

November 2007

Publishing Original Content in an Institutional Repository

Paul Royster

University of Nebraska-Lincoln, proyster@unl.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/librarianscience>



Part of the [Library and Information Science Commons](#)

Royster, Paul, "Publishing Original Content in an Institutional Repository" (2007). *Faculty Publications, UNL Libraries*. 126.
<http://digitalcommons.unl.edu/librarianscience/126>

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.

Publishing Original Content in an Institutional Repository

Paul Royster
University of Nebraska–Lincoln

Institutional repositories (IRs) have been developed and promoted primarily as a means to re-publish scholarly content previously published elsewhere—usually in journals, *festschriften*, or collections of articles. This essay discusses the use of IRs as the originating publisher of materials not previously published elsewhere, and assesses their potential use as a viable “first resort” for scholarly publication and an already existing alternative to traditional commercial or university presses.



“Freedom of the press is guaranteed only to those who own one.”

—A. J. Liebling

One of the great surprises in managing the infant but growing IR at the University of Nebraska–Lincoln¹ has been the overwhelming popularity of a number of works that had been or logically would be deemed unsuitable for ordinary (i.e. paper) publication. While such content makes up only a tiny percentage of the repository’s content, its usage is far out of proportion to that of the usual and “koshered” or peer-reviewed articles the IR was designed to archive, preserve, and render accessible. A review of any month’s “most downloaded documents” will usually show that half or more of the top ten most-downloaded works are documents published originally in the IR and unavailable anywhere else.

¹ <http://digitalcommons.unl.edu/>

This suggests a role for the IRs beyond that of archival storage and accessibility enhancement: in fact, they are well suited to become online publishers giving voice to a wide range of authors normally excluded, put off, or ill-served by the vagaries, idiosyncrasies, delays, obligations, and hoops-jumping of the conventional publication routes. To illustrate the possibilities, we discuss four examples: a large reference work, an edited document series, an undergraduate project, and a set of doctoral dissertations.

I. A large reference work: Soon after the university's IR went live in July 2005, its manager (the author) was touring the state museum's parasitology lab in one of many campus visits to try to recruit IR content. There, among rows and racks and cases and drawers of worms in jars of alcohol, was a typescript about 18 inches tall. Upon inquiry, it turned out to be a dictionary of invertebrate zoology compiled over a 15-year period by the lab director, along with a (retired) colleague from the University of California and that colleague's (now-deceased) wife. Of course, there was a sad story attached to it: it had been submitted to a large western university press, peer-reviewed, accepted, revised, and put under contract; then, without warning, and apparently the very week it was to be shipped off to the publisher, the contract was cancelled, as the press had decided (after several loss-producing years) to get out of the area of zoology publishing. Some other presses had been approached, but one already had a competing product, and others were deterred by the length of the manuscript and the limited commercial appeal. The lab director mentioned they were "thinking of maybe publishing it online" on their lab's homepage, but had so far only got a quarter of the way through the A's. He was entreated to give the IR manager a shot at it, and a few days later the IR received 98 WordPerfect files. These were translated to MS Word and concatenated, and the process of copy-editing and formatting began. About 2 weeks and 250,000 clicks later, a book emerged which was electronically published in the IR September 6, 2005, as *The Online Dictionary of Invertebrate Zoology*.² It was almost 1,000 half-letter pages long and weighed in at nearly 10 Mbytes. It was a pdf file, translated from the MS Word document by the IR's automatic pdf generating filter.

² <http://digitalcommons.unl.edu/onlinedictinvertzoology/>

Almost overnight, the work became responsible for a third of the IR's downloads. It turned out that the authors were well-connected in the field (socially and electronically), and that the work had been anticipated anxiously for quite some time. It was downloaded nearly 1,200 times in the first month online, and has continued to be downloaded around 1,000 times per month ever since. About six weeks after its initial publication, a slicker version was prepared with Adobe InDesign and pdf'ed with Adobe Acrobat, reducing the file size to 3.4 Mbytes. To date (27 months out), there have been nearly 30,000 downloads.

