are happy to participate.

While the Catherwood Library is committed to doing all of this work, the Web and Digital Projects Manager agreed that if a faculty member expressed interest in having the control of administering his/her own content in the IR (perhaps with the assistance of an administrative assistant), the Project Manager would provide the training. Almost none have expressed such interest, aside from a couple of exceptions, which will be described later.

Staffing

Because of Catherwood's commitment to do much of the work involved in getting a document into the repository, DigitalCommons@ILR has involved a large investment of library staff resources – more than anticipated. Due to centralization of the technical services function in the Cornell University Library system, Catherwood was able to realign some technical services to web and digital projects, including the IR development. Table 1 describes all of the individual and group roles that have evolved to support DigitalCommons@ILR.

Table 1. Staffing in support of DigitalCommons@ILR

Web and Digital Projects Manager	Investigate and choose IR platform Oversight of all aspects of IR Supervision of 1.5 FTE support staff Chair DigitalCommons@ILR Working Group Main technical administrator Primary contact with Proquest/bepress Acts as final negotiator for copyright permissions Primary trainer for those administering the IR
Project Assistant	Uploads documents and enters all metadata
Project Assistant (part-time)	Uses established workflow to check copyright permissions on publications.
Liaisons (3 existing library staff assigned to different ILR School groups)	Serve as a single point of contact for faculty who are interested in participating, and provide education on scholarly communication and copyright issues.
DigitalCommons@ILR Working Group (project manager, the Reference Services Coordinator, the Collection Development Librarian, and the Outreach Librarian)	Works on policies and workflow issues. For example, worked together to create the Collection Development Policy, the faculty Publications Workflow procedure, and the Digital Repository Submission Agreement
DigitalCommons@ILR Advisory Group (larger group consisting of	Provide institutional oversight for the IR, allow opportunities for input from stakeholders, and generate ideas for

all of the above, plus staff outside of the library representing Faculty, Staff, and ILR Web Team	promotion and marketing.
Selectors for Catherwood Library collection	A Reference librarian has the primary responsibility for the selection of documents in the library's Key Workplace Documents portion of the IR, but others may also identify appropriate documents for this collection. Archivists from the Kheel Center for Labor-Management Documentation and Archives select material that may be appropriate for the Kheel Center collection.
ILR Web Team staff (not library staff)	Advise on technical issues; Assists with web design customization, in order to match the branding of the ILR School; developing tools to harvest metadata from DigitalCommons@ILR for use on local ILR School web pages.

Funding and associated costs

Clearly, there is a cost to the library when all of the staffing resources described above are devoted to implementation and maintenance of an IR. It is difficult to measure the total staffing cost, since all of the staff involved also have other duties within the library. Currently, the amount of time spent on the IR development changes as a result of shifting priorities within the library (for example, DigitalCommons@ILR did not get as much focus when the Catherwood Library was launching a new web site).

Because hosting of the IR was outsourced, Catherwood does have one fixed cost, which is the contract with Proquest/bepress. At this time, the cost of this contract is being paid for by income from a Catherwood Library endowment. Equipment is another fixed cost, including the initial purchase of a computer and scanner, and the purchase of second monitors for those actively working on the administration of the site in order to ease the strain of going back and forth between metadata screens and document information. There will be additional, future costs related to the preservation of the DigitalCommons@ILR content, including migration to other formats, and this is not currently built in to the library budget.

Creating win-win partnerships

Partnerships both inside and outside of the ILR School have been crucial to the development of this repository. At the time of writing, of the almost 2,000 documents in the repository, about one-third are published by faculty and staff within ILR, while the remaining two-thirds fall into the born-digital or gray-literature category. This section will describe some of these partnerships and the potential benefits gained by these partners.

Public Relations and Increased Visibility

From the point of view of the School's administration, there is a strong public relations incentive for having faculty make their work more widely available and discoverable through web search engines. They believe it can be used as a powerful marketing tool for advancing the research of the institution. This is often cited as a rationale or incentive for participation in IRs (Crow, 2002; Johnson, 2002). At ILR, this incentive seems most attractive to assistant professors, who will be facing a future tenure review, and they are the only examples of ILR faculty wanting to upload "in press" articles. Interestingly, an Assistant Dean recently inquired whether DigitalCommons@ILR could aid in reviewing the publications of a faculty member being considered for tenure. The answer was "yes and no," since the faculty member being reviewed does have publications in DigitalCommons@ILR (making them convenient to access), but it is not a complete representation of that person's works, due to copyright restrictions. A different professor reported receiving more phone calls from the press, since making publications available in DigitalCommons@ILR.

