University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Faculty Publications, UNL Libraries | Libraries at University of Nebraska-Lincoln

10-1-2009

Case Study: DigitalCommons@University of Nebraska–Lincoln

Paul Royster
University of Nebraska-Lincoln, proyster@unl.edu

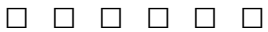
Follow this and additional works at: http://digitalcommons.unl.edu/libraryscience

Part of the Library and Information Science Commons

Royster, Paul, "Case Study: DigitalCommons@University of Nebraska–Lincoln" (2009). *Faculty Publications, UNL Libraries.* Paper 264.
http://digitalcommons.unl.edu/libraryscience/264

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Published in: *Starting, strengthening, and managing institutional repositories: A how-to-do-it-manual,* edited by Jonathan Nabe (New York: Neal Schuman Publishers, 2009). Copyright © 2009 Neal Schuman Publishers. Used by permission.



DigitalCommons@University of Nebraska–Lincoln

Paul Royster
University of Nebraska–Lincoln

The institutional repository at University of Nebraska–Lincoln (UNL) began operation in April of 2005. The decision to implement an IR was made by the Dean of Libraries, Dr. Joan Giesecke, in 2004, and after a study of the available platforms, the Library selected the DigitalCommons software developed by Berkeley Electronic Press (bepress) of California, and (at that time) marketed and licensed through ProQuest, Inc., of Ann Arbor, Michigan. The perceived advantages of this system were 1) its out-of-the-box functionality, 2) its single-price arrangement for software, server space, and ongoing support, and 3) its offering by a known vendor with whom the library had an ongoing relationship. Policies, rationale, and mission statements were adopted wholesale from existing models as suggested by the bepress developers.

The IR (http://digitalcommons.unl.edu) was initially stocked with the complete set of UNL Ph.D. dissertations since 1955, which UNL contracted with ProQuest to digitize and mount on the site. These were digitized from microfilm copies held at ProQuest/UMI and mounted as page-image PDF files, with most document files running from 10 to 20 Mb. By contract with ProQuest, full-text access to this content is free for UNL users; non-UNL users have free abstracts and 24-page previews, with electronic copies available for purchase for about $40. Approximately 9,400 of these dissertations went online in April 2004 and were the only content in the IR for the first three months. About 1,400 have been added since, representing new degrees awarded and also extending the coverage back to the 1890s, when the first Ph.D.s were granted at UNL.

The staff consisted of one full-time librarian (this author), holding the title of Coordinator for Scholarly Communications, who began in June 2005 and was hired specifically to manage and recruit content for the new IR. I had no previous library experience, but did have a background of 25 years in scholarly publishing that included editorial, production, and copyright

experience. At that time, I had been on UNL campus about 2½ years, with the English Department and the university press, so I had some faculty contacts and a general sense of the university's organization and its research strengths. The position I filled was created by the retirement of a longtime employee with collection development responsibilities, which were mostly shifted elsewhere. I reported directly to the Dean; I had no staff and no steering committee—just an office, a computer, access to the library's resources, a great deal of independence, and a mandate to discover and do what was necessary to make the IR "successful." Just what "success" meant wasn't clear, but I set a goal for the first year of 1,500 articles—which was based on one article for each member of the UNL faculty.

Many members of the library faculty and staff were extremely helpful, offering materials, contacts, and advice, and helping spread news of the program across campus, especially when they saw an opening that might prove fruitful. Associate Dean Beth McNeil introduced me to Stephen Vantassel and the Internet Center for Wildlife Damage Control, who have since contributed over 4,000 articles; Technical Services Librarian Sue Ann Gardner connected me with Scott Gardner, who has contributed both the *Online Dictionary of Invertebrate Zoology* and an extensive backfile of the *Journal of Parasitology*; and Mary Bolin not only contributed but also uploaded the backfiles of *Library Philosophy and Practice.* For my part, I followed up every lead and never said "No, we can't (or don't) do that."

I began by visiting the websites of UNL's many departments, schools, colleges, programs, centers, and institutes, and by creating "communities" and "series" within the repository to hold the content I aimed to recruit. I organized the repository's collection alphabetically by subject (Agronomy, Biological Sciences, Chemical Engineering, etc.) rather than by name (Institute for, Center for, Department of, etc.) or by organizational structure. Like any institution that has existed for 150 years, UNL has grown by accretion as much as by plan: it has two separate entomology departments, at least three departments of computer science, two centers for materials science, etc. After about six weeks, I had built a large (but mostly empty) repository with hundreds of community and series pages, each appropriately titled, linked to the webpage of the academic unit it was intended for, and marked with an appropriate graphic to help distinguish one from another.

