Institutional Repositories:

Partnering with Faculty to Enhance Scholarly Communication

RICHARD K. JOHNSON

From its earliest days, SPARC (the Scholarly Publishing and Academic Resources Coalition) has explored strategies to unleash the power of the digital networked environment to enhance the process of scholarly communication and address the serious economic problems that plague it. During the past year, we have been following the promise and progress of early-stage institutional repositories -- digital collections capturing and preserving the intellectual output of a single or multi-university community. We believe that institutional repositories are a practical, cost-effective, and strategic means for institutions to build partnerships with their faculty to advance scholarly communication.

Institutional repositories build on a growing grassroots faculty practice of posting research online, most often on personal web sites, but also on departmental sites or in disciplinary repositories. This demonstrates a desire for expanded exposure of, and access to, their work. In addition, digital publishing technologies, ever-expanding global networking, and enabling interoperability protocols and metadata standards are coalescing to provide practical technical solutions that can be implemented now. The convergence of these interrelated strands indicates that institutional repositories merit serious and immediate consideration from academic institutions and their constituent faculty, librarians, and administrators.

This belief is reinforced by SPARC's recent experience in bringing together stakeholders to discuss the prospects for institutional repository building. The evident energy and activity give cause for optimism that institutional repositories are an emerging dimension of scholarly communications.

In a recent close examination of the topic (*The Case for Institutional Repositories: A SPARC Position Paper* <http://www.arl.org/sparc/IR/ir.html>), SPARC examined the strategic roles institutional repositories serve for colleges and universities. In the following comments, I will attempt to present key elements of the case.

Why Institutional Repositories?

The rationale for universities and colleges implementing institutional repositories rests on two interrelated propositions: one that supports a broad, pan-institutional effort and another that offers direct and immediate benefits to each institution that implements a repository.

New Scholarly Publishing Paradigm
While institutional repositories centralize, preserve, and make accessible an institution's intellectual capital, at the same time they will form part of a global system of distributed, interoperable repositories that provides the foundation for a new disaggregated model of scholarly publishing. This model unbundles the principal functions of scholarly communication, thus presenting the potential to realize market efficiencies previously hidden by the vertically integrated publishing model that now characterizes academic journal publishing.

Altering the structure of the scholarly publishing model will be neither simple nor immediate. The stakes are high for all the well-entrenched participants in the system—faculty, librarians, and publishers—and the inertia of the traditional publishing paradigm is immense. In the near-term, large journal publishers have both the power and the incentive to maintain the status quo: the prestigious journals they control appear integral to the very structure of academic professional advancement. However, digital publishing and networking technologies, harnessed by an increasingly dissatisfied library market—as well as by authors themselves—are now driving fundamental changes to this publishing model at an accelerating pace. And new communications paradigms, especially when constructed by the scholars themselves, can eliminate seemingly insurmountable publisher advantages in relatively short order.

Institutional Visibility and Prestige
Institutional repositories, by capturing, preserving, and disseminating a university's collective intellectual capital, serve as meaningful indicators of an institution's academic quality. Under the current system of scholarly communication, much of the intellectual output and value of an institution's intellectual property is diffused through thousands of scholarly journals. While faculty publication in these journals reflects positively on the host university, an institutional repository concentrates the intellectual product created by a university's researchers, making it easier to demonstrate its scientific, social and financial value. Thus, institutional repositories complement existing metrics for gauging institutional productivity and prestige. Where this increased visibility reflects a high quality of scholarship, this demonstration of value can translate into tangible benefits, including the funding—from both public and private sources—that derives in part from an institution's status and reputation.

The current system of scholarly communication limits, rather than expands, the readership and availability of most scholarly research (while also obscuring its institutional origins). Rounds of journal price increases and subsequent subscription cancellations act to reduce the audience further. In this context, the role of alternative scholarly communications models, such as institutional repositories, in breaking the monopolies of publishers and increasing the awareness of university intellectual output grows increasingly clear. Further, institutional repositories can serve this function whether they are implemented on individual campuses or in collaborative consortial projects.

Essential Elements of an Institutional Repository

Stated broadly and in the context of SPARC's focus, a digital institutional repository can be any collection of digital material hosted, owned or controlled, or disseminated by a college or university, irrespective of purpose or provenance.

Other types of institutions that generate substantial bodies of research or other intellectual property could establish repositories as well. These might include government departments or agencies, NGOs and IGOs, museums, independent research organizations, federations of societies, and (theoretically at least) commercial entities -- any organization that wishes to capture and openly disseminate its intellectual product, thus contributing to scientific/scholarly discourse and benefiting from the resulting organizational visibility.

Here, however, we will narrow our definition to focus on a particular type of institutional repository—a digital archive of the intellectual product created by the faculty, research staff, and students of an institution and accessible to end users both

within and outside of the institution, with few if any barriers to access. In other words, the content of an institutional repository is:

- Institutionally defined;
- Scholarly;
- Cumulative and perpetual; and
- Open and interoperable.