In response to marketing efforts, there has been enough faculty interest in DigitalCommons@ILR to keep Catherwood busy checking vitas and copyright permissions and uploading documents with accompanying metadata. Approximately 67 out of 110 faculty are currently not participants. Why participation has been low is not readily apparent, because there has been no formal assessment of the reasons for non-participation. As Catherwood nears

the end of processing the documents of "early adopters," non-participants are being contacted. In many cases, these non-participants are willing to have their vitas checked by Catherwood once Catherwood staff has made personal contact with them.

The Potential for Increased Sales

Another partner that was approached for inclusion in DigitalCommons@ILR was the ILR Press[6], which was founded in 1946 as the publishing division of the ILR. While it is now an imprint of the Cornell University Press, the ILR Press retains editorial autonomy with an editorial board drawn from the ILR School. The ILR Press focuses on books concerning workplace issues, labor management, and social policy.

The ILR Press was approached after ILR faculty who have book chapters published in ILR Press publications asked if they could have those chapters included in DigitalCommons@ILR. A proposal was submitted to the ILR Press for such inclusion, as well as to use the IR to create a series that would feature ILR Press titles much the way that Amazon offers "samples" of a book for potential consumers.

The ILR Press granted permission for book chapter postings and for selected ILR Press imprints to be "sampled" within DigitalCommons@ILR. Content in the IR includes the book cover, table of contents, and first twenty-five pages of text. Each book entry includes an acknowledgement consisting of the full title, author name, and copyright line as it appears in the book, as well as the following statement: "Used by permission of the publisher, Cornell University Press. All rights reserved to and by the publisher." To date, it is too early to tell if this added exposure has increased sales for the ILR Press.

Metadata Harvesting and International Relationships

As part of the ILR School, the Employment and Disability Institute (EDI) conducts research, provides continuing education opportunities, and offers

[6] http://digitalcommons.ilr.cornell.edu/ilrpress/

technical assistance on numerous aspects of disability and the workplace. EDI has a long history of publishing and was already investigating strategies for making documents accessible online when the Catherwood Library decided to develop an IR. EDI is one of the partners given administrative rights to manage its portion of DigitalCommons@ILR (including entering of metadata and uploading of documents), but library staff still do the copyright permission checking. In October 2005, the EDI "Publications Page"[7] was launched, and it was the first ILR School web site page to pull metadata from DigitalCommons@ILR in order to display documents organized by topic in the context of the EDI web site. When visitors choose a topic, they are then brought to an automatically generated search results page within DigitalCommons@ILR.

Related to EDI are the Global Applied Disability Research and Information Network (GLADNET)[8], which was established in 1997 as an initiative of the Disability and Work Programme of the International Labour Organisation (ILO), a specialized agency of the United Nations. GLADNET promotes disability policy and program reform with an emphasis on integrated training and employment options for adults with disabilities. The Director of EDI is currently the Chair of the GLADNET Board of Directors, and EDI hosts the GLADNET web site. When an opportunity came up to move an already existing GLADNET infobase of documents into DigitalCommons@ILR, the close relationship of the organization with the ILR School facilitated the move. This is an active collection to which new items are added, as selected by GLADNET staff or Catherwood Library staff. GLADNET was pleased to have a new home for its documents, and the Library was happy to add this important workplace-related collection to its repository.

Online Scholarly Journal Access

Founded in 1947, the *Industrial & Labor Relations Review* (*ILR Review*)[9] is a quarterly publication that provides interdisciplinary research on all aspects of employment, as well as reviews of books in the area of industrial and labor

[7] http://www.ilr.cornell.edu/edi/m-pubs.cfm
[8] http://www.gladnet.org/
[9] http://digitalcommons.ilr.cornell.edu/ilrreview/

relations. In March 2005, the *ILR Review* Managing Editor was exploring options for making the full-text of the journal's recent content available digitally (older content is available through Ebsco's Business Source Premier, Hein Online, and JSTOR). Exploration of this option coincided with Catherwood staff approaching the *ILR Review* editor with the option of participating in DigitalCommons@ILR. It soon became apparent that this was a win/win situation for both DigitalCommons@ILR and the *ILR Review*. The costs associated for the *ILR Review* were minimal (no out-of-pocket costs for digitizing the materials) and were limited to staff time. Updating DigitalCommons@ILR with new batches of articles and books reviews takes only a few hours. The editors of the *ILR Review* were very pleased to have a new forum for their publication. The DigitialCommons@ILR Working Group was naturally thrilled to have a journal included in the overall content.

