Institutional Repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace

Philip M. Davis pmd8@cornell.edu (corresponding author)
Matthew J. L. Connolly mjc12@cornell.edu
Final Draft. Last Revised Jan 08, 2007
Forthcoming, *D-Lib Magazine*, Volume 13, Issue 3/4, 2007
http://www.dlib.org/dlib/march07/davis/03davis.html

Abstract

Problem: While there has been considerable attention dedicated to the development and implementation of institutional repositories, there has been little done to evaluate them, especially with regards to faculty participation.

Purpose: This article reports on a three-part evaluative study of institutional repositories. We describe the contents and participation in Cornell's DSpace and compare these results with seven university DSpace installations. Through in-depth interviews with eleven faculty members in the sciences, social sciences and humanities, we explore their attitudes, motivations, and behaviors for non-participation in institutional repositories.

Results: Cornell's DSpace is largely underpopulated and underused by its faculty. Many of its collections are empty, and most collections contain few items. Those collections that experience steady growth are collections in which the university has made an administrative investment, such are requiring deposits of theses and dissertations into DSpace. Cornell faculty have little knowledge of and little motivation to use DSpace. Many faculty use alternatives to institutional repositories, such as their personal Web pages and disciplinary repositories, which are perceived to have higher community salience than one's affiliate institution. Faculty gave many reasons for not using repositories: redundancy with other modes of disseminating information, the learning curve, confusion with copyright, fear of plagiarism and having one's work scooped, associating one's work with inconsistent quality, and whether posting a manuscript constitutes "publishing".

Conclusion: While some librarians perceive a crisis in scholarly communication as a crisis in access to the literature, Cornell faculty perceive this essentially as a non-issue. Each discipline has a normative culture, largely defined by their reward system and traditions. If the goal of institutional repositories is to capture and preserve the scholarship of one's faculty, institutional repositories will need to address this cultural diversity.

Introduction: Building the case for Institutional Repositories

The digital revolution has affected how scholars create, communicate and preserve new knowledge. While the technologies exist for scholars to manage their own digital content, faculty are typically best at creating, not preserving, new knowledge [1]. As a consequence, most faculty host their digital objects on a personal website, where their long-term preservation is not secure. If institutions truly value the content created by their faculty, they must take some responsibility for the long-term curation of this content.

Clifford Lynch, Director of the Coalition for Networked Information, defines an institutional repository as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution." [1]

Two philosophical camps

There are two opposing philosophical camps among those who work to justify institutional repositories: one that views IRs as *competition* for traditional publishing, the other that sees IRs as a *supplement* to traditional publishing.

In 1994, Stevan Harnad wrote his Subversive Proposal for Electronic Publishing, in which he argued that all academics should make their research articles publicly available through open repositories [2]. This collective effort would help to reduce the power wielded by publishers who have built economic barriers to limit scholars' access to the literature.

Similarly, Raym Crow, writing the position paper for the American Research Libraries, argues that increasing access to the literature is but one goal of institutional repositories [3]. He posits that, by taking at least some control over the dissemination of scholarship, repositories can increase competition in the marketplace and reduce the monopoly power of journals. Crow believes that there is no reason that institutional repositories cannot provide all of the functions of traditional publishing (registration, certification, dissemination, and archiving), in effect taking the role of scholarly publishing out of the hands of third-party publishers and placing it back in the hands of the academy.

In opposition, Clifford Lynch views IRs as *supplements*, not primary venues for scholarly publishing, and warns against assuming the role of certification in the scholarly publishing process. He argues that "the institutional repository isn't a journal, or a collection of journals, and should not be managed like one" [1]. Lynch fears that viewing IRs as instruments for undermining the economics of the current publishing system discounts their importance and reduces their ability to promote a broader spectrum of scholarly communication. Institutional repositories may better serve to disseminate the

so-called "grey literature": documents such as pamphlets, bulletins, visual conference presentations, and other materials that are typically ignored by traditional publishers [4].

DSpace was not conceived as competition to commercial publishers, but as a resource to capture, preserve and communicate the diversity of intellectual output of an institution's faculty and researchers [5-9]. It was designed specifically to deal with a wide range of content types including research articles, grey literature, theses, cultural materials, scientific datasets, institutional records, and educational materials, among others [7].

Evaluating the Success of Repositories

There have been several previous studies that attempted to measure the participatory use of repositories:

Surveys of Scholars

A 2001 survey of scholars randomly chosen from nine scientific disciplines from colleges and universities in the United States and Canada sought to determine faculty participation in depositing e-prints into digital archives [10]. Physics and astronomers reported the highest participation, followed by mathematicians and computer scientists, engineers, cognitive scientists and psychologists, and biological scientists. Chemists reported no participation. Those who reported participation cited the dissemination of research results, visibility, and the author's exposure as reasons for depositing their work. Reasons for non-participation included publisher policies, relevance to their field, and technological constraints. A large, international survey of senior authors in 2005 demonstrated a general low level of knowledge and motivation to use institutional repositories [11]

