Proposal to Mellon for Exploratory Research:

Identifying Factors of Success in CIC Institutional Repository Development

Carole L. Palmer, Investigator
Associate Professor of Library and Information Science
Associate Director, Library Research Center

On behalf of the Committee on Institutional Cooperation

18 October 2006

Introduction
With the general support of the CIC members, the University of Michigan and the University of Illinois at Urbana-Champaign have been working together to share ideas and accomplishments of their institutional repositories (IR) – Deep Blue and IDEALS. Their collaborative goals include creating interoperable systems and learning from one another’s mistakes and successes. They hope later to offer their knowledge to the CIC as a whole, thence to the UC system, the AAU, and the academic world.

Although there are numerous technical hurdles to overcome in IR development, some of the greatest challenges are behavioral, social, and organizational: How do libraries get faculty and graduate students to put material into an IR? How do they manage the rights environment to allow material to be readily and appropriately deposited and disseminated? What types of material should be prioritized in collection development? What can libraries learn from scholarly communities that have adopted IR-compatible models for exchange of research papers?

Paul Courant (University of Michigan) suggests that:

The ideal experiment would be to saturate a field (or the universe of scholarly material across all fields) such that essentially everything would be available at no marginal cost to the user. Theoretical physics has come close to this ideal; everyone puts their work up on the arXiv\(^1\) and everyone in the field reads what interests them. The important peer review takes place online, as people read and comment on each other’s work. Publication per se still matters, but is often viewed as a formality. This state of affairs is hard to replicate, because few fields have as much agreement on methods, and as much clarity about what is good work and what is not. But the physics example is instructive in that it is quite clear that were all fields of academic inquiry organized like physics, the work of

librarians would be to harvest the material and assure that it was well cataloged and preserved.

The success in physics is aligned with the culture of the discipline. Within the field, preprint exchange was established long before it migrated to the digital medium, and there was early adoption of electronic documents as the preferred mode of production, in a large part because they were so easy to exchange. Interestingly, Pinfield (2001) found that arXiv.org was used primarily by the physicists who contributed to it and that it had not replaced formal journal publication, but instead served as an additional means of distribution. The level of adoption has no doubt increased considerably since Pinfield’s study, again due to practices that are commonplace in the physics discipline. In recent years it has become widely recognized that information technology development needs to be responsive to the critical differences among research communities in terms of the types of information valued and the various roles information plays in the production of research. At the same time, to realize economies of scale, IR developers must also understand the commonalities across scholarly and scientific fields.

In a recent discussion of IRs that was held in Ann Arbor, the point was made that any success that was specific to a field would likely be interpreted as being idiosyncratic to that field. On the other hand, if an institution were able to get broad participation across fields, the rest of the country would likely beat a path to their door to find out how. Our proposal is to try to determine success factors for IR participation among faculty and the applicability of these factors across fields. The essential key to success, we believe, is to provide faculty with good reasons to participate and to make it very easy for them to do so. How can we construct IRs that have the degree of draw and cachet that arXiv has in the physics community?

Background
In the UK, Europe, and Australia, repository implementations in academia have been more plentiful, in part because they have benefited from central government support and coordination. There has also been considerable dialogue on the value of mandates for populating IRs. The 2006 report produced by the Australian Department of Education, Science, and Training encouraged both “hard and soft” mandates at the national and institutional levels, as well as from funding agencies, to ensure deposit of funded research.

---


and to sustain long-term success of IRs. Pinfield (2005) asserted that such mandates “would simply help to overcome quickly the cultural and managerial barriers that currently exist; something that would otherwise take a number of years.” Without such mandates, IR initiatives at MIT and the University of Glasgow have been slow to foster faculty participation, and analyses across academic disciplines show low depositing rates, especially in the humanities fields. Yet there remains concern among university administrators that mandates could lead to resentment among faculty at their institutions, and that differing social and cultural practices among disciplines are not being accommodated. Thus, disciplinary opinion leaders may be key to promoting IRs in specific fields, and uneven adoption across fields is likely to be the normal course of development, as has been the case with other scholarly technologies.