II. An edited document series: The author of this article came to IR management with an interval of 20 years between his graduate degree (in English) and his return to a "faculty" position—most of that time was spent in scholarly publishing in a variety of production, editorial, and management roles. Of course, some of the first content to be put in the IR was his old scholarly articles, mainly biographical dictionary entries published in the 1980s on obscure figures in early American literature. Naturally these did not attract much usage, partly because the works for which these figures were known were long (i.e., hundreds of years) out of print and completely unavailable. So it seemed to him at the time that bringing those works back into print electronically might help create more interest in the biographical essays about their authors. This scheme failed miserably, but it did begin a project, which turned into a series (now numbering 65), of "Electronic Texts in American Studies" that republishes new authoritative editions of original works from the sixteenth, seventeenth, eighteen, and nineteenth centuries, and these have been surprisingly well-received.³

It has long been a handicap to the study of early American literature that the texts are largely unavailable. Anthologies include a few samples (usually abridged) from a handful of sources, many of which **were not even published at the time** (e.g., Winthrop's *A Modell of Christian Charity*, Bradford's *Of Plymouth Plantation*, Edward Taylor's *Preparatory Meditations*, Samuel Sewall's *Diary*, etc.). Some works were republished in various proceedings series scattered throughout the nineteenth century, but most have been accessible only in microfilm, microfiche, microprint, and more recently electronic digitized format (as page images)

³ <http://digitalcommons.unl.edu/etas/>

from Early English Books Online (EBO), Eighteenth-Century Collections Online (ECCO), Evans Early American Imprints, etc. These are sometimes defective to the point of illegibility, access is limited to the largest research institutions, and the series are circumscribed by dates of publication or country of origin. As commercially published collections, they are available only to institutional subscribers; as page images, they are not searchable or excerptable—and yet the underlying works have long, long since passed into the public domain. Conventional publishers have no interest in this material—it is not copyrightable, has a limited audience, and (for the most part) unproven appeal.

It has become the author's "scholarly output" to create and publish (online in the IR) reliable editions of early American texts that are generally unavailable. To date (about 24 months since the publication of the first text), more than 36,000 have been downloaded. The most popular, by a large margin, has been Benjamin Franklin's 1734 edition of James Anderson's 1723 work *The Constitutions of the Free-Masons*, the first Masonic book printed in America.⁴ Also popular are Cotton Mather's *The Wonders of the Invisible World* (1693),⁵ John Smith's *A Description of New-England* (1616),⁶ Increase Mather's *A Brief History of the Warr with the Indians in New-England* (1676),⁷ and *David Cusick's Sketches of Ancient History of the Six Nations* (1828).⁸ Part of the aim is to exhibit the variety of subjects and treatments of the literature of this era, in contradiction to the general impression of uniformity and consensus that has often prevailed. Works in the series discuss Masons, magic, witches, Indian wars and captivity, extinct mammals (Steller's sea cow), astronomy, sodomy, sexual perfectionism, transvestitism, slavery and abolition, persecutions, and massacres.

These works are made available in free open-access electronic format for the use of researchers, teachers, students, and other interested parties. They give instructors a much wider selection of texts than was previously available, and they offer an opportunity for a broad-based reassessment of a field that has long been underserved by the publishing industry.

⁴ <http://digitalcommons.unl.edu/libraryscience/25/>

⁵ <http://digitalcommons.unl.edu/etas/19/>

⁶ <http://digitalcommons.unl.edu/etas/4/>

⁷ <http://digitalcommons.unl.edu/libraryscience/31/>

⁸ <http://digitalcommons.unl.edu/libraryscience/24/>

III. An undergraduate project: In the summer of 1898, The Trans-Mississippi and International Exposition was held in Omaha, Nebraska, and music was played daily at a variety of venues within the fairgrounds. By studying programs and newspapers from the period, a UNL undergraduate in the Music Department was able to reconstruct most of the musical programs presented each day over the five-month period (June 1 through October 31) that the Exposition was in session. She developed an interactive website that showed who played what where on each day. Having no place to permanently host the site, she was referred to the IR by her professor; and her html files were combined into an illustrated, searchable, text-based pdf document that was posted in the IR in a series developed for student research.⁹ While the “flat” (106-page) pdf version sits in the IR, it is also attached to an archived downloadable “zipped” html version that reconstitutes the original interactive website. (The Digital Commons system allows for the attachment of any number of “related” or supplemental files in any file format.) To date (15 months after initial publication) the work has been downloaded 953 times, an average of twice daily over that period. The work would not really have been publishable in any other format, and would have remained stored on a CD in a box or on a shelf, unavailable to the world.