Environmental Scan

In 2003, an environmental study of 45 institutional repositories running the EPrints software supported widely held speculations about the success of institutional repositories [12]. Because of low faculty participation, IRs have only been able to collect a small fraction of an institute's research output. Grey literature and theses make up the majority of content in these systems and the number of new records added to the repositories falls dramatically after the first couple of months. In addition, IRs have replicated the subject bias found in disciplinary repositories and have failed to attract contributions from medicine and the clinical sciences, chemistry and law. Mark Ware, the author of the study, argued that there was little evidence that IRs are leading a reform in scholarly publishing; and, given their patchy coverage, it was not clear how IRs were adding to a long-term preservation agenda. In a 2006 report, Ware concludes, "At present it appears that the large majority of authors are either ignorant of or indifferent to the potential benefits of self-archiving." (p.24)[13]

Participation in PubMedCentral

On May 1st, 2005, a policy was enacted which recommended, not required, that all researchers receiving grant monies from the National Institutes of Heath deposit final copies of their manuscripts in PubMed Central (PMC), a free digital archive of biomedical and life sciences journal literature. PMC offers many valuable services to authors, such as indexing in Medline (the primary literature index for the biomedical and life sciences), as well as dynamic links to the published version of their article. After eight months, the participation rate remained a dismal 3.8% [14]. Lack of awareness of the policy was not cited as contributing to the low compliance rate. On December 14th, 2005, Senator Joseph Lieberman introduced the CURES Act (S.2104), which would require (not recommend) mandatory deposit of final manuscripts within the first six months after publication. This bill has been referred to a special committee.

Survey of Institutions

The success of institutional repositories has been somewhat spotty. In a 2005 survey of university and liberal arts colleges in the United States, 40% of universities and 6% of colleges had operational IRs. Of those that didn't, 88% of the universities and 21% of the colleges were planning to participate in a consortial IR system [15]. DSpace was the dominant content management package listed by the respondents. The reported sizes of these IRs ranged from hundreds of thousands of objects (over 10 terabytes of space) to a few dozen (less than one gigabyte), although there was confusion from the respondents on what constituted an "object." Some respondents may have counted a database as a single object, or each record in the database as an object. The formats of materials stored in these repositories was diverse, including e-prints, electronic theses and dissertations, digitized special collections, multi-media, course materials, and datasets. Participation from an institution's faculty was, in all cases, a voluntary prospect and generally perceived to be very low.

A similar survey in 2005 was undertaken at universities in ten European countries – Belgium, France, the United Kingdom, Denmark, Norway, Sweden, Finland, Germany, Italy and the Netherlands – as well as in Canada and Australia [16]. The number of institutional repositories varies from as low as 1.5% of universities (a single instance) in Finland to as high as 100% in Germany, Norway and the Netherlands. The focus on acquisition of content in IRs was almost exclusively (with the exception of Australia and the U.S.) on collecting faculty publications.

Like the U.S. study, the European survey also reported low faculty participation in storing objects in their IRs. In their article, Van Westrienen and Lynch [16] identified several reasons for non-participation from faculty, including:

- Difficulties informing faculty and convincing them to participate
- Confusion and uncertainty about intellectual property issues
- Scholarly credit and how the material in IRs would be used
- The perception of Open Access content being of low quality, and
- A lack of mandatory policies for depositing manuscripts.

Faculty Reward Systems

Understanding the academic values of faculty and their reward system is essential for evaluating institutional repositories and predicting their future success. A Mellon-sponsored study of scholarly communication within the rubric of academic values was recently undertaken at the University of California, Berkeley [17]. While the researchers reported that scholars were generally in support of the idea of making knowledge available for the public, scholars were fundamentally more concerned with issues of advancement and stature in their own scholarly field such as peer review. The authors concluded that "approaches that try to 'move' faculty and deeply embedded value systems toward new forms of archival, 'final' publication are destined largely to failure in the short term" [17].

Purpose

While much emphasis has been placed on building institutional repositories, there has been little work on evaluating their outputs. Understanding the reasons for non-participation from an institution's faculty can assist developers and implementers of repositories in making enhancements to the software, developing an educational outreach program to encourage future use, or incorporating faculty submissions as part of the publication process. The purpose of this study will attempt to answer these questions:

1. How has DSpace been adopted at Cornell University? For what purposes are individuals and communities using DSpace?

2. What are the reasons that deter or discourage faculty from using their institutional repositories?

Methodology

Data collected from Cornell's DSpace interface, along with Web server log files detailing visits to DSpace, were processed and analyzed in order to calculate descriptive statistics for the repository (number of objects, types of objects, number of communities and collections, participation rates, and metadata view counts). To preserve user anonymity, log file events were grouped by IP address and the IP address itself was dropped from the data set.

To provide a basis for comparison with other instances of DSpace, permission was obtained to perform the same type of automated collection from seven other institutions that use DSpace as a primary IR. Server logs, which contain much more sensitive data, were not requested from other institutions. The data collected was the same as for Cornell with the exception of the metadata views count, which is a custom Cornell addition.

In order to better understand the attitudes, motivations and behaviors behind non-participation, eleven semi-structured interviews with selected faculty members across different Cornell departments were conducted between September and October 2006.

Taken together, this three-part methodology attempts to provide context and meaning to the results of the Cornell DSpace study.

Results

Cornell DSpace

DSpace is organized into *communities*, a high-level organizational structure whose only purpose is to divide collections into related groups. Each community contains one or more *collections*, which are containers for related *items*. An item is a deposited object of any type: a published article, an image, audio, or video file, notes, a presentation, etc

As of late October, 2006, Cornell DSpace held 2,646 items, which were organized into 196 collections within 193 communities and sub-communities. Cornell's DSpace was implemented as a university-wide structure and launched with an established skeleton of communities and collections already in place. At the time of this study, 57 (29%) of these collections remain empty.