I will amplify and qualify each of this definition's elements below. However, the purpose in doing so is not to prescribe the precise requirements necessary to qualify as an institutional repository. In practice, institutional repositories can assume many forms and serve a variety of purposes. The technical and administrative infrastructures developed by academic institutions for existing digital library initiatives might often be modified or repurposed to serve the requirements of an institutional repository. Similarly, our more narrowly defined institutional repository might form a component of a more comprehensive institutional initiative, one encompassing virtually all of an institution's digital assets. Nevertheless, we need to identify essential defining elements to bound a meaningful discussion of the organizational, technical, financial, and cultural issues relevant to implementing an institutional repository.

Institutionally Defined
In contrast to discipline-specific repositories and subject-oriented or thematic digital libraries, institutional repositories capture the original research and other intellectual property generated by an institution's constituent population active in many fields. Defined in this way, institutional repositories represent an historical and tangible embodiment of the intellectual life and output of an institution. And, to the extent that institutional affiliation itself serves as the primary qualitative filter, this repository becomes a significant indicator of the institution's academic quality.

Depending on the university, an institutional repository may complement or compete with the role served by the university archives. University archives often serve two purposes: 1) to manage administrative records to satisfy legally mandated retention requirements, and 2) to preserve materials pertaining to the institution's history and to the activities and achievements of its officers, faculty, staff, students, and alumni. Compared to institutional repositories, which aim to preserve the entire intellectual output of the institution, university archivists exercise broad discretion in determining which papers and other digital objects to collect and store. Still, the potential overlap of roles of the two repository types merits consideration at institutions that support both.

Developing institutional repositories does not require that each institution act entirely on its own. For many colleges and universities, existing state or regional institutional or library consortia will provide a logical infrastructure for implementing institutional repositories via collective development. Such cooperation could deliver economies of scale and help institutions avoid the needless replication of technical systems. Indeed, consortia might well prove the fastest path to proliferating institutional repositories and attaining a critical mass of open access content.

Scholarly Content
Depending on the goals established by each institution, an institutional repository could contain any work product generated by the institution's students, faculty, non-faculty researchers, and staff. This material might include student electronic portfolios, classroom teaching materials, the institution's annual reports, video

recordings, computer programs, data sets, photographs, and art works—virtually any digital material that the institution wishes to preserve. However, given SPARC's focus on scholarly communication and on changing the structure of the scholarly publishing model, we will define institutional repositories here—whatever else they might contain—as collecting, preserving, and disseminating scholarly content. This content may include pre-prints and other works-in-progress, peer-reviewed articles, monographs, enduring teaching materials, data sets and other ancillary research material, conference papers, electronic theses and dissertations, and gray literature.

To control and manage the accession of this content requires appropriate policies and mechanisms, including content management and document version control systems. The repository policy framework and technical infrastructure must provide institutional managers the flexibility to control who can contribute, approve, access, and update the digital content coming from a variety of institutional communities and interest groups (including academic departments, libraries, research centers and labs, and individual authors). Several of the institutional repository infrastructure systems currently being developed have the technical capacity to embargo or sequester access to submissions until the content has been approved by a designated reviewer. The nature and extent of this review will reflect the policies and needs of each individual institution, possibly of each participating institutional community. Sometimes this review will simply validate the author's institutional affiliation and/or authorization to post materials in the repository; in other instances, the review will be more qualitative and extensive, serving as a primary certification.

Cumulative and Perpetual
Essential to the institutional repository's role both within the university and within the larger structure of scholarly communication is that the content collected is both cumulative and maintained in perpetuity. This has two implications.

First, whatever the content submission criteria for a repository, items once submitted cannot be withdrawn—except in presumably rare cases involving allegations of libel, plagiarism, copyright infringement, or "bad science." This removal would be the functional equivalent of revoking the registration initially granted to the contribution on accession into the repository. This does not necessarily mean that all content will be universally accessible in perpetuity. Institutions must develop criteria and policies—and implement rights management systems—for allowing access to a repository's content, both inside the institution and from outside, that balance the goal of the broadest available access with the reality of encouraging faculty participation. The cumulative nature of institutional repositories also implies that the repository's infrastructure is scaleable. While initial processing and storage requirements might prove modest, institutional repository systems must be able to accommodate thousands of submissions per year, and eventually must be able to preserve millions of digital objects and many terabytes of data.

Second, institutional repositories aim to preserve and make accessible digital content on a long-term basis. Digital preservation and long-term access are inextricably linked: each being largely meaningless without the other. Providing long-term access to digital objects in the repository requires considerable planning and resource commitments. The institution needs to balance the desire to accept the farrago of file formats popular with various disciplines, in order to simplify content submission and encourage faculty participation, with the complications that migrating some of those formats or media might present as new standards evolve. While it is possible for an institution to dictate digital formatting standards for students—in the submission of electronic theses and dissertations, for example—

prescribing such formats for faculty, for both attitudinal and practical reasons, proves far more problematic.