Various negotiations were necessary to bring this idea to fruition. The *ILR Review* had to work with EBSCO to allow permission for material to be added to DigitalCommons@ILR, and with Proquest to allow for some restricted access to materials (all visitors can view and download any book review and any archived article that is more than one and a half years old, while subscribers can view and download recent and current articles). Licensing agreements also had to be worked out, but ultimately, this partnership was a perfect example of multiple entities (the *ILR Review*, Proquest, bepress, EBSCO and the Catherwood Library) working together to ensure that the needs of all were met. The ILR Web Team was able to draft a presence in DigitalCommons@ILR that looks exactly like the *ILR Review's* own web site and the journal officially joined DigitalCommons@ILR in September 2005. As of this writing, nearly 60% of the *ILR Review* subscribers have requested online subscription access to the current issues.

Born-Digital, Increased Access, and Serendipity

The Catherwood Library and Kheel Center Collections within DigitalCommons@ILR comprise a wide-range of documents including, workplace-related government documents, librarian-produced bibliographies, publications about the ILR School, and digital versions of primary source documents housed in the Kheel Center. The loss of workplace-related documents that are born digital or are seemingly here today and gone tomorrow has long been a concern for Catherwood Library, as it is considered to be a "library of last resort" in the subject area of Industrial and Labor Relations. With the commitment to host and develop the IR, library staff was eager to take advantage of the repository technology to serve other library goals, such as preserving digital information and increasing access to resources currently only available in print. Documents that the library may have once selected in print are now received digitally and put into DigitalCommons@ILR if copyright permissions allow. Selected documents from the Catherwood and Kheel collections have also been scanned and made available in the repository, based on perceived demand.

While more than 95% of users visiting DigitalCommons@ILR are referred from Google (from available referrer statistics), once a researcher is in DigitalCommons@ILR, her or she may choose to browse or search within Catherwood's workplace-related subject repository. For this reason, ILR administrators are pleased that non-institutional, workplace-related documents are also included in DigitalCommons@ILR, potentially increasing the visibility of these works ILR administrators view the Catherwood Library as part of the ILR School institution, and therefore inclusion of Catherwood collection documents is seen as a logical extension of the IR.

Conclusion

At the time of writing, DigitalCommons@ILR includes almost 2,000 documents, and receives about 6,300 full-text downloads of repository content per week. One of the biggest lessons learned was the need to actively work with partners, (whether they are faculty, publishers, or workplace-related organizations), in order to promote the potential benefits of participation in DigitalCommons@ILR. Catherwood Library staff has been able to build on already existing relationships with faculty to get early participation. Creating additional contacts and seeking out new opportunities, has allowed staff to develop new relationships that can only increase the relevance of the library to the ILR community. Maintaining and growing DigitalCommons@ILR requires a large amount of staffing resources. An ongoing commitment to such resources will be crucial to the continued development of DigitalCommons@ILR.

References

Allard, S., Mack, T.R. & Feltner-Reichert, M. (2005), "The librarian's role in institutional repositories: A content analysis of the literature", *Reference Services Review,* vol. 33, no. 3, pp. 325-336.

Buehler, M.A. & Boateng, A. (2005), "The evolving impact of institutional repositories on reference librarians", *Reference Services Review,* vol. 33, no. 3, pp. 291-300.

Crow, R. (2002), "The case for institutional repositories: a SPARC position paper", available at: http://www.arl.org/sparc/IR/ir.html (accessed October 10, 2006).

Foster, N.F. & Gibbons, S. (2005), "Understanding faculty to improve content recruitment for institutional repositories", *D-Lib Magazine,* vol. 11, no. 1, pp. 1-12.

Johnson, R.K. (2002), "Institutional repositories: partnering with faculty to enhance scholarly communication", *D-Lib Magazine*, vol. 8, no. 11., pp 1-7, available at: http://www.dlib.org/dlib/november02/johnson/11johnson.html (accessed October 10, 2006).

Lynch, C. (2003), "Institutional Repositories: Essential infrastructure for scholarship in the digital age", available at: http://www.arl.org/newsltr/226/ir.html (accessed October 10, 2006).