

The Value Proposition in Institutional Repositories

In the education and research arena of the late 1970s and early 1980s, a struggle developed between those who advocated centralized, mainframe-based computing and those who advocated distributed computing. Ultimately, the debate reduced to whether economies of scale or economies of scope are more important to the effectiveness and efficiency of information technology, with the answer being that both are critical. A similar tension is mounting today in the “institutional repository” movement. The question is: “Should digital object repositories be individual-focused or institution-focused?” And like the centralized-versus-distributed debate of thirty years ago, there is a developing realization today that although institutional repositories must have institutional organization, coordination, and investment, they will be successful only when they achieve broad and voluntary participation by individuals in the communities they serve.

What Are Institutional Repositories?

Institutional repositories (IRs) are critical to developing, managing, and leveraging enterprise-wide digital content and bringing greater value to institutional output. Clifford A. Lynch, executive director of the Coalition for Networked Information (CNI), defines an IR 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.”¹ Geoff Payne, manager of the ARROW (Australian Research Repositories Online to the World) Project, defines an IR as “a managed collection of digital objects, institutional in

scope, with consistent data and metadata structures for similar objects, enabling resource discovery by the ‘Communities of Practice’ for whom the objects are of interest.”² Payne notes that an IR needs to allow reading, inputting, and exporting of objects to facilitate resource sharing, while respecting access constraints. Repositories are meant to be sustainable over time and support preservation strategies. Lynch’s broad definition stresses the central role of the institution, and Payne’s definition begins to address the value to individual contributors.

The open source movement has adopted the idea of institutional repositories. EPrints (<http://www.eprints.org/>) and Greenstone (<http://www.greenstone.org/cgi-bin/library>) are early examples of open source software for use with IRs. DSpace (<http://www.dspace.org/>) and Fedora (<http://www.fedora.info/>)³ are more recent examples. Free for users, open source software does generate concern regarding support and maintainability. Addressing this concern, the VITAL project from VTLS (Visionary Technology in Library Solutions) builds additional capabilities on the Fedora repository architecture and provides long-term maintenance and support (<http://www.vtls.com/>).

What Is the Value Proposition to Institutions?

An institutional repository focuses on developing, enhancing, and protecting the *value* in the creative output of the members of the sponsoring institution. The deployment of IRs can be expensive. In creating such a repository, an institution makes an implied commitment that it will

provide resources to manage the repository and will keep the contents preserved and accessible. Yet, ignoring institutional repositories may turn out to be even more expensive than deploying them.

IRs are meant to be broad in scope and to cover all objects created in the enterprise. Repositories house quality digital content and the associated mechanisms to interconnect this content. Examples of content in IRs include the following: multimedia objects (e.g., PowerPoint files, audio, video, graphic, photo, animation); datasets; scientific visualizations of datasets; Electronic Theses and Dissertations (ETDs); electronic portfolios; written materials in the form of text files, spreadsheets, doc files, RTFs, PDFs; computer executable code (e.g., a simulation used in a course); session-capture recordings (e.g., Silicon Chalk, Centra Symposium); software tutorials (e.g., on-site Element K content); administrative content (e.g., grades, transcripts, architectural drawings); archive-specific materials (e.g., all contents of a past online course).

The value to the institution comes from the collocation, the interconnection, the archiving, and the preservation of the intellectual output of the institution. However, consistent and controlled exposure of the content will perhaps provide the greatest value. This suggests that individuals from outside the institution should have relatively easy and, wherever possible, open access to these repositories. IRs that lead to partnerships and collaborations with other national and global institutions also increase the exposure, utility, and value. To facilitate this broader access, it is expected that metadata from these repositories will be har-

vested by others to create search-and-access databases that may be organized along subject and/or geographic lines. *Subject-oriented repositories*, created by scholarly societies or commercial enterprises, may collect materials on a given subject from a large number of IRs. Similarly, *region-oriented repositories*, created by regional consortia or state and national organizations, may aggregate data along geographic lines. Obviously, the extent to which these special services grow will depend on the value they bring to specific audiences.

What Is the Value Proposition to Individuals?

