Scherlen, Allan. (2007) "Institutional Repositories: A Good Idea for North Carolina," ***North Carolina Libraries***, Vol. 65: no. 2 (Fall/Winter 2007): pp. 80-82. The version of record can be accessed at: http://www.ncl.ecu.edu/index.php/NCL/issue/archive

Institutional Repositories: A Good Idea for North Carolina

Allan Scherlen

ABSTRACT

Librarians at universities in North Carolina are beginning to consider whether to establish electronic repositories where faculty and students can deposit copies of their scholarship for preservation and world-wide access. This article addresses a number of questions and concerns that arise, especially for librarians at smaller institutions, as they consider implementing an institutional repository (IR) program. Does a given institution have enough scholarly content to warrant building an IR? What does an IR provide that is not already available from publishers and database providers? Why would anyone search an IR? (and) Is an IR too costly for a small institution with a limited budget to set up and maintain? The author argues that while building an IR collection requires a significant commitment in staff resources, the outcome of making the collective scholarship of North Carolina open access through IRs will be immensely beneficial to scholars, hosting institutions, students, and citizens of North Carolina and beyond.

Institutional Repositories: A Good Idea for North Carolina[1]

Introduction

Much of the scholarship at North Carolina institutions is documented in digital form but is scattered and stored in various places without adequate metadata to find it or preservation measures to assure its existence into the future. Librarians at institutions of all sizes in North Carolina are recognizing that they have a role to play in exercising stewardship over these resources and in facilitating and archiving digital scholarly communication.[2] One way to offer this service is to provide local scholars access to a Web-accessible digital archive or institutional repository (IR), as it is called in the library literature. When we examine any of the classic definitions of IRs or common descriptions of the purpose of such archives, we can immediately see that their functions and purposes are not limited to major research institutions. All institutions that produce research and scholarship of any quantity can benefit greatly by either creating an IR or finding a means of providing access to an IR for their researchers.

Realization of the importance of IRs for the dissemination and preservation of scholarly communication is spreading rapidly across the academic community. A survey distributed to 123 ARL member libraries in January 2006 revealed that 43% of the respondents already have an institutional repository (IR) in place, and 35% said that they are planning for one by 2007.[3] Thus, with over 70% of ARL libraries very likely making

IRs operational within the coming year or two, we are already seeing smaller university libraries following with plans to develop a vehicle of their own for preserving and disseminating the scholarship of local faculty and students. For example, at a meeting in February 2007 sponsored by University of North Carolina System Librarians Advisory Council (ULAC), librarians and technical staff from member libraries throughout the state began discussions on how universities of all sizes throughout the North Carolina system could create institutional repositories for their authors and scholars.

North Carolina libraries are confronted by a number of questions as they consider the challenge of moving into this arena:

- Is starting an institutional repository a prudent move for small- to medium-size institutions?
- Does our university have enough published scholarship to justify investment in an institutional repository?
- Why do we need to put our publications in an IR anyway?
- Isn't it redundant and unnecessary to duplicate what publishers are already doing very well?
- Who would go to our little IR to search for content?
- Aren't IRs expensive to set up and operate?
- What will this cost in personnel time and salaries?
- Will faculty and administration buy into the process of archiving scholarship?

Many persuasive reasons have been given for why universities should create IRs and how libraries play a central role in both implementation and operation. Reasons for establishing an IR range from broad goals such as a library seeking "to move beyond a custodial role to contribute actively to the evolution of scholarly communication"[4] to specific pragmatic reasons such as providing an alternative electronic place faculty can deposit a copy of their published scholarship for readers who lack access to the official publications. The main reasons for establishing an IR given by a sample of ARL libraries surveyed was "to increase the global visibility of, preserve, provide free access to, and collect and organize the institution's scholarship."[5]

A researcher's work is valuable, whether it is produced at a major research institution or at a smaller regional university. Important work is being done at institutions of all sizes; and while a particular location may not produce vast quantities of publications, its output, though less in number, is equally deserving of preservation and dissemination through open access. Most importantly, the cumulative effect of a number of North Carolina institutional repositories going online to make much of their content open access to the world will be a significant boon for North Carolina institutions and their scholars.

Does our institution have enough locally published scholarly to justify investment in an institutional repository? Aren't such repositories more suited for big research institutions?