Figure 1 shows the distribution of collection sizes for those collections with at least one item. Almost 80% of collections contain fewer than 10 items; less than 5% contain 100 or more. The largest 10 collections include five image collections from the Cornell Plantations, closed theses and dissertations, an archived engineering newsletter, a collection of senior seminars, and articles and presentations by Cornell Library staff. This is not to say that DSpace is not active. Figure 2 shows the growth of DSpace as a whole over time, from its inception in 2002 through October 2006. The sharpest period of growth, though, visible in the first half of 2005, is primarily due to a one-time load of large image collections.

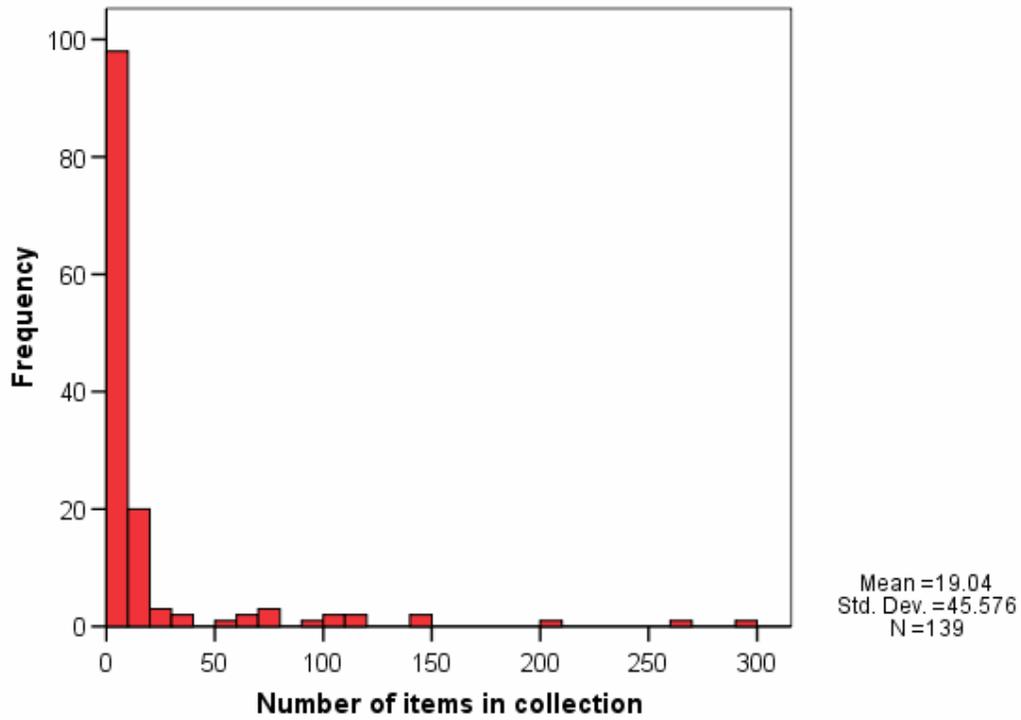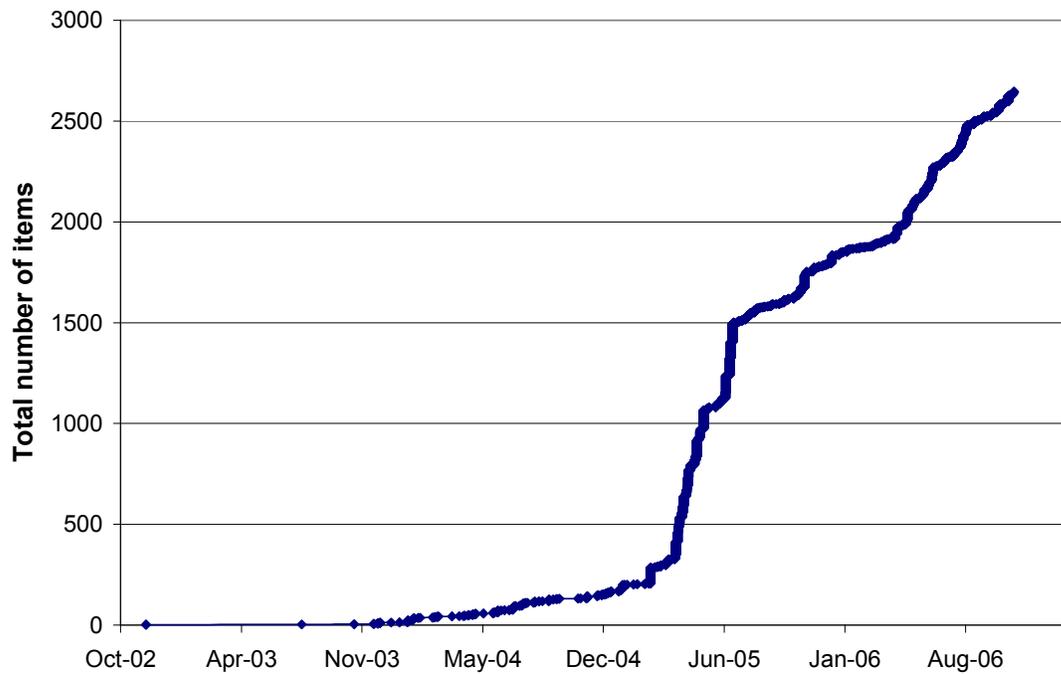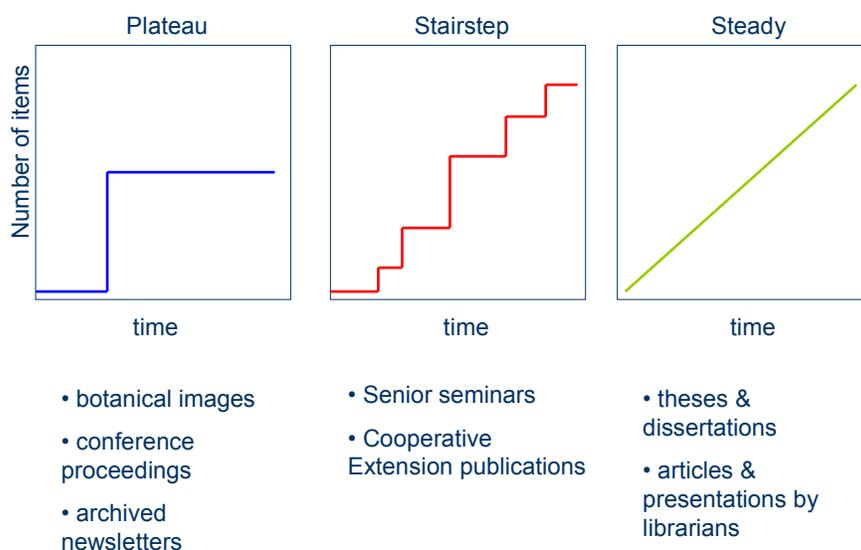**Figure 1. Histogram of Items in Collections**