In a recent survey of institutional repository deployment in thirteen countries, van Westrienen and Lynch (2005) reported that the greatest barriers to deposit included confusion about intellectual property issues, time-consuming submission processes, and a lack of mandatory deposit provisions by institutions or funding organizations. Factors cited as stimulators to uptake included convenient and easy submission process (library involvement in the submission was viewed as a strong positive support mechanism), propagation of materials from local repositories to national or disciplinary repositories, and indexing in Google or other methods of ensuring more widespread access to

---


repository content.\textsuperscript{11} In the UK, JISC has supported numerous efforts to stimulate repository creation and uptake among scholars and to understand factors that have determined success thus far. In particular, Swan and Awre (2006) developed a framework of user-centered services for a linked repository environment. Their study corroborated van Westrienen and Lynch’s (2005) findings on stimulators for uptake and also showed that faculty placed high value on specialized retrieval services, customized views of content according to topic or specialty, and indexing of the content in Google for broader coverage.\textsuperscript{12}

A study funded by the Institute of Museum and Library Services (IMLS) conducted at the University of Rochester aimed “to understand the current work practices of faculty in different disciplines in order to see how an IR might naturally support existing ways of work.”\textsuperscript{13} The investigators conducted twenty-five videotaped interview sessions with faculty from 6 academic departments to explore current scholarly work practices. They concluded that much more IR content was needed before use would increase, which in turn resulted in a plan for librarians to help IR staff encourage faculty to supply content. Work practice studies are essential in coordinating development of digital content and functionality with the actual daily and long-term needs of scholars and scientists. However, the chain of inference required to determine the implications of such studies can be greatly shortened if investigators directly examine how IR technology and applications are being received and adopted as they are actively developed and introduced at multiple sites.

Our proposal to Mellon comes approximately 2 years after the Rochester study and aims to bypass the extended chain of inference from scholarly work practice to suggested implications. The University of Michigan already holds significant content in its IR (approximately 24,000 publications), and both Michigan and Illinois have hired “sherpas” or similar library staff who will work in a variety of ways with faculty to help build both content and use of the institutional repositories. A third study site, Purdue University, is making considerable progress on working with faculty on curation and ingest of data sets for their IR. Their case introduces the opportunity to examine how behavioral, social, and technical factors differ for data as compared to documents in IR development.

These three sites (Michigan, Illinois, and Purdue) represent sophisticated IR operations that benefit from strong institutional support and management by dedicated professionals.


Each is applying varying strategies and priorities for working with faculty and promoting development and use of their IR. These are prime conditions for empirical, comparative case studies of IR development. Moreover, CIC Provosts strongly endorse this effort and are looking to the results to guide the development of institutional repositories across the CIC. (See attached letter of support). The Provosts have convened two summits to discuss and examine issues of scholarly communication, and the proposal for this experiment is a direct outcome of and recommendation from those summits. The Provosts will review and act upon the recommendations from this experiment at their 2007 meetings.

Approach and Research Questions
This pilot study will explore factors that contribute to successful development of IRs, using Michigan, Illinois, and Purdue to represent differing institutional IR development profiles. These institutions make a productive sample in that they are applying very different strategies, incentives, and priorities to populate their IRs and encourage participation. Each takes a different perspective on how new professional roles can best advance the creation and use of IRs. In general terms:

- Michigan is applying a critical mass content strategy complemented by an investment in dedicated professional lines for work with faculty. They have been interested in how to seed content without the direct input of faculty, and with their new sherpa positions will be focusing more on faculty awareness and outreach strategies as well as rights issues.

- Illinois is relying on the more traditional role of departmental liaison librarians with assistance and guidance from IR management and technical professionals. They are interacting directly with department heads to develop mechanisms for faculty submission that are linked to established routines, such as annual review and promotion processes. At the department level, IR participation is also being sold as a way for faculty to increase their exposure and citation (“hit”) rates.  