IV. Doctoral dissertations. Publishing doctoral dissertations in an IR is not unusual; in fact, they make up the largest single document type in a recent survey of IRs in the United States.¹⁰ But it is especially worth noting the great difference that is made by making them open-access as opposed to limited-access or subscription-access. The experience at our IR has shown this difference to be about a 60-fold increase in downloads (and thus in impact) of open-access as opposed to limited access dissertations.

The UNL IR was seeded with over 9,500 electronic theses and dissertations when it began. These were digitized by ProQuest/UMI from their microfilm collections. The university paid a fee for perpetual on-

⁹ <http://digitalcommons.unl.edu/musicstudent/4/>

¹⁰ C. S. McDowell, “Evaluating Institutional Repository Deployment in American Academe Since Early 2005,” *D-Lib Magazine* 13:9–10 (September/October 2007); doi:10.1045/september2007-mcdowell.
<http://www.dlib.org/dlib/september07/mcdowell/09mcdowell.html>

campus electronic access to the digitized ETDs, and ProQuest/UMI retained the rights to distribute them for sale in all other formats and venues. Until ProQuest returned the administration of the Digital Commons to BEPress in the summer of 2007, our usage reports showed the number of downloads for these subscription-based dissertations (free on-campus, for-pay off-campus). Downloads of the most recent ETDs averaged 7.20 copies per year (24,770 downloads of 1,527 dissertations over a 27 month period). The ratio of actual downloads to cover-page “hits” of all dissertations was 16.0% (61,507 downloads out of 383,473 hits; there were 9,391 ETDs in April 2005 and 10,446 ETDs in June 2007). In other words, 5 out of 6 browsers who saw the title and abstract of an ETD decided **not** to download it (at a cost to off-campus users of around \$30). [For the OA portion of the IR, the ratio of downloads to hits ranges between 70% and 75%.]

In 2006, the university began to require the electronic deposit of dissertations (thereby relieving the library of binding and storage costs), but there were concerns in the Graduate School that all Ph.D. candidates might not be able to generate an appropriate pdf file. The IR manager (this author) offered to provide assistance to those needing pdf help, and found that about 10% of the roughly 200 new Ph.D.’s needed help converting their files into pdf’s (these tended to be older [i.e., 35+] students in music, modern languages, and English). The ProQuest/UMI dissertation deposit agreement is non-exclusive; i.e., the candidate retains the right to publish or post the dissertation elsewhere in any format. So each time the IR manager finished helping the new Ph.D. generate a proper pdf file, he also evangelized the candidate on the benefits of electronic publishing in the open-access side of the IR (immediate publication, Google searching, permanent linkable URL, worldwide 24/7 access, etc.). Most candidates agreed to post their dissertations in the open-access IR (probably out of gratitude more than anything else). This has produced a sample of 66 ETDs that are completely open-access and published in the IR.

These open-access dissertations have been downloaded like gangbusters: they account for half of the top “most-downloaded” works each month; and their popularity puts the faculty-authored, peer-reviewed, journal-published articles entirely in the shade. Their average number of downloads per year has been 424 (25,658 downloads over 726 total months of availability). This compares to an average of 7.2 per year

for the non-open-access dissertations, an increase of 5889%, or a factor of almost 60. Somewhat surprisingly, the most popular (downloaded 3,759 times over 19 months) is entirely in Spanish: *Poética de lo Soez: Luis Rafael Sánchez: Identidad y Cultura en América Latina y en el Caribe*, by Julio César Sánchez Rondón.¹¹ (And they said literary criticism was dead!)

Even compared to the other open-access materials, the dissertations are more popular: the open-access portion of the repository averages between 3 and 5 downloads per item per month; the OA dissertations' average exceeds 35.

Outreach

"If a tree falls in a forest and no one is around to hear it, does it make a sound?"

—Philosophical riddle (unknown origin).

Clearly, there is more to publishing than simply issuing a work, whether in paper *or* electronic form. If potential users remain unaware of a work's availability, then the falling tree hasn't made a sound.