Figure 2. Growth of Cornell DSpace as a Function of Time

To gain a better understanding of how individual collections were developing, a plot of number of items (population) as a function of time was created for each non-zero collection in DSpace. Each plot was then manually sorted into one of three growth patterns (plateau, stairstep, steady) or "uncatagorizable" (see Figure 3).

Of the non-zero collections, the majority (77% or 107 collections) fell into the plateau category, 19% (26) exhibited the stairstep pattern, and only 3% (4) demonstrated sustained, steady growth. These 4 collections were Theses and Dissertations (open and closed), Cornell Library staff articles and presentations, and Multimedia and Videos.

Figure 3. Growth Patterns for DSpace Collections

Plateau	Stairstep	Steady
• botanical images	• Senior seminars	• theses & dissertations
• conference proceedings	• Cooperative Extension publications	• articles & presentations by librarians
• archived newsletters		

Extremely popular items in collection

Metadata downloads for each item were regressed with the days they spent in the repository to see if a general relationship between downloads and time could be established. From this analysis, six outliers, representing unusually high downloads were identified and analyzed.

Two of these extremely popular items were videos, one a biography of a famous Nobel Laureate in Physics and the other a laboratory demonstration of nonlinear dynamics used in a first course in mathematics. Two other widely popular items were scanned images of classic books no longer in print. These books are used in college classes and linked to from organizations and society websites. The last two objects were reports produced by Cornell University librarians, one a controversial report on the economics of Open Access publishing and the other a (similarly controversial) report on the changing nature of the library catalog commissioned by the Library of Congress.

Participation in Depositing Items

By analyzing the data from the DSpace log files, we derived a picture of how widespread contributions to DSpace were over the covered time period (15 months). During that time, items were deposited into the repository from 519 unique IP addresses.

Of these, 50% deposited only a single item, and only 32 IPs were responsible for 10 or more items each. The greatest number of items deposited from a single IP was 82. It is important to note that an individual working in a department at Cornell might submit items from several IP addresses depending on how the network is configured, and indeed, how workstations are shared within an office. Likewise, a faculty member working from home may connect to the Internet through a dynamically assigned (i.e., variable) IP address, making it appear that several individuals are contributing.

These data reveal much about the way that DSpace is being used at Cornell. Although a university-wide structure exists, much of it remains in skeletal form, with many collections empty or meagerly populated. The DSpace repository as a whole is enjoying steady growth, with approximately 1,000 items added over the past year; however, only a small number of collections display a steady growth pattern. Instead, it appears that most collections are being used to build archival collections as either one-time deposits or periodic batch additions of material. The collections that do exhibit steady growth are largely supported by active policies or guidelines that dictate that items *will* be deposited into DSpace, such as the case of theses and dissertations.

There is little evidence to suggest that individual faculty are making significant contributions of regular scholarly output to the repository. Although the breakdown of submissions by IP address is not conclusive, it is echoed by the growth patterns exhibited by the majority of collections.

Comparison with other DSpace implementations

Seven additional institutions were studied in an attempt to draw comparisons with Cornell's IR. The age of their DSpace installations ranged from 8 to 35 months, with item counts from 500 to 32,676. Table 1 summarizes statistics on community and collection structure across the seven schools. These numbers reveal that there is a wide spectrum of DSpace structures employed, with Cornell falling midrange for most values. DSpace at Institution #6, for example, contains a relatively low number of items in a large number of collections and communities. This is an institution where a university-wide DSpace skeleton has been put in place, as at Cornell. The price paid for this structure, however, is a disproportionately high percentage of empty collections: 58% percent! At the other extreme is Institution #1, which contains no communities and very few collections, but an unusually large number of deposited items.