So far, there are few IRs in full production. Those that are in operation have lower-than-expected faculty adoption, resulting in a small amount of available content. A critical mass of content is essential to success. A recent study conducted by a team at the University of Rochester, River Campus Libraries, reveals a number of reasons why faculty may be reluctant to “surrender” their research content to an institutional repository.⁴ Faculty tend to prefer (and are accustomed to) maintaining complete control over their content. In doing so, they use a number of tools, but they have little interest in how the tools work, only that the tools did work and were available as needed. Faculty research is a closely guarded and personal enterprise. The term *institutional repository* has little appeal to faculty because it implies that the system is designed to support and achieve the needs and goals of the institution, and not necessarily those of the individual.

There are many ways in which an IR might add value for faculty and students. For instance, digital objects associated with a course or Web site could be utilized by individual faculty on Web pages or by a course management system or by a content store for Web sites dynamically generated by a content management system. A digitized map could be used by various faculty—for example, in several sections of a geography course—saving storage resources and improving the efficiency of the course management system. Repository objects such as large datasets or dataset visualizations generated in research projects could be archived for later

access by distributed members of a research project.

In concluding its study, the University of Rochester team developed two strategies to address faculty’s lack of interest in and understanding of the repository concept. The first strategy involved approaching faculty on their turf and using terminology that stressed the benefits and capabilities of the repository for individuals (e.g., the term *personal digital repository* can be used to highlight the individually customizable aspect). The second strategy dealt with creating ways in which faculty could use the repository to showcase their achievements and to participate in the design of the digital space.

In addition, another issue can inhibit the broad adoption of IRs by individuals: the constant reminder of loss of control inherent in the process imposed by institutions for submitting content to the repository. In most cases, individual contributions to repositories are mediated. Professionals responsible for managing the repository validate the content and ensure the existence of the appropriate metadata. This metadata includes descriptive metadata, structural metadata, technical metadata, and administrative metadata, with the latter consisting of two parts: rights management metadata and preservation metadata.⁵ The creation of metadata during this mediation process is slow and time-consuming and does not scale. If IRs are to be heavily used, the metadata creation needs to be as automatic and as transparent as possible. A self-submission process that automatically captures as much of the required metadata as possible, without too much additional effort on the part of the submitter, would be ideal but is easier said than done. Such a process requires additional development, starting with properly defined content models for different object types. Work on this has begun.

The value to individuals is a repository that goes beyond being a safe, reliable, and accessible archive. The IR facilitates exposure and promotes interconnection and collaboration.

Conclusion

Institutional repositories capture, retain, and leverage the value in the knowledge products of institutions and their mem-

bers. Much like institutional computing, IRs will yield maximum value to institutions only if economies of scale and economies of scope are fully leveraged. Institutions must commit to coordinating and assembling tools for the development and management of repositories, to ensuring long-term data security and integrity, and to underwriting the associated costs. In addition, supporting the goals of institutional repositories through content submission is extremely important. IRs will be successful only if a large percentage of the institutional community voluntarily subscribes to the concept of the repository and routinely participates in it. The perceived cost to these individuals will be in yielding an element of control over the output from their work, research, and scholarship. Therefore, faculty, staff, and students must perceive sufficient value at negligible cost, measured in time and effort. Addressing this “personal value” issue is the opportunity for those advocating and sponsoring institutional repositories, and for those building the required enabling tools.

Notes

1. 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>>.
2. Geoff Payne, “ARROW Institutional Repositories: A Report on the Decisions and Experiences of the ARROW Project,” presentation to *Information Online*, Sydney, Australia, February 1, 2005, <<http://arrow.edu.au/docs/>>.
3. See <<http://sakaiproject.org/>> for the potential use of Fedora within the Sakai framework.
4. Nancy Fried Foster and Susan Gibbons, “Understanding Faculty to Improve Content Recruitment for Institutional Repositories,” *D-Lib Magazine* 11, no. 1 (January 2005), <<http://www.dlib.org/dlib/january05/foster/01foster.html>>.
5. See National Information Standards Organization (NISO), *Understanding Metadata* (Bethesda, Md.: NISO, 2004), <<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>>.

Erv Blythe is Vice President for Information Technology at Virginia Tech. He serves as Editor of the *EDUCAUSE Review* New Horizons department. Vinod Chachra is CEO of VTLS Inc. He previously served as Vice President for Computing and Information Systems at Virginia Tech and is a past president of CAUSE.