Interoperability and Open Access
Providing no- or low-barrier access to the intellectual product generated by the institution increases awareness of research contributions. The goals motivating an institution to create and maintain a digital repository—whether pan-institutional, as a component in the changing structure of scholarly communication, or institution-centric—require that users beyond the institution's community gain access to the content.

For the repository to provide access to the broader research community, users outside the university must be able to find and retrieve information from the repository. Therefore, institutional repository systems must be able to support interoperability in order to provide access via multiple search engines and other discovery tools. An institution does not necessarily need to implement searching and indexing functionality to satisfy this demand: it could simply maintain and expose metadata, allowing other services to harvest and search the content. This simplicity lowers the barrier to repository operation for many institutions, as it only requires a file system to hold the content and the ability to create and share metadata with external systems.

Given the disparate publishing practices amongst academic disciplines, an institution's content accession and access policies need to accommodate legitimate researcher concerns about access to pre-publication material deposited in the repository. Institutional repositories typically do not permit content to be removed once submitted. However, a variety of legitimate circumstances might require an institution to limit access to particular content to a specific set of users. These circumstances might include copyright restrictions, policies established by a particular research community (limiting access to departmental working papers to members of that department, for example), embargoes that an institution's Sponsored Programs Office might require to keep the institution in compliance with the terms of sponsor contracts, and even monetary access fees for certain data. Implementing these policy-based restrictions requires robust access and rights management mechanisms to allow or restrict access to content—and, conceivably, to parts of digital objects—by a variety of criteria, including user type, institutional affiliation, user community, and others.

What's In It for Faculty and Researchers?

The greatest obstacle to any change in the fundamental structure of scholarly communication lies in the inertia of the traditional publishing paradigm. And nowhere is that inertia more profound -- and understandable, given the professional stakes -- than amongst academic faculty. Unlike trade publishing, academic authors rarely receive direct compensation for the research articles they publish. Rather, they publish for professional recognition and career advancement, as well as to contribute to scholarship in their discipline. Accommodating these faculty needs and perceptions -- and demonstrating the relevance of an institutional repository in achieving them -- must be central to content policies and implementation plans.

The principal author benefits of online open access to their research pertain to enhanced professional visibility. This visibility and awareness is driven by both broader dissemination and increased use. No library can afford a subscription to every possible journal -- regardless of publication quality -- rendering much of the research literature inaccessible to many researchers. The OAI Metadata Harvesting Protocol creates the potential for a global network of cross-searchable research

information. By design, networked open access repositories lower access barriers and offer the widest possible dissemination of a scholar's work. Further, departmental overlay bulletins and journals can increase the visibility and status of an entire academic department, in addition to the status of its constituent faculty. Another related author benefit derives from the increased article impact that open access articles experience compared to their offline counterparts. Research has demonstrated that, with appropriate indexing and search mechanisms in place, open access online articles have appreciably higher citation rates than traditionally published articles. This type of visibility and awareness bodes well for both the individual author and for the author's host institution.

Additionally, value-added services such as enhanced citation indexing and name authority control will allow a more robust qualitative analysis of faculty performance where impact on one's field is a measurement. The aggregating mechanisms that enable the overall assessment of the qualitative impact of a scholar's body of work will make it easier for academic institutions to emphasize the quality, and de-emphasize the quantity, of an author's work. This will weaken the quantity-driven rationale for the superfluous splintering of research into multiple publication submissions. The ability to gauge a faculty member's publishing performance on qualitative rather than quantitative terms should benefit both faculty and their host institutions.

Institutional repositories can serve another function currently served by print journals: that of registering the priority of ideas and intellectual property. By removing the physical page constraints that pertain in print, digital publishing expands the amount of worthy research that can be made available for review. In this way, institutional repositories provide a venue for a greater proportion of researchers to register their work in a recognized forum. Another implication of removing page constraints affects faculty as readers-consumers: progress in most academic disciplines relies largely on the amount of available information. All things being equal, more prior research translates into more and better scholarship. Thus the ability to locate and retrieve more relevant research more quickly and easily online will improve scholarly communication and advance scholarly research.

Besides the benefits for faculty as authors, institutional repositories also deliver benefits to teaching faculty. By including non-ephemeral faculty-produced teaching material, the repository serves as a resource supporting classroom teaching. These materials might include concept illustrations, visualizations, models, course videos, and the like -- much of the material often found on course web sites. This benefit should help extend the appeal of institutional repositories across a broader audience of research and teaching faculty.

Conclusion

Institutional repositories offer a strategic response to systemic problems in the existing scholarly journal system -- and the response can be applied immediately, reaping both short-term and on-going benefits for universities and their faculty and advancing the transformation of scholarly communication over the long term.

***