Table 1. Cornell DSpace compared to other implementations

Institution	No. Items	No. Communities	No. Collections	% Empty Collections
Cornell	2,646	193	196	29
#1	32,676	0	18	0
#2	500	6	10	0
#3	1,438	13	32	19
#4	3,111	24	65	45
#5	801	70	140	41
#6	1,542	390	282	58
#7	1,535	49	81	34

The division of growth patterns seen in Cornell's DSpace is generally repeated at other institutions (Table 2), as the majority of collections fall into either the Plateau or Stairstep categories. Institution #3 (and to a lesser degree #2) show an unusually high percentage of steady growth collections. These institutions have also implemented a simple community-collection structure (Table 1). While the number of institutions in this comparison is quite limited, there may be reasons why these two patterns might co-occur. Institutions with narrowly targeted collections may have rigid policies that dissuade participation, and a potential DSpace contributor may be reluctant to deposit materials into a collection with little or no content. Conversely, a large and actively growing community may be perceived as having higher value to a potential contributor and encourage participation. For these communities, a Cumulative Advantage Process may be in effect, where success breeds success. [18]

Table 2. Growth patterns compared across implementations

Institution	% Plateau	% Stairstep	% Steady	% Other
Cornell	77.0	18.7	2.9	1.4
#1	60.0	30.0	0	10.0
#2	46.2	34.6	7.7	11.5
#3	38.9	36.1	16.7	8.3
#4	73.2	23.2	3.7	0
#5	54.6	40.3	2.5	2.5
#6	37.5	56.3	0	6.3
#7	74.5	23.6	0	1.8

Faculty Surveys

During the months of October and November 2006, we interviewed eleven faculty members from selected departments across campus (Communication, Applied Economics and Management, History, Mathematics, Ecology & Evolution, Molecular Biology & Genetics, Microbiology, Computer Science, Romance Literature, Electrical and Computer Engineering, and High Energy Physics). The interviews were semi-structured, focusing on the use of digital repositories but allowing the interviewer to explore tangential topics as they arose. The interview length ranged from 30 to 60 minutes, depending on the interest and availability of the faculty member.

We attempted to solicit interviews from both male and female faculty (M=6, F=5) at different stages of their career (4 assistant professors, 3 associate professors, and 4 full professors). In order to protect the anonymity of these individuals, we will refer to them by their professional title, (i.e. "the economist"), and because some departments have very low numbers of women among their ranks, we will use a masculine form (he, him, his) when gender words are unavoidable.

While this group of faculty interviewees should not be considered to be a representative sample from the entire university community, it does include members from the life and physical sciences, social sciences, and humanities, and reflects the diversity of departments commonly found at most research universities. The goal of these interviews was not to make generalizations across each discipline, but to better understand the diversity of attitudes, motivations and behaviors regarding the use of digital repositories.

Results

Access a non-issue for faculty

In the library community, the "crisis in scholarly communication" is often framed as a crisis in *access* brought about by a combination of high journal price inflation and a stasis in library funding, resulting in journal cancellations [19]. In stark contrast, access to the literature appeared to be a non-issue for all faculty interviewed save one.

The Cornell faculty are undoubtedly privileged to a more comprehensive set of resources than faculty at other institutions, yet the library collections are by no means complete. More importantly, we were interested in three different dimensions of access: faculty as authors, readers, and public citizens. As a result, we framed our access questions in three different ways: (1) Did the faculty member feel that he/she had adequate access to the work of his/her colleagues; (2) did the faculty member feel that other members in their community had adequate access to *his/her* work; and (3) did the faculty member perceive public interest in his/her work[1].

"By and large, the [research] community has adequate access to my work, either through the journal, the arXiv, or my web page," explained the engineer.

"Most of the people who are interested in my work are at institutions that have access to the journal literature," responded the ecologist, "the ones who don't go to my website or contact me directly."

These sentiments were also echoed by the microbiologist, who admitted belonging to an institution with exceptional resources. "I have access to a good library, but others may not." The microbiologist did receive periodic email, mainly from researchers in Eastern Europe and Asia, for copies of articles. In addition, the primary publisher in the field of microbiology, the American Microbiological Society, makes all articles published in their

---

[1] By public, we meant the non-research community.

journals available for free after six months. This practice reduced the obligation on the microbiologist's part to make work more public. "[As a result], I'm not terribly motivated to deposit articles [in a repository]."

The high energy physicist noted that his colleagues are located around a small number of geographically dispersed particle accelerators. In order to do research, one needs access to one of these accelerators. Similar to astrophysics, doing research in particle physics has two main barriers to entry: access to the literature and access to equipment. The second barrier is much more significant than the first and is the principle limitation to doing –and publishing– research [20]. As a result, most particle physicists are *de facto* located at institutions with large resources and excellent information resources. Access to the literature for these physicists is essentially a non-issue.

