First We Build Them, Then What? : The Future of Institutional Repositories

Carol Hixson
Head, Metadata and Digital Library Services
University of Oregon Libraries
chixson@uoregon.edu

In the United States, many academic libraries became involved in the institutional repository movement as a response to the prohibitive journal price increases of the last few decades. In the seminal SPARC position paper on institutional repositories[1] Raym Crow articulates a vision for institutional repositories (IRs) that many have subsequently tried --and failed-- to create. Although the introduction pays lip service to a broad vision for IRs, noting that "it is only to be expected that academic institutions would take an interest in capturing and preserving the intellectual output of their faculty, students, and staff," the focus of the paper is on providing a mechanism for capturing, archiving, and providing open access only to faculty output. The emphasis on faculty output is justified by the author with the hope that IRs will lay the groundwork for a new scholarly publishing paradigm, breaking the control that traditional publishers have come to exercise over scholarly content. The author most clearly expresses this hope in the conclusion, stating: "Institutional repositories represent the logical convergence of faculty-driven self-archiving initiatives, library dissatisfaction with the monopolistic effects of the traditional and still-pervasive journal publishing system, and availability of digital networks and publishing technologies." It's understandable that SPARC, an organization whose mission is to change the nature of the traditional scholarly publishing model, would nurture this hope. For many, the fate and mission of IRs are inextricably tied to the success of the open-access movement. For some, it has even taken on the tone of a moral crusade: the public good of open access to information can win out over publisher or vendor greed, if we all work together.[2] Some of us who have been working to develop these archives have come to see this as little more than a naïve wish that greatly underestimates faculty and institutional inertia to adopting this model.

What is the reality of institutional repository development after several years of intensive work around the world? Westrienen and Lynch[3] summarize and comment upon the findings of a survey conducted by the Coalition for Networked Information (CNI), the UK Joint Information Systems Committee, and the SURF Foundation in the Netherlands looking at IR deployment in thirteen countries. In spite of acknowledged flaws in the collection of the survey data, the authors are able to make a number of interesting observations about the current state of institutional repository development. One such observation is that there is great variety in the types of materials being collected in IRs around the world, ranging from books, theses, articles, primary data, audio-visual objects, course materials, or a variety of other types. The bulk of content currently being collected in IRs is text-based, although those in the United States contain a significant amount of non-textual materials. The survey further revealed that disciplinary coverage encompasses all fields, with heavy emphasis on the humanities and social sciences in many countries but with very strong representation for materials in the life sciences, natural sciences, and engineering in others. The survey also attempted to gauge how

widely such archives were accepted by academics in the thirteen countries by looking at the number and percentage of academics with at least one record in an institutional repository. While few countries were able to answer the question definitively, it was nevertheless clear that the percentage of total academics contributing to IRs is still very low, with the possible exception of the Netherlands and Germany. In spite of the relative failure to date of IRs worldwide to bolster the open-access movement, the authors speculate that in most of the countries surveyed "open-access issues in scholarly publishing may well be the key drivers of institutional repository deployment, at least in the very short term." [3] This has, indeed, been the primary motivation of many who began to build an institutional repository, a fact noted by numerous authors contributing to a recent issue of *Reference Services Review* that focused on institutional repositories.[4]

Reporting on a survey of CNI member institutions in the United States conducted in February 2005, Lynch and Lippincott[5] provide a more in-depth look at the state of institutional repository development at academic institutions in the United States. While acknowledging the limitations in the focus of the survey, the authors provide some useful observations about current trends. One such observation is that IRs are becoming well-established parts of university infrastructures in the United States, with 40% of respondents having some type of functional IR and with 88% of the remainder stating that they plan to start their own or participate in some form of consortial system. Another important observation made in this article and the companion piece cited above is that the development of IRs varies from country to country, depending upon government policies and the existing national context. Given the lack of a coordinated national policy, it is likely that the use of institutional repositories in the United States will continue to be completely voluntary, although the increasing requirements of funding agencies for data management and archiving could spur some growth in their use. Another revealing result of the survey is that a significant number of institutional repositories in the United States are collecting a far greater variety of content than just faculty e-prints. It is also clear that libraries are playing the lead role in developing IRs on their campuses, with 80% of respondents reporting that the library has sole administrative responsibility for their repository.