- Purdue has adopted a multi-disciplinary team approach oriented toward assisting faculty in managing and integrating data and information for their research projects. They are making important strides in determining more explicit roles for the research library in customizing curation of content to assist in long-term scientific data management, a challenge that looms large at most universities.

---

14 In addition, we will be considering one particular incentive-building strategy being developed at UIUC. The BRAIN project (John Unsworth, investigator) has developed a current awareness system for IR depositors. In the course of the study we will assess the viability of this technology as an incentive device and for possible deployment in IRs. See the BRAIN website at http://brain.lis.uiuc.edu/wiki/index.php/Main_Page.
These varying organizational profiles are ideal for a comparative case study that allows both within-case and cross-case analysis. The scope of the project is too constrained for full case development. However, the range of specific strategies being applied by the sherpas and other library professionals across institutions will be identified in the first stage of the project and those that emerge as the most promising will be targeted for more complete data collection as the project progresses. Moreover, while we will begin with a focus on instances of success, our experience shows that the problems or roadblocks encountered in technology development can also be highly instructive. Therefore, we may also explore emergent problems and problem-solving processes in IR development that can help us collocate, assess, and elaborate IR development strategies and their outcomes.

The primary research question is:

Do librarian intermediaries make a substantial contribution to the IR development process?

To address this question we will

- assess the influence of interventions applied by sherpas, liaison librarians, and data support specialists on faculty participation and attitudes, as well as on IR development team efforts more generally.

- determine the impact of systematic departmental initiatives and incentives as compared to more consultative and individualized liaison work performed by library personnel.

- identify potentials for leveraging informal colleague networks to build awareness and possibly increase submission patterns.

A secondary question will also be interrogated:

To what degree do scholar “demographics” influence faculty interest and activity IR development?

We will examine how seniority, experience in the field, and need for publication and recognition relate to participation, and especially the dynamics associated with those who do not characterize themselves as “early adopters” of technology.

The focus of this pilot project is twofold: what have librarians done and how have faculty responded and why? Subsidiary questions about IR use and points of access in searching are beyond the scope of this study and will likely require more fully developed IRs before we can learn anything substantive about use behavior at CIC institutions. However, using the case study approach we do expect to gain insights into additional institutional and
disciplinary variables of significance in IR planning and implementation that will serve as the basis for a larger research project discussed briefly below.

Methods Sketch
Case studies will be developed through face-to-face interviews and analysis of documentary evidence. Interviews will be conducted with sherpas and the other librarians involved with promoting IR participation among faculty, as well as technical and management operations. Selected faculty who have deposited works in the IRs at Michigan and Illinois and those involved in data experiments at Purdue will also be targeted, as will department heads who have become involved in promoting IR submission with their faculty. The interviews will be held at 3 points in time over a period of a year using a standard set of semi-structured questions for each group: sherpas and other IR personnel, faculty, and academic administrators. The documentary evidence analyzed for each case will depend on availability, but will be essential for providing context on previous stages and future plans of IR work at the institutions. We expect to be able to gain access to materials such as records of internal IR operations and decision making, collection development policies, and correspondence between IR developers and departments and faculty, for comparison across cases.

Post-Pilot Research
Results from the proposed project will serve as a base for additional IR research. During the course of the pilot project we will be able to identify and explicate factors impacting IR development across institutions related to specific technological implementations and advances, particular requirements for ingest and functionality (i.e., documents vs. data requirements), progress on policy development and application, and related issues. Differences and commonalities in these factors across disciplines are also likely to emerge. These observations will provide evidence for developing working hypotheses for an extended project on the more global dynamics of IR development and scholarly communication.

The next phase of research will require collaboration with representative scholarly societies and research library constituencies. Its aim will be to investigate how IRs can be developed in concert with other digital content efforts to effectively advance research and scholarly communication, and in particular to determine how institutional repositories can best complement and integrate with parallel disciplinary efforts. Currently, interaction and understanding of activities and commitments among developers of institution-based and discipline-based initiatives is extremely limited and will slow advances in the management of digital content and the associated tools, and ultimately research progress.