The Romance Studies professor described that access to the journal literature in South America was woefully lacking and described the disparity as a "digital divide" between the northern and southern hemispheres. Unlike the other fields, where doing research meant being located at a well-funded institution with adequate access to the literature, romance studies did not necessarily have that association, which resulted in a large disparity between the haves and have-nots. As a solution, the sharing of IDs and passwords with individuals at other institutions was a common way to partially relieve access issues.

All but the Romance Studies professor believed there was little or no public interest in their work. All conceded that they work in very narrow fields with little (if any) public interest in their primary research, although there may be some public interest in digested forms of their work, such as a website that translates and summarizes medical research into layperson's terms.

Personal or Group Web Pages

All eleven faculty members had personal Web pages, either maintained by their departments or by themselves, and every faculty member, with the exception of the two humanities professors (History and Romance Studies), used them for disseminating some version of their scholarly work. Many of the scientists also participated in laboratory or research group Web pages. The computer scientist seemed most open about sharing work. His team's website is used to disseminate conference proceedings, journal articles, datasets, programming code, and data annotation instructions.

The reasons cited for using a personal or group Web page to post digital objects were ease and control. For some departments, Web pages were designed to give all faculty pages the same look and feel and were used to promote the school and its faculty to potential graduate students and to the general public.

Reasons for using a digital repository

If a personal, departmental, or laboratory group Web page can serve as a resource for dissemination of digital objects, why do faculty use digital repositories?

Permanence

The mathematician responded that permanence was especially important. "One's Web page can change. With the arXiv, I don't have to manage it." The computer scientist cited data migration, especially document format migration, as an important function of a repository. While the PDF is the standard for documents today, there was the assumption that the sponsoring organization would take responsibility in transferring these documents into formats that could be read by future software.

Policy of granting agencies and publishers

The molecular biologist and ecologist mentioned that granting agencies, such as the NIH, are putting more pressure on authors to deposit manuscripts or supplemental datasets as a condition of funding. With regard to depositing objects, the molecular biologist acknowledged, "I know that I should do it, but I'm hoping that it is being done for me automatically." This sentiment was echoed by the microbiologist.

Timeliness

While not limited to digital repositories, the issue of timeliness for disseminating scholarly works was cited by many faculty as reason for using repositories or their own Web page for disseminating draft or accepted manuscripts. The ecologist admitted that "The field of ecology moves slowly compared to, say, molecular biology, so getting information out as soon as possible isn't as important." Publishers who are releasing articles before print (described as "early view" or "online early") are reducing the publication lag of journal and making the manuscript version of one's article less important, according to the ecologist.

Registration

The engineer described the registration feature in the arXiv as a key reason for its adoption. By putting a date stamp on every deposited manuscript, the arXiv fulfils the first function of scholarly publishing – the registration of new ideas[2]. The engineer described that researchers would often submit new findings to the next conference, however remote, if they were concerned about being the first to report a discovery. This was much faster than waiting for final publication to stake a claim and reduces the risk of being scooped. In this way, the registration feature of the arXiv is more efficient and cheaper than the conference route. It is also more timely and speeds up dissemination of research findings. To avoid having to keep multiple versions current on both his Web page and in the arXiv, his personal Web page simply points to documents hosted on the arXiv.

---

[2] DSpace performs this function as well.

Reasons for not using a digital repository

While permanence, granting policy and registration of ideas were described as reasons for using a digital repository, there were many reasons for their non-use.

Learning curve

The Communication researcher described that there was a learning curve required for using any new technology and didn't see the reason for using a new system that was not perceived as adding additional value. This sentiment was shared by the economist, who stated that "[my Web page] meets my need for professional standards and recognition."

Copyright concerns

The economist, molecular biologist and microbiologist all were unsure about copyright issues and what was permitted by the publisher. The molecular biologist voiced a certain amount of confusion about what authors can do with their papers and stated that the rules were constantly changing. It was safer, therefore, not to make articles available from sources outside the publisher's website. This sense of caution was not shared by all. The ecologist routinely puts up final published versions of articles on a personal website, and asked rhetorically, "Is it legal? I don't care."

Publishing original work

There was also some question about what is considered "published." The molecular biologist asked rhetorically, "Is a preprint in a repository considered a publication?" If the answer is "yes", then this has implications for having one's work accepted in some journals that prevent previously published work from being submitted. Consequently, there is some reluctance to use a new system like a repository if it could possibly jeopardize one's publication success.

Quality association

Quality association was a reason used by the economist for not participating in a digital repository. The most common repository in economics (ERN) "has some good material, but it also has some material of questionable quality," and this economist didn't want to associate his own work with a site perceived to have variable quality. He also perceived that some researchers used ERN for "self-aggrandizement," which is frowned upon in the culture of economics. "You don't want to be known as an SSP [shameless self-promoter]." The best sign of quality work was getting published in one of the top five journals in economics.

Fear of plagiarism and being scooped

The benefits of timeliness were clearly not felt by either of the humanities faculty. In fact, early dissemination of one's work came with very real risks.

"Timeliness is not an issue in History. Historians work on topics for a very long time." In addition, there appeared to be little if any incentive to make one's work available before formal publication. "[In history], disseminating work before it is finished and published will not help you [professionally]. There is absolutely no incentive to make

your work broadly available electronically before it appears in print." Both the historian and Romance Studies faculty mentioned that plagiarism was a real concern and a reason for not disseminating their work early in electronic form. The molecular biologist was afraid of unscrupulous use of data and results, and held a general paranoia of releasing results before an article was received for publication. Releasing results before formal publication was equated with giving away one's competitive advantage over discoveries.

