

# BENEFITS OF AN INSTITUTIONAL REPOSITORY

Several compelling reasons exist for why an organization would want to establish an IR. If an IR is right for your organization, the reasons likely include some of the following ones.

## Stewardship

At any organization digital documents of enduring value exist that need to be preserved, as well as shared. The responsibility of the care and preservation of a digital document, however, usually falls to the document's author, such as a researcher storing a conference paper on a computer's hard drive.

Without deliberate attention and effort, that conference paper can become inaccessible due to obsolescence of format (such as Word Star), obsolescence of storage medium (such as 5-1/4-inch disks), or loss of information due to degradation of the digital file itself. An IR shifts the burden and responsibility of stewardship from the individual's level to the institution's level.

To date, no agreed-on, bulletproof methodology for the preservation of digital formats exists. Many techniques such as migration, emulation, LOCKSS, and others have been proposed, and likely the best method combines all of these.

Some people use this uncertainty in digital preservation methods as a reason to delay the establishment of an IR. They are waiting for the day when digital preservation is as well-defined as the techniques for preserving a medieval manuscript.

Unfortunately, this day may never come and, in the meantime, valuable digital content is lost daily due to complete neglect. Doing nothing to try to preserve digital works of enduring value guarantees their loss.

Tasks as seemingly mundane as identifying, collecting, and storing digital content are excellent first steps for any preservation strategy, and an IR provides an infrastructure for those first steps.

## Efficiencies

Although the establishment of an IR requires the institution to incur an ongoing financial burden for the staffing, equipment, and preservation, most institutions find an overall net savings by centralizing functions that are currently occurring in a decentralized fashion throughout the institution.

Early in the development of DSpace, an institutional repository system developed through an MIT Libraries and Hewlett-Packard partnership, MIT conducted a survey of its faculty to learn about their perceptions and anticipated uses of DSpace. Included in the survey was the question "If you distribute preprint articles, how do you distribute them?"

LOCKSS, <http://lockss.stanford.edu>

DSpace, [www.dspace.org](http://www.dspace.org)

Of those who responded (n=82), 43% reported posting preprints on their own websites, 60% had sent preprints to colleagues by e-mail, and a surprising 55% reported mailing paper copies (Barton & Walker, 2002). If these results are extrapolated to the entire MIT campus, just imagine the time, money, and effort being spent in the distribution of locally authored documents.

Institutional repositories represent an opportunity to realize tremendous efficiencies by eliminating or centralizing these activities. Once the author deposits the working paper into the IR, the need no longer exists to post the document to a personal website, send copies as e-mail attachments, or send by mail carrier.

In the best case, users discover and obtain the documents on their own, without the authors' involvement, by searching the IR directly or using search services such as OAIster and Google.

The worst-case scenario has the authors providing the users with the persistent URLs of the documents in the IR, which is still far more efficient than the activities represented in the MIT survey.

IRs also provide efficiencies by removing information technology barriers. When introducing the concept of an IR, a frequent question is the difference between the IR and personal Web page.

Many people see little difference between the two because the end product is the same—a Web page from which digital documents can be accessed. The differences are in how the authors post these digital documents.

Most IR systems have a form-based submission process to deposit a document into the repository, where it quickly becomes accessible to other users. The submission process requires the completion of some basic metadata fields and then the uploading of the digital file.

Posting this same document to a personal homepage necessitates knowledge of HTML, authorized access to the file servers, or specialized software, such as an HTML authoring tool.

Controlling access to the documents on a website, a core function of an IR, requires even more technical knowledge or the assistance of a systems administrator. Therefore, a reason to establish an IR is to lower significantly the barrier to document distribution over the Internet.

### Showcase

An IR also can showcase the research, teaching, and scholarship at an institution. In a university setting, an IR provides a centralized digital showcase through which community members can highlight their work. Through an IR, prospective students and faculty can obtain a robust picture of the types and areas of scholarship in progress in a given department.

The institutional repository also serves in this showcase function by exposing its content's metadata to metadata harvesters. Documents in an IR are showing up in the search results of scholarly search engines, such as OAIster, as well as Google, Yahoo, and other general-purpose search engines. Digital scholarship, once relegated to the hard drive of a person's computer, is now being discovered, read, and cited because it was deposited into an IR.

## Wider distribution

A 1998 survey by the Association of Learned and Professional Society Publishers found that "communication to the widest possible audience" was the top objective of authors when publishing scholarly works, with 93.6% indicating that it was either "very important" or "important" (ALPSP, 1999, Appendix A).

IRs can partially fulfill this objective, as supported by the research by Steven Lawrence (2001), who found that "[o]n average there are 336% more citations to online articles compared to offline articles published in the same venue" (p. 521).

The electronic theses and dissertation (ETD) collection of Virginia Tech is a case in point. A self-examination of Virginia Tech's ETD collection found that the online availability of dissertations exponentially increased their use.

In 1996, Virginia Tech received about 175 interlibrary loan (ILL) requests per month for copies of paper theses or dissertations. In the 2002-2003 academic year, the ETDs of Virginia Tech were accessed more than 7.3 million times. That number is more than 600,000 per month or an increase of an astounding 342,000% since 1996. What students would not want to give their dissertation the potential for such use and exposure?