As Lynch and Lippincott make clear, the open-access movement and the institutional repository movement are not one and the same thing. The dream of providing an outlet for the open distribution of faculty output in digital form is a subset of a broader vision. I have come to believe that an institutional repository can and should have a broader mission, one that more closely matches the vision articulated by Lynch in a 2003 article:

> "a university-based institutional repository is 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."[6]

Although the University of Oregon's institutional repository, named Scholars' Bank[7], began with the intention to provide an archive and distribution system for faculty research in digital form, it now closely resembles Lynch's broader vision. Approximately 18% of the content archived in Scholars' Bank has been authored by faculty of the University of Oregon, and the library continues to market the service heavily to the faculty. However, the archive contains a wide array of other material, such as: campus and departmental newsletters; scholarly journals authored or edited by members of the UO faculty; student terminal projects, class papers, honors theses, or dissertations; campus administrative records; campus planning documents; Oregon city and county planning documents harvested from local government web sites; finding aids to manuscript collections owned by the library; electronic texts of Renaissance materials; and much more. More than a repository for specific types of content, the archive is a suite of services that we offer our campus community for managing and disseminating digital materials that support the scholarly mission of the University of Oregon and, by extension, scholarly research around the world.

What are some of the services that an academic IR can and should offer? Many are, not surprisingly, the standard services that libraries have been offering for years. Familiarity with the service model is perhaps one reason that so many libraries have committed to developing IRs for their campuses. Some of the standard and new services are elaborated on below.

Identifying and acquiring valuable content

Library staff can identify potential content for an institutional repository by surveying departmental and faculty web sites; talking with academic and administrative departments about their output and publications; reading campus newsletters to learn about conferences, presentations, and lectures that might merit inclusion in the archive; and reviewing print publications and contacting editors to see if they are willing to archive the digital versions from which almost all print publications originate today. The initial vision of IRs as a place to capture finished faculty output was too limited. Such a vision places these archives in direct competition with traditional publication models and expects faculty and university administrators to abandon a model they know and trust for an uncertain one that seems to require more effort on their part with a less certain outcome. Jenkins, Breakstone, and Hixson[8] discuss some of the cultural barriers that many developers have faced in trying to implement faculty-focused IRs and outline strategies for overcoming these obstacles. Likewise, Foster and Gibbons[9] discuss faculty resistance to such repositories at the University of Rochester and focus on one particular strategy for overcoming that resistance.

One of the primary services that IRs can provide is to acquire materials that would otherwise have been lost, have been inadequately archived and indexed, or were known only to a limited audience. In this category are student class papers, terminal projects, and honors papers, as well as formal theses and dissertations. Such materials have often languished for in faculty offices (or now on departmental web sites) before being lost or discarded; sometimes they are collected by university archives where they are seldom

cataloged and are difficult to discover and gain access to. Also within this category are campus newsletters that often contain unique and valuable information and that are seldom organized, indexed, or made available over the long-term to a wide audience. More frequently, this category of ephemera now includes U.S. federal, state, or local government publications that are made available for short periods of time on unstable and constantly changing web sites. Capturing the wealth of grey literature or ephemera produced, supported, or needed by the academy is a unique service that more libraries should consider providing through their IRs.

Making content available in a systematic, standardized fashion

A primary service of an institutional repository is to provide access to content in an organized fashion and to tie that content to existing standards so that it can be widely shared. One of the key advantages of a repository over individual web sites is that content in an IR is described and indexed using common principles or standards. Although there is generally room for flexibility in the description and indexing of the content, certain key data elements are usually presented in a consistent fashion, such as authors or contributors, titles, and some key words, abstracts, or full-text indexing that provides an indication of the disciplinary focus of the materials. Many of the existing IRs map their data elements to the Dublin Core Metadata Elements Set or otherwise commit to making their content comply with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Lynch and Lippincott[10] observe that even though there are only a few experimental services that are currently harvesting from IRs, the institutions building the repositories seem ready to become part of national or international systems that would make their content widely available to other institutions and researchers around the world. Building bridges to other related or similar content is a key service that library-supported IRs can provide.

Preserving content

One of the primary services that the University of Oregon Libraries emphasizes about its repository, Scholars' Bank, is that of assuming responsibility for the long-term preservation of the content deposited in the archive. Digital preservation is a much more proactive process than the preservation of analog materials. The standard permission file for Scholars' Bank grants the Libraries permission to migrate content to new file formats in order to preserve it. This implies an active program of tracking file degradation and identifying appropriate file formats for an eventual migration. It also implies an understanding that digital preservation entails much more than simply backing files up. It is important for those working on repository development to become conversant with the ongoing efforts to develop standards for digital preservation, such as the RLG-OCLC report on trusted digital repositories[11], the work of the PREMIS Working Group[12], and others. Any institution that has developed or is considering the development of an institutional repository has a responsibility to learn as much as possible, to stay informed, and to document –and follow-- local policies and practices. At this early stage of digital archiving, institutions must be careful about making absolute promises, but they should

nevertheless strive to earn the same trust for digital content that libraries have long enjoyed for their handling of analog materials.

Offering instruction on how to find and cite content

Another key service for the host of an institutional repository is providing instruction on how to find and cite the content within it and within other similar archives. Libraries have a long tradition of providing instruction in the use of indexing tools, as well as in the evaluation, citation, and use of the content that they own. This understanding of the materials from different disciplines and of the indexing tools, coupled with the commitment to educating users to become self-sufficient in their use, is a library service that developers of IRs should be prepared to continue.