Reputation and the importance of accuracy

In contrast to the physical scientists and engineer, the life scientists were much more reluctant to share work that had not gone through peer review and been accepted for formal publication. The molecular biologist remarked, "I'm careful not to clutter the world with mistakes." This was almost identical to the remark by the microbiologist, "I don't want erroneous information to be propagated." Misinformation can have lasting implications on their fields of research. Both faculty members also mentioned that rigor and accuracy were tied to one's professional reputation.

Reasons for not using Cornell's DSpace

Of the eleven faculty members interviewed, only four knew about the existence of DSpace, and only one member (the historian) had deposited items in it.

Use of subject repositories

Those faculty who used a subject repository felt these services filled their needs, making DSpace redundant. Three faculty (mathematician, engineer, and high energy physicist) used the arXiv, the molecular biologist used PubMed Central, and the microbiologist had used GeneBank for depositing gene sequences. The computer scientist cited a specific disciplinary repository, which for the purposes of protecting his anonymity, cannot be stated.

Lack of DSpace functionality

The historian who had experience using DSpace did not use it for personal scholarly output, but instead for depositing scanned primary documents for the purpose of teaching. A permanent home and URL were the reasons for using DSpace. "I never wanted a dead link." While the historian was very grateful for the existence of DSpace, there was a perceived lack of software functionality. Categories are inflexible, and one could not delete, move objects, or cross-list objects across categories[3]. The historian noted that researchers approach the primary literature in various ways (by date, topic, region, historical figure names), and the inflexibility of DSpace made it necessary to design an additional search interface to the DSpace documents. This was not a perfect solution, as one cannot link directly to a DSpace object -- only the description page for that document. This resulted in an additional step required by a researcher, and because of it, "we're probably losing a lot of people at this step [who don't reach the document]."

---

[3] DSpace version 1.2 has since addressed cross-listing.

Community Salience

When asked about which community they associated more strongly with, all faculty described an international community of researchers working in a narrow discipline. As the economist most aptly stated, "Academics are still very isolated within their own discipline." Given the choice of associating one's work with an institution or a research community, all chose the research community, although several did point out that certain types of documents (teaching materials, theses and dissertations, and descriptions of local resources) do associate better with an institution.

Repositories as Islands

All faculty, with the exception of two, perceived an institutional repository as a single island completely isolated from other institutional repositories. In this sense, there was a general perception that one would have to know *a priori* where to find relevant information. The molecular biologist remarked that faculty move, and it was unrealistic to have material in multiple repositories and require someone to search them separately. The microbiologist had a similar response in that any institutional repository would be biased in that they don't aggregate material around a topic. In this sense, many faculty still perceived repositories as isolated and unique resources. In contrast, the computer scientist and engineer had correctly identified that the metadata in most institutional repositories were indexed and could be searched by tools such as Google™. It didn't really matter where a document was being stored, as long as it could be easily found using a standard search engine.

Culture of disciplines

Cultural norms were regularly brought up as justification for behavior. The mathematician essentially took a normative and pragmatic approach to how he communicated his work, "I would use the repository that is used by the rest of my community. If an institutional repository is not coming up regularly in a search, I would not put my papers there."

Just as the culture of many scientific disciplines has defined the articles as the unit of scholarly dissemination, the culture of history focuses on the monograph. "In our profession, we are expected to publish books, and that is where our reputation is built." And where many of the scientists were rewarded with collaborative work, it is still rare in the humanities. "Coauthorship is extremely rare in history" lamented the historian, "It is a very lonely profession."

While some disciplines have adopted repositories as normative behavior, it was important to ask what function(s) the journal provides in the scholarly communication process. According to the mathematician, "Getting published [in a journal] conveys a stamp of quality. It has nothing to do with dissemination. Journals [also] convey a certain status, something that the arXiv cannot do, at least not at present."

It was clear from the conversation with the molecular biologist that funding was responsible for a lot of the normative behavior in the field. Specifically, the decline in

funding from the National Institutes of Health has been making researchers more competitive with each other. Researchers are spending more time writing grants but fewer are getting funded. This has made researchers more guarded about sharing their results until they are formally published.

If most current work in High Energy Physics is so freely available in disciplinary repositories like the arXiv, what is the role of journals like *Physical Review*? According to the physicist, a published journal article can serve as a "sign of accomplishment" and "an accumulation point of past work." Building a new particle accelerator is so complicated and takes decades to plan and implement. Some people dedicate their entire lives to building accelerators, and as such, so much of the work is to get the accelerator working so one can start doing science. As graphically described by the physicist, "We publish in concrete and steel."

The Future

From these interviews, it is clear that faculty hold various perceptions about the functions, risks, and benefits associated with using digital repositories, and that these perceptions may be largely defined by disciplinary norms and their reward structure. It would be unfair, however, to suggest that the norms of a discipline are rigidly defined and immutable.

"It is not clear in my mind the future of journals versus the arXiv," stated the engineer. "The old school of paper journals and digital repositories are on a collision course."