Some efforts have been made to alert possible users to the existence of these resources, mostly in the form of **online links**. The most effective at generating traffic has been The Online Books Page¹²; a listing there usually triples the daily downloads almost immediately. Wikipedia¹³ is another useful and popular site, although users should consult and abide by their external links guidelines.¹⁴ Other resources that have proved useful are the MLA International Bibliography,¹⁵ and Intute.¹⁶ Discipline- or area-specific web pages are also useful, and recognition and dissemination of our American studies texts has been helped by the Society of Early Americanists resource pages.¹⁷ There is, however, a wide variety of

¹¹ <http://digitalcommons.unl.edu/modlangdiss/1/>

¹² <http://onlinebooks.library.upenn.edu/>

¹³ <http://en.wikipedia.org>

¹⁴ http://en.wikipedia.org/wiki/Wikipedia:External_links

¹⁵ <http://www.mla.org/bibliography>

¹⁶ <http://www.intute.ac.uk/>

¹⁷ <http://www.societyofearlyamericanists.org/links.html>

policies and updating practices among online resource sites; many will only list texts that they themselves host and often insist on html file format; others proclaim themselves to be the latest thing, but haven't updated since 2001.

Cataloging original works (by our library's Technical Services division) creates MARC records for our own online catalogue,¹⁸ and these are harvested and included in WorldCat, where they appear as "Online resources."¹⁹

Search engine optimization (SEO) is another key element in outreach. It is important to get the subject into the title and to utilize the abstract metadata to its fullest. In the case of the *Online Dictionary of Invertebrate Zoology*, we created additional files for each separate letter, and loaded the defined terms (minus their definitions) into the abstracts, so that someone searching in Google for "androsynhesmia" (=a group of males gathered together during mating season) will find the ODIZ at the top of their results.

Respectability

Online electronic publication does suffer from a relative lack of respectability. It is sometimes difficult to achieve the type of recognition that the material merits. Even our own Graduate School Office of Research declines to include electronically published book-length works in their annual survey and report of university scholarship. Among the faculty in general, online publication seems to be regarded as only about a half-step up from blogging.

The only answer for this may be that time, the increase of online scholarly materials, and the changing of generations will ultimately turn the tide. It is significant that Grisha Perelman was recently awarded the Fields Medal (the highest honor in mathematics) for his solution of the Poincaré conjecture—work that was published only in an online open-access repository, arXiv.²⁰

¹⁸ <http://iris.unl.edu/>

¹⁹ <http://www.worldcat.org/>

²⁰ <http://arXiv.org/>; the papers are at <http://arxiv.org/abs/math/0307245>, <http://arxiv.org/abs/math/0303109>, and <http://arxiv.org/abs/math/0211159>.

L' Envoi

It should not be inferred that the activities described here represent (in the author's opinion) the last word or the cutting edge of electronic publishing or digital scholarship. Indeed, he feels much more like an old man driving down the "information highway" at 35 miles an hour with his blinker on. The works published are "flat" pdf files, containing at most an occasional hyperlink, not interactive or hypertextual. The work is not grant supported, does not develop new digital tools, or analyze, subvert, or revolutionize the craft and methods of either traditional or electronic scholarship.

What is potentially revolutionary, however, is that it makes publication (or access to the means of scholarly production) widely available to all faculty, a working class who are surely in need of new relations of production. What other profession requires its members to give away the products of their labor (their copyrights) for the sake of retaining their jobs? Oh, but they receive royalties!—Some may, but most do not; and few receive anything more than a token pittance, except for the occasional widely-adopted introductory textbook.

What of the publishers' traditional defense that they provide peer-review and serve a vital "gate-keeping" function? Peer-review is provided not by the publishers but by the faculty. As for gate-keeping, what is being kept out (or made to stand in long lines awaiting admission) are those works that don't fit the income needs of conventional publishers. The actual direct costs of developing a book-length monograph (most of it done by free-lancers and outside services) are in the range of \$5,000; but it must support a publisher's overhead costs of \$20,000 or more, and so needs to generate sales income of \$50,000 or more just to approach breaking even. In such an environment, publishers cannot afford to publish in many areas, and scholarship gives way to "publishing decisions" that have everything to do with the self-preservation of the publisher, not the faculty. Publishing—even scholarly publishing—is a book-selling business; and scholarship is packaged into salable units as defined by market forces.

But what if authors could just publish for themselves? What if every faculty member had his own press, and the freedom to use it as he or she saw fit? Would the result be anarchy, as traditionalists predict? Would there be riots in the quad if John Wiley & Sons (say) didn't decide what

gets published? How would libraries decide which (free) resources to collect and preserve? How would tenure committees know who to reappoint? These are questions too deep or too broad to be answered by an IR administrator at a midwestern college best known for its football program. Nonetheless, he feels that open-access online publishing of faculty scholarship has potential that has only begun to be realized, and that the scholarly world of 2015 or 2020 will look very different from that of today.