In the intervals, I uploaded my own papers (about a half-dozen) and developed a set of Microsoft Word templates designed to provide improved onscreen display of the manuscripts that faculty would ultimately be depositing. (These templates have had relatively little impact and almost no use, but they were an attempt to address the translation of 8.5″ × 11″ vertical sheets of paper to a horizontal computer screen with a 4:3 aspect ratio.) School was not yet in session, so I called on what few faculty I knew and could find and offered to upload their eligible papers for them. I did this because I needed 1) something to do, and 2) some content to

show as examples when I began to make the rounds of department meetings with a recruitment pitch.

When the fall semester began in August, I began to seek appointments with department chairs and invitations to faculty meetings, to which I went armed with handouts and a PowerPoint presentation on the rationale of IRs, the benefits of online open access, and the ease of self-archiving. The faculty were generally polite and attentive and seemed interested, but the response rate was well under 5% —I got an average of about one inquiry per department, and only about half the inquiries produced any deposited content.

But meanwhile, other avenues were turning out to be much more productive. On a visit to the Nebraska State Museum's Parasitology Lab, I discovered that the director had a large "dictionary of invertebrate zoology" manuscript that had been peer-reviewed and contracted by the University of California Press but subsequently cancelled for financial reasons. He was struggling to publish it online on their own website, but had so far only managed to get part-way through the A's. At my persistent urging he agreed to let me have a go at it, and several days later sent me the text in 99 WordPerfect files, which I concatenated and began to copyedit, design, and typeset. After about 2½ weeks (and roughly (25,000 or 250,000?) clicks), we published it in pdf format as *Online Dictionary of Invertebrate Zoology* (http://digitalcommons.unl.edu/onlinedictinvertzoology/) , and the response was astounding. Co-authors Armand Maggenti and Scott Gardner were well-connected in their field, and the work had long been anticipated. By its second week online it was accounting for 40% of the IR's traffic, and it remains our most-downloaded work. Spending almost 3 weeks full-time on a single document seemed somewhat counterproductive at the time, but the Dean was supportive, and as it turned out, perspicacious. We did get some additional documents out of it, and made for improved searching and downloading as well, by splitting out letters as separate files.

Around the same time, I first met with Stephen Vantassel, webmaster for the Internet Center for Wildlife Damage Control—a multinstitutional (Cornell, Clemson, Utah State, & UNL) project that was engaged in digitizing "all" the conference proceedings in their field going back to the early 1960s. The problem was they lacked a good platform for hosting this content; mounting it on their website required labor-intensive html coding and constant redesigns and revisions of their web pages, and created server space issues. We had the platform, they had the content, and funding to pay for digitizing more; it was a great and lucky match. They would furnish book-length rtf or pdf files, and I would split them into individual articles and post them to the various series created within their IR community. This connection has so far supplied more than 4,000 articles, about 20% of our total content.

Not everything worked out so well. My hopes of harvesting the Computer Science Department's online archive of technical reports and gray literature came to naught when they decided to write a robot program to automatically transfer all existing and new works; the program never got written and their online archive has since disappeared in a website redesign. The Honors College raised my hopes of collecting 800-1,000 undergraduate honors projects annually, but this too has yet to come to pass.

Still, by the middle of the fall semester I was encouraged to find that the monthly downloads of the several hundred open-access articles I had managed to recruit were already exceeding the downloads from the 9,400 toll-access dissertations. Nonetheless, I was dissatisfied with the rate of uptake among the faculty, so toward the end of the first semester I began offering "for a limited time only" to gather, clear permissions, and upload all the eligible articles for anyone who would send me a vita or a publication list. This did improve the response rate from about 5% to around 10%-15%, which was encouraging and did much to keep me more busy.

Toward the end of the spring semester of 2006, I met with a group of professors from the Department of Physics and Astronomy and boldly made them this same offer. One said, "Do you want everything?" "Absolutely ! Everything." I replied. I did not realize then what I have come to know since: 1) that physics has a tradition and wide acceptance of the practice of online archiving, and 2) physicists publish a ton of articles. About ten members of the department sent me their publication lists; the shortest had 150 articles, the longest more than 400. Fortunately, the major physics journal publishers have very favorable archiving policies, so most of these several thousand articles were eligible for posting as publisher's pdfs. But I was still distraught over how to deliver on the promises I had so rashly made.