The reward structure established by each discipline largely defines the motivation behind faculty behavior. As eloquently stated by the economist, "While we are going through a digital revolution – in the way we teach and communicate with each other – the reputation of being published in the print journals is still the strongest incentive for motivation." This position was largely echoed by the engineer, who stated "what is holding us to the journal is the promotion procedure. This is about a problem of measurement with how Cornell evaluates my work."

That said, there are real risks associated with changing one's practices, especially when one assumes the role of an early innovator. As the communication faculty member summarized, "There has to be a better way than the current system, but I'm not willing to be on the leading edge in using that system."

Conclusions

Cornell's DSpace is largely underpopulated and underused by its faculty. Its complex organization is seen at comparable institutions, but may discourage contributions to DSpace by making it appear empty. In addition, faculty have little knowledge of and no motivation to use DSpace. Each discipline has a normative culture, largely defined by their reward system and inertia. If the goal of institutional repositories is to capture and preserve the scholarship of one's faculty, IRs will need to address this cultural diversity.

Acknowledgements

References

1.   Lynch, C.A.: Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. ARL Bimonthly Report (2003) 1-7. <http://www.arl.org/newsltr/226/ir.html>

2.   Harnad, S.: Scholarly Journals at the Crossroads: A Subversive Proposal for Electronic Publishing. An Internet Discussion about Scientific and Scholarly Journals and Their Future. Chapter 1. The Subversive Proposal. Association of Research Libraries, Washington D.C. (1994) <http://www.arl.org/scomm/subversive/sub01.html>

3.   Crow, R.: The Case for Institutional Repositories: A SPARC Position Paper. SPARC, Washington, DC (2002) 37. <http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf>

4.   Correia, A.M.R., Neto, M.D.: The role of eprint archives in the access to, and dissemination of, scientific grey literature: LIZA - a case study by the National Library of Portugal. Journal of Information Science **28** (2002) 231-241.

5.   Smith, M., Bass, M., McClellan, G., Tansley, R., Barton, M., Branschofsky, M., Stuve, D., Walker, J.H.: DSpace: An Open Source Dynamic Digital Repository. D-Lib Magazine **9** (2003) <http://www.dlib.org/dlib/january03/smith/01smith.html>

6.   Tansley, R., Bass, M., Smith, M.: DSpace as an open archival information system: Current status and future directions. In: Koch, T., Solvberg, I.T. (eds.): Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science, Vol. 2769. Springer, New York (2003) 446-460.

7.   Smith, M., Rodgers, R., Walker, J., Tansley, R.: DSpace: A year in the life of an open source digital repository system. In: Heery, R., Lyon, L. (eds.): Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science, Vol. 3232. Springer, New York (2004) 38-44.

8.   Smith, M.: Exploring variety in digital collections and the implications for digital preservation. Library Trends **54** (2005) 6-15.

9.   Tansley, R., Smith, M., Walker, J.H.: The DSpace open source digital asset management system: Challenges and opportunities. In: Rauber, A., Christodoulakis, S., Tjoa, A.M. (eds.): Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science, Vol. 3652. Springer, New York (2005) 242-253.

10.   Lawal, I.: Scholarly Communication: The Use and Non-Use of E-Print Archives for the Dissemination of Scientific Information. Issues in Science and Technology Librarianship **36** (2002) <http://www.istl.org/02-fall/article3.html>

11.   Rowlands, I., Nicholas, D.: New journal publishing models: an international survey of senior researchers. (2005) 75. <http://www.ucl.ac.uk/ciber/ciber_2005_survey_final.pdf>

12.   Ware, M.: Institutional repositories and scholarly publishing. Learned Publishing **17** (2004) 115-124.

13.   Ware, M.: Scientific publishing in transition: an overview of current developments. Mark Ware Consulting, Ltd. Commissioned by the International Association of Scientific, Technical & Medical Publishers, and the Association of Learned and Professional Society Publishers, Bristol (2006) 30. <http://www.alpsp.org/news/STM-ALPSPwhitepaper.pdf>

14.   National Institutes of Health: Report on the NIH Public Access Policy. In: Department of Health and Human Services (ed.): (2006) 8. <http://publicaccess.nih.gov/Final_Report_20060201.pdf>

15.   Lynch, C.A., Lippincott, J.K.: Institutional Repository Deployment in the United States as of Early 2005. D-Lib Magazine, Vol. 11 (2005) <http://www.dlib.org/dlib/september05/lynch/09lynch.html>

16.   van Westrienen, G., Lynch, C.A.: Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005 D-Lib Magazine, Vol. 11 (2005) <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>

17.   King, C.J., Harley, D., Earl-Novell, S., Arter, J., Lawrence, S., Perciali, I.: Scholarly Communication: Academic Values and Sustainable Models. Center for Studies in Higher Education, University of California, Berkeley, Berkeley, CA (2006) 124. <http://cshe.berkeley.edu/publications/docs/scholarlycomm_report.pdf>

18.   Price, D.S.: A General Theory of Bibliometric and Other Cumulative Advantage Processes. Journal of the American Society for Information Science **27** (1976) 292-306.

19.   English, R.: Scholarly Communication and the Academy: The Importance of the ACRL Initiative. Portal: Libraries and the Academy **3** (2003) 337-340.

20.   Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., Murray, S.S.: The effect of use and access on citations. Information Processing and Management **41** (2005) 1395-1402. <http://arxiv.org/abs/cs.DL/0503029>