The question of how much research is accomplished and published at a given university is an empirical one which can be measured to some degree. Certainly, we can begin to get indications of the archivable research of a university by perusing our departmental and faculty Web sites, examining department annual reports, requesting vitas, and searching article databases by the names of faculty members. *Web of Science*, for example, which contains *Science Citation Index Expanded* (1900-present), *Social Sciences Citation Index* (1956-present), and *Arts & Humanities Citation Index* (1975-present) contains approximately 8,830 journal titles considered the core journals in their respective fields indexed from 230 disciplines. Since *Web of Science* can also be searched by organization and city affiliation of authors, the database provides a handy way to compile a quick list of significant local research.

I searched the *Web of Science* for content at a number of selected North Carolina universities that do not yet have an institutional repository in place as of March 2007. The purpose of the search was simply to illustrate in a simple, straightforward way that North Carolina universities, especially those that are not generally regarded as "research" institutions, are producing significant scholarship in major journals worthy of archiving.[6] Using the search phrases noted in the column beneath the institution's name, I searched for both the number of overall items associated with that institution (far right column) and the number of items limited to only articles and abstract of published item (center column). The total numbers indicate (right column) all possible achiveable items (such as reports, reviews, etc) and (left column) the number of potential peer-reviewed published material for inclusion in an IR.

Table 1:

Searching *Web of Science* for content at select NC institutions for inclusion in an IR

NC Institutions Searched	Number of items tagged as 'article' & 'abstract of published item'	Number of all potential items in *Web of Science* to consider for inclusion
Western Carolina University (OG=Western Carolina)	**1065**	**1599**
Appalachian State University (OG=Appalachian State)	**1734**	**2868**
Univ N Carolina, Wilmington (OG=Univ N Carolina and CI=Wilmington)	**2170**	**3305**
Univ N Carolina, Greensboro (OG=Univ N Carolina and CI=Greensboro)	**4231**	**7222**
Univ N Carolina, Charlotte (OG=Univ N Carolina and CI=Charlotte)	**4592**	**6769**
East Carolina University (OG=E Carolina Univ)	**7401**	**12005**
Total potential items for inclusion	**21,193**	**33,768**

The results of the search provide a snapshot of one kind of potential content. Granted, many of the items resulting from this search may not be permissible by their publishers' policies for archiving, while others may not be suitable for inclusion for one reason or another. But the numbers show that these regional universities within the NC University System have a significant number of important publications published in major journals to consider for an institutional repository collection.

Another excellent way to determine the extent of potential local scholarship for an IR is to compile a faculty scholarship citation database. For example, Appalachian State University Library has been building a Web-accessible faculty publication database for several years which links citations to the library catalog and subscription databases.[7] As of January 2007, the database contained over 4500 records. An ASU Library staff member is just beginning (March 2007) to add *Web of Science* citations to the database which, as indicated in the table above, show well over 2000 additional items to consider for possible inclusion.

A growing number of institutions have created and publicly posted their faculty's research on Web pages. Barbara Blummer discusses these databases at length and argues that they show a rich variety of content.[8] As the faculty publication database grows, a more complete picture of the scope and depth of local research become apparent. Such an effort can also be a valuable proactive step in acquiring metadata for a future IR.[9]

Why do we need to put our publications in an institutional repository anyway? Isn't it redundant and unnecessary to duplicate what the publishers and database vendors are already doing very well?

When scholarship becomes accessible through an IR, it benefits authors and readers as well as the institution which hosts it. One of the primary benefits can be summarized in two words—providing access. Scholars publish articles for the purpose (besides serving the requirements of tenure and promotion) of reaching other scholars and engaging in the development of knowledge within the discipline. How well one is reaching one's peers can be gauged by the number of times an article is viewed, interlibrary-loaned or downloaded, and how often it is cited in other works. The latter measure is known as the impact factor. Studies have shown that articles available open access through such means as an IR are cited 50% to 300% more often than non-open access articles from the same journal and year.[10]

It can be argued that scholars who publish in lesser known or lower circulating journals may be especially keen to push the impact of their work and deliver their scholarship to as broad an audience as possible through open access in an IR. One can understand the appeal to scholars of small niche journals to make their work open access to the world of fellow scholars, students, practitioners, and general readers. Now that most academic journal publishers (between 80 and 90+ percent) permit some form of self-archiving by authors, institutional repositories of North Carolina can provide academic authors of the state the opportunity to make their publications accessible to the world and provide proper preservation of that material.