The Dean, however, came up with a modest work-study budget, and when school began the next August a student was hired. I began gathering the postable physics articlesfor the student to upload, and in about ten weeks we had worked through the entire backlog. So I went back to Physics and recruited some more faculty, who were impressed (and perhaps a bit envious) that their colleagues had all this content online and were getting monthly reports showing how often it was being downloaded.

This second year I began to notice that the numbers of downloads were spiraling—each month's usage was triple that of the same month from the previous year; and this increase continued through the third year. Now, in our fourth year, it seems to have slowed, and we are merely doubling the downloads from 12 months ago.

To date, I would estimate that our student worker has uploaded at least 12,000 of the roughly 20,000 articles in the repository, whoseapplication and efficiency have created another kind of

challenge: I now have to find 30 to 50 postable article a day just to provide enough work. This has spurred me to develop new strategies for recruiting or finding potential content.

- We have mined the scholarly and research publications coming from within the university, such as the *Nebraska Beef Cattle Reports*, *Nebraska Swine Reports*, *Cornhusker Economics*, *Great Plains Research, Nebraska Studies in Language and Literature, Mid-West Quarterly*.
- We have taken advantage of the federal employee exception to the copyright laws—that works authored by US government employees in the performance of their duties are not subject to copyright, making the published pdf's eligible for posting regardless of the publisher's policy. UNL is a land-grant institution with numerous associated USDA centers.
- We have recruited entire publication backlists, whose publishing organizations lacked seviceable hosting platforms.
- We have digitized out-of-copyright materials with special historical relevance to the university, such as works by Roscoe Pound, Louise Pound, Charles Bessey, et al.
- We proactively explore the lists of the publishers who allow posting of their pdf's, and when we find articles by a UNL faculty member, we approach that person by email, saying "We have recently seen your article ... May we have your permission to archive it, and any other eligible ones? And do you have a publication list you could send?"—the positive response rate from this approach is close to 90%, making it by far the most effective recruitment method we have yet found.

In short, where I had managed to add about 2400 articles, the help provided by the work-study students (and the pressure to keep them busy) made it possible to add over 6,000 articles in the second year, and more than 7,000 articles the third year. We have never done any batch uploads—all content has been added "brick by brick."

So much for the story, in abbreviated form, of our growth from nothing into the (currently) third-largest IR in the United States. With that out of the way, let me just try to address some of the issues covered in the previous chapters.

Staffing & Budget

The first year we had a staff of one; but I was full-time, and devoted exclusively to the IR (apart from serving on 2 committees and provide general copyright advice for the campus).
We currently have a staff of 1½ , plus three work-study students working 10 hours/week each during the academic year. Two of the students do scanning and pdf file preparation, the other does uploading. After the federal share of their work-study wages, they cost the library about $15 to $20 per week, each.

The DigitalCommons package is billed as a flat annual fee, based on FTE.The pricing is confidential, but its cost is less than the salary of one entry-level programmer or technician.

Platform

We are enthusiastic proponents of the bepress DigitalCommons package. Its greatest advantage is that it allows us to spend almost 100% of our time in recruiting and developing content. It worked from the moment it was installed (by bepress), and is almost never down. The annual fee includes customer support, maintenance, and unlimited server space; there is zero load on our IT department or web developers. UNL owns the content files and can take them with us if we should come to a parting of ways.

The monthly download reports to authors generated automatically by the DigitalCommons system are an extremely effective re-selling and recruitment device. This provides feedback to authors that they do not get from journal publishers, and many are astonished at the interest their work receives. Reports available to administrators include numbers of downloads, cover page hits, and referring domains and countries—all of which contributes to a sense of worldwide involvement in the program.

Policies, Preservation

Most of our policies were adopted from those suggested by Bepress. We have made some refinements or clarifications, such as:

- Any co-author may authorize an article's deposit; but any co-author may request its removal.
- We do not deposit articles in copyright without permission of an author or of the copyright holder.
- Depositors should have some connection to UNL; but "connection" is interpreted as broadly as possible.
- UNL Libraries is committed to maintaining its archive permanently, and to migrating file formats should need and general practice require.

Marketing (i.e. recruitment)

Self-archiving is a wonderful concept, but does not appeal to most faculty, at least not enough for them to practice it. I essentially abandoned it as a marketing message after about 3 months.