Educating on copyright and intellectual property

A newer service that libraries have been developing is that of educating their users on issues of copyright and intellectual property. Librarians have developed substantial expertise in this area as *users* of content through the implementation of robust interlibrary loan and reserves services. With libraries licensing and facilitating access to more digital content, this expertise has had to expand to the digital realm where the lines are less clearly drawn. With institutional repositories, there is now also a need to educate authors about their rights as *creators* of content. A growing service that the University of Oregon Libraries offers within the context of its institutional repository is providing information for authors on how to negotiate with publishers to retain control of their intellectual property and how to investigate what their publishers will allow them to archive in an open digital archive. While it would be risky to attempt to determine on behalf of individual authors if they had the right to make a digital copy of their published work available in an institutional archive, it is becoming increasingly common for repository developers to point their contributing authors to sources of information where they can make that determination for themselves.

Assist with publication

One thing that almost all of the content in the University of Oregon's institutional repository has in common is that it was contributed to the archive by library staff on behalf of the authors or copyright owners. In this respect, implementers of IRs may find themselves performing a role that is very like publishing. This is a service that libraries have not often performed on behalf of a third party but it is an important one that developers of IRs can provide to their communities. Such assistance may also take the form of performing digitization and optical character recognition of digitized content for inclusion in the repository. Taking on the role of publisher can help assure compliance with some basic standards regarding metadata and file formats and can therefore increase the usability and preservation of the content.

Conclusion

The future of IRs is uncertain. The costs of developing and maintaining them are not well known, nor is it certain how committed individual institutions will remain to the effort in the long-term as the costs associated with IRs are better understood or increase. There is so far little consensus on the types of materials that are appropriately stored in such repositories and little practical development of federated searching across different repositories. The development of IRs in different countries appears to follow different paths, depending on national policies and infrastructures. In spite of the uncertainty of the purpose of IRs and our relative inexperience with the new services they require, they nevertheless show tremendous promise. That promise, however, may not be in fomenting a revolution in scholarly publishing, as many hope, but rather in transforming scholarship by emphasizing and collecting the material "at the edge", as Paul Gherman describes it.[13] Identifying and capturing more of the ephemera –the grey literature-- and making it available to a wider audience is where the future of IRs seems most assured of making a lasting contribution. If the hoped-for revolution in scholarly publishing happens, it will take considerably longer and will require more than the uncoordinated establishment of isolated institutional repositories at academic institutions around the world.

References

1) Raym Crow, "The Case for Institutional Repositories: A SPARC Position Paper"

2) David C. Prosser, "The Changing Face of Scholarly Communication", *BiD, textos universitaris de biblioteconomia i documentació*, no. 11 (desembre 2003).

3) Gerard van Westrienen and Clifford A. Lynch, "Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005", *D-Lib Magazine*, vol. 11, no. 9 (September 2005)

4) Ilene F. Rockman, editor, "Reference Librarians and Institutional Repositories", *Reference Services Review,* vol. 33, no. 3 (2005).

5) Clifford A. Lynch and Joan K. Lippincott, "Institutional Repository Deployment in the United States as of Early 2005", *D-Lib Magazine*, vol. 11, no. 9 (September 2005)

6) Clifford A. Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age", *ARL Bimonthly Report,* no. 226, February 2003. http://www.arl.org/newsltr/226/ir.html

7) Scholars' Bank, so named because it is for the use of the UO scholarly community and because the word "bank" conveys to U.S. constituents the idea of a secure place in which to store valuable assets, is available at: https://scholarsbank.uoregon.edu

8) Barbara Jenkins, Elizabeth Breakstone, and Carol Hixson, "Content In, Content Out: The Dual Roles of the Reference Librarian in Institutional Repositories", *Reference Services Review*, vol. 33, no. 3 (2005), p. 312-324.

9) Nancy Fried Foster and Susan Gibbons, "Understanding Faculty to Improve Content Recruitment for Institutional Repositories", *D-Lib Magazine*, vol. 11, no. 1 (January 2005)

10) Clifford A. Lynch and Joan K. Lippincott, "Institutional Repository Deployment in the United States as of Early 2005", *D-Lib Magazine*, vol. 11, no. 9 (September 2005)

11) *Trusted Digital Repositories: Attributes and Responsibilities: an RLG-OCLC Report* (Mountain View, CA: RLG, 2002). There has recently been a draft *Audit Checklist for Certifying Digital Repositories* released for public comment that is available at: http://www.rlg.org/en/page.php?Page_ID=20769

12) PREMIS (PREservation Metadata: Implementation Strategies) Working Group site available at: http://www.oclc.org/research/projects/pmwg/

13) Paul M. Gherman, "Collecting at the Edge—Transforming Scholarship", in *Collection Management and Strategic Access to Digital Resources* (Haworth Press, 2005), p. 23-34.