Many scholars in North Carolina are publishing in areas of interest to people outside academia who do not have access to journal databases or even interlibrary loan. The town manager in a poor rural area of the United States who seeks current studies in planning, a health worker who needs to keep abreast of medical research, and citizens

with access to the Internet who want to be better informed can all benefit from open access to IRs. As more North Carolina scholars become conscious of the social good of making their publications open access, whether they are at a major research institution or a small liberal arts college, many will expect their institutions to provide a means of placing their work where it can best be accessed and serving those readers who need to read it.

Universities that have already implemented IRs have discovered the value of gathering various miscellaneous collections that reside on office computers around campus: newsletters, out-of-print local publications, music performances, images, data, reports, presentations, and other documents. The IR can make these "collections" more accessible by giving them more searchable metadata and can help to assure their preservation into the future.[11] IRs already in place are revealing other benefits as well. For example, an IR contributes to the prestige of an institution by showcasing the scholarship of its members attracting the interest of legislators, donors, job candidates and students considering enrollment. It also provides a single accessible location for faculty to refer scholars and students to all of their publications.

"Who would go to our little IR to search for content?"

Many misconceptions about IRs still circulate at universities considering implementation of an IR. For example, some people assume that since IRs are created as discrete archives, they must therefore be searched individually. In fact, if an IR's content is hosted according to Open Access Initiative Protocol standards, the IR can be searched among hundreds of others as one global digital archive.[12] OAI Protocol compliance

assures that all IRs are interoperable, i.e., part of one great interconnected collection of IRs around the world). People with access to the Internet will find the content of all those little IRs – along with the content of large ones – as they search across the network of IRs using such search tools as Google Scholar or specialized IR search engines such as OAIster.[13]

Aren't IRs expensive to set up and operate?

Studies have shown that the cost of developing and maintaining an institutional repository need not be unmanageable. Peter Suber, among others, has argued that setting up an IR need not present an overwhelming financial challenge, for open-source software can be utilized to build and maintain them.[14] An ARL task force report (*SPEC Kit 292: Institutional Repositories,* July 2006), which only looked at ARL institutions working independently, found a wide range of start-up costs ranging from $12,000 to $160,000, with a mean of about $81,667 and a medium of $75,000. They found the cost on ongoing operations to range from $8,600 to $500,000. The cost of software and hardware was low. The most common IR software choice was the open source DSpace software. For the majority of implementers, salaries and benefits accounted for the largest portion of the budget – 63% of start-up budgets and 68% of ongoing budgets, on average.[15]

When looking at the documented costs of setting up and operating IRs, we must take into consideration that many of the institutions with established IRs were pioneers in the field, accruing the expenses of exploration, spending salary time in careful discussions and decision making, tweaking and debugging early software versions, and generally trying to explore and develop a new information management resource. The

costs of set up and operation for second and third generation adopters of IRs should progressively decline or at least be more predictable as systems and practice become more standardized.

A number of smaller institutions can coordinate efforts to share personnel costs, especially in the areas of system design and modification, to bring the cost for each individual institution to a minimum. But IR developers have to be willing to invest resources in ongoing operations, just as they have been willing in the past to hire catalogers and collection development librarians to serve their traditional library collections. And the IR cannot be expected to populate itself. Recent literature argues that the common strategy for recruiting content adopted by so many IRs in the early years – assuming authors will self-service their submissions – simply is proving to be ineffective.[16] Libraries must be willing to allocate the staff time necessary to make a robust IR possible. An IR program that includes proactive mediation of the IR content by designated library staff will assure not only that an adequate quantity of the institution's scholarship will be archived but also that the content will be properly coded and preserved. Ultimately, it is becoming clear that the real challenge will not be so much how to pay for the IR but how to implement a successful plan for systematically acquiring content.

Dealing with costs of an IR – sharing, absorbing, and dedicating funds

Efforts are now underway by a number of universities in North Carolina to find ways to share costs of IR implementation. The cost of IR design and programming customization of the IR can, perhaps, be shared among a number of institutions. Each

institution will still require local staff to handle many of the functions of the IR that are best dealt with locally, such as user support, advocacy and outreach, acquisition of local content and metadata entry.