Some few authors have made sporadic attempts at archiving their own papers, and I encourage those who want to try, but I also find that checking (and often correcting) the self-archived articles is more time-consuming than gathering and preparing them myself, and giving them to a student to upload.

I began with the assumption that younger faculty would be the most likely audience and most enthusiastic users, but the reverse has turned out to be the case—older faculty with long publication lists are better candidates than younger ones, and the emeriti faculty often turn out to be the best of all.

A wise strategy is to find partners who have large amounts of digital or digitizable content.Without sufficient staff and equipment, it is advisable to become the host, not the digitizer. The Lester Larsen Tractor Test and Power Museum at UNL had a roomful of file cabinets stuffed with technical tractor test reports going back to 1920. They acquired a desktop scanner and used museum volunteers to digitize and upload the entire set, over 2,250 reports. This has now become our largest and most popular series (http://digitalcommons.unl.edu/tractormuseumlit/).

Finally (and despite the cliché), it is necessary to think outside the box. It is important not to stop at the campus border—much of our content comes from places other than traditional academic departments: museums, research institutes, USDA-affiliated centers, extension offices, etc. We have never turned down anyone who was interested in participating. Moreover, it is counterproductive to be afraid to "push the envelope"—to take a reasonable chance or a calculated risk: in the online environment, something can always be easily and quickly taken down if proves a problem (a luxury that print publishers don't have).

Collection development

Organization of the repository matters more to depositors than to users, most of whom (about 75%) reach their destination via Google searches. In some areas we have set up series specific to individual authors; in others we have a more minimal division of 1) faculty publications, 2) dissertations and student research, and 3) department materials (newsletters, program reviews, etc.). A series with only one or two papers can look rather sad and lonely, but one with several thousand papers can be unwieldy, both in presentation and for technical management.

Usage

I think the best way to build usage is to assemble a large collection. Online content has a gravity-like property—its attraction increases exponentially with its mass. Of course, the best

articles are not necessarily the ones that attract the most usage. Further, usage has very little to do with recentness, and everything to do with relevance to common searches. Intriguingly, online bibliographies and open-access dissertations seem to be far more popular than their print counterparts.

Some ways we have found to build usage include cataloging book-length works in our online catalogue (which also places them in WorldCat), listing them with the Online Books Page (http://onlinebooks.library.upenn.edu/), placing links in relevant Wikipedia articles (but see http://en.wikipedia.org/wiki/Wikipedia:External_links), and listing or linking them in appropriate subject-based websites.

Overall, I think usage is facilitated by posting articles with full-length and keyword-rich abstracts: it helps search engines find them and encourages finders to go ahead and download the article.

My best advice is "Don't wait." Start gathering and posting something as soon as possible. I try not to let a day go by without posting *something*. One thing will lead to another, and that to another, and so on. Waiting for just the right time or just the right content seems pointless; upload what you can get, and go on to the next thing—the browsing public on the internet will decide what it wants, and out of those billion people, the odds are most things will be of interest to someone.

Further Reading

Royster, Paul. "Publishing Original Content in an Institutional Repository," *Serials Review* 34:1 (2008), 27–30; special issue on "Open Access Revisited." doi:10.1016/j.serrev.2007.12.002 Open-access preprint at http://digitalcommons.unl.edu/libraryscience/126/

Royster, Paul. "The Institutional Repository at the University of Nebraska–Lincoln: Its First Year of Operations," *OCLC Systems & Services: International Digital Library Perspectives* 23:2 (Spring 2007), 183-189. http://digitalcommons.unl.edu/libraryscience/58/

Royster, Paul. "Value Added Services: Publishing Continuum at the University of Nebraska–Lincoln" (with Joan Giesecke), SPARC Digital Repositories Conference, Renaissance Harborplace, Baltimore, MD, November 17, 2008. http://digitalcommons.unl.edu/library_talks/46/ or http://www.arl.org/sparc/meetings/ir08/valueadded.shtml (with audio)

Royster, Paul. "How to Fill Your Institutional Repository, or, Practical Lessons I Learned by Doing," American Library Association Convention, Anaheim, California, June 30, 2008. http://digitalcommons.unl.edu/library_talks/40/

Royster, Paul. "Research Showcase: UNL's Institutional Repository," American Library Association Convention, Anaheim, California, June 30, 2008. http://digitalcommons.unl.edu/library_talks/41/