Technical services personnel are naturally suited to working with the content of a digital repository. Some institutions may wish to reallocate time to incorporate the work of IR content "acquisition" and metadata "cataloging" into the technical services workflow. Student assistants can also be utilized to perform routine data entry tasks.

Funding for such a revolutionary project can also be sought by broadening institutional "buy-in." Enthusiastic faculty members who have bought into the IR idea can help to spread the word about the benefits of archiving scholarship and expand support among their colleagues. With increased faculty interest and support, administrators can be influenced and educated about how the IR is an essential component of the academic system worthy of full institutional support. Administrative support can translate into increased funding and even mandates tied to the tenure, promotion, and merit raise process.

Conclusion

North Carolina's libraries and institutions of higher education are now facing an exciting challenge to build partnerships with faculty and students to preserve local scholarship and advance scholarly communication. Regardless of the size of the institution or the amount of its published scholarship, every scholar within those institutions deserves an opportunity to preserve his or her scholarship and to maximize its distribution around the world through archiving in a Web-accessible institutional repository. The cumulative

effect of the scholarly contributions of North Carolina, large and small, to the global IR network will (like the effect of the cumulative content of the Internet itself) be a significant contribution to access and preservation of the world's scholarship. Since librarians have the expertise to properly collect, describe, and manage information, North Carolina libraries should take a lead role in their affiliate institutions in planning to collect and preserve the scholarship produced by their faculty and students.

---

[1] This paper is based upon a presentation given to the 16th Annual North Carolina Serials Conference "Serials at Warp Speed: Navigating Transitions " held at The William and Ida Friday Center for Continuing Education, Chapel Hill, NC, March 29, 2007.

[2] For an eloquent argument for the value of institutional repositories see Clifford Lynch, "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," portal 3:2 (April 2003).

[3] Charles W. Bailey, *Spec Kit 292; Institutional Repositories*, *July 2006*. Association of Research Libraries (2006).

[4] Raym Crow, "The Case for Institutional Repositories: A SPARC Position Paper." ARL No. 223 (Aug. 2002). http://www.arl.org/sparc/bm~doc/ir_final_release_102.pdf (March 20, 2007).

[5] Bailey, 20.

[6] This is not to imply that only scholarship published in major journals is worth archiving, but merely to illuminate this growing body of archivable material.

[7] Appalachian State University Faculty Publications Database, which is currently undergoing a major upgrade, can be accessed at http://facultypublications.appstate.edu/

[8] Barbara Blummer's article, "The availability of faculty publications databases from Library Web pages" *Journal of Web Librarianship* 1:2 (2007)) contains a list with URLs of various faculty publications databases.

[9] North Carolina State University, one of two universities in the North Carolina University system to have set up a faculty scholarship archive before 2007, integrated its database of faculty research citations with its IR to form a dynamic NCSU Scholarly Publications Repository that shows both archived faculty research as well as citations with links through its link resolver to full text in library databases.

[10] S. Hitchcock, "The Effect of Open Access And Downloads ('Hits') On Citation Impact: A Bibliography of Studies." Southampton archive, Last updated 8 March 2007 http://opcit.eprints.org/oacitation-biblio.html (March 27, 2007).

[11] For more information on both implementing an institutional repository and the kinds of local material that may be included, see Carol Hixson's presentation, "Implementing an Institutional Repository" at http://scholarsbank.uoregon.edu/dspace/handle/1794/3947

[12] Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) compliance assures that all IRs are interoperable.

[13] It should also be expected that indexing and abstracting databases, as well as aggregator databases that students and scholars use in libraries to find articles by subject, will see fit to eventually include links not only to the official published version of an article but also to the version preserved in an institutional repository. Some databases, such as *Web of Science*, already have links to some e-repositories.

[14] Peter Suber, *Open Access Overview: Focusing On Open Access to Peer-Reviewed Research Articles And Their Preprints*. Web site (2006) http://www.earlham.edu/~peters/fos/overview.htm (December 10, 2006), p. 6.

[15] Bailey, 16.

[16] For a recent discussion of the problem of faculty nonparticipation in a self-service IR content acquisition model, see Philip M. Davis and Matthew J.L. Connolly, "Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace," D-*Lib Magazine* 13:3/4 (March/April 2007) http://www.dlib.org/ (March 30, 2007)