Scholarly communications, such as ETDs, preprints, technical reports, white papers, and even formal articles, when deposited into an IR with unfettered access, are available to a wider audience of readers than materials in subscription-based publications.

For faculty members, though, local practices of defining tenure-worthy publications can outweigh the benefits of potential wider distribution in an IR. IRs provide a vehicle for wider distribution, but local practices and expectations, as well as disciplinary customs ultimately dictate how the vehicle can be used.

## Scholarly communication crisis

IRs also can respond to the current crisis in scholarly communication. As most librarians are aware, universities, research institutes, and federal and local granting agencies pay enormous sums of money to support scholarship throughout all disciplines.

Often a product of this scholarly inquiry is a journal article. In the current scholarly communication paradigm, the author of the article surrenders copyright to the publisher of the journal.

The author's host institution's library must then purchase a subscription to the journal to provide a copy of that article back to the rest of the community members. In other words, the institution must buy back access to scholarly inquiry that it funded in the first place, often at exponentially escalating subscription prices.

Several organizations have been established to try to bring about a change to this scholarly publishing crisis, including the Association of Research Libraries' SPARC (Scholarly Publishing and Academic Resources Coalition), the Open Archives Initiative (OAI), and the Budapest Open Access Initiative. Organizations such as these view IRs as one way to provide the creators of scholarly works alternatives to publishing their findings in for-profit journal publications.

(Source: <http://scholar.lib.vt.edu/theses/data/somefacts.html>)

SPARC, [www.arl.org/sparc](http://www.arl.org/sparc)

OAI, [www.openarchives.org](http://www.openarchives.org)

Budapest Open Access Initiative, [www.soros.org/openaccess](http://www.soros.org/openaccess)

University of Rochester project, <http://docushare.lib.rochester.edu/docushare/dsweb/View/Collection-331>

As a theoretical example, Professor Smith authors an article about superconductors. He may decide to publish the article in the hypothetical, for-profit *Journal of Superconductor Research*, which has a reasonable subscription rate of \$7,500 annually.

If, however, Smith's university has an institutional repository, he has some alternatives. He could deposit a copy of the manuscript into the IR *in lieu of* publication in the journal. He could deposit a preprint of the article in the IR *in addition to* publication in the journal.

Or, Smith, frustrated by the increases in the subscription rates of the journal and the eight- to 12-month submission review wait time, decides, along with several of his colleagues, to create a new, online journal to serve as a lower-priced alternative to the *Journal of Superconductor Research*. This new online journal is stored and distributed by the university's IR.

### Words of caution

Even with all these reasons, simply building an IR does not guarantee its success or ensure it will be used. Unfortunately, "if you build it, they will come" does not yet apply to an IR.

Recognize that, for most people, using an IR means a change in their work practices and perhaps even the culture of their academic disciplines and subdisciplines. A common woe heard from the majority of early adopters of IRs is that the recruitment of content, not the technology, is the greatest barrier to success.

The University of Rochester, River Campus Libraries, with the support of an Institute of Museum and Library Services (IMLS) grant, has been studying the work practices and cultures of faculty in different disciplines to better understand how an institutional repository might be used.

An anthropologist, who conducts extensive interviews and observations of faculty, leads the project. She asks questions that paint a fuller picture of how faculty members conduct their research, with particular emphasis on the use, generation, and storage of gray literature. These findings inform modifications to the University of Rochester's DSpace institutional repository to better align it to the needs and work practices of the faculty.

Although only six months into the one-year project, Rochester already has initial findings that help explain why recruiting content for IRs is so difficult. The first reason focuses on difference between institutional and individual benefits of an IR. Many of the benefits of an IR outlined above apply primarily to the institution and are not sufficiently compelling to an individual.

At most institutions, the majority of people would agree that an institutional repository is a good thing and likely would encourage their institutions to establish one. But when they must find the time and alter daily work practices to make use of the repository, participation rates are quite low.

The many compelling reasons for establishing an IR, as detailed above, do not yet directly translate into participation—the difference between theory and practice. As evidence, a survey of 45 institutional repositories found that the average number of documents per repository was only 1,256 (Ware, 2004b).

Institutional affiliation is generally not the strongest affiliation of faculty members. In the absence of being able to identify themselves with colleagues in the same research field, who are usually at other institutions, faculty members see themselves as independent scholars with only administrative ties to those in their department and institution. The concept of an *institutional* repository is neither attractive nor practical.

More likely the scholars' interests are in a personal digital repository as described by Ganel, Katz & Metros (2004, p. 48). This interest does not require an abandonment of the centralized IR for disparate, disaggregated personal digital repositories. Rather, it suggests that for an IR to be attractive and compelling, it must have a personal focus, both in its collection development policies, as well as its user interface.

IRs impose an artificial structure and order on their content, exemplified by the action of depositing a personal document into a predefined collection. People generally do not think of their works of scholarship as necessarily belonging in one container or another. Instead, they have developed several personal, organic organizational schemas to fit their own unique needs, which seldom exactly match the schemas employed by another.

Likewise, when an IR is first established, the earliest content usually comes in the form of pre-existing collections migrated from someplace else, such as a working paper series on a departmental website. These collections are natural fits for an IR because some compelling reason existed to compile the group of documents together in the first place.

Recruiting deposits of currently "collection-less" materials is far more difficult, since IR systems require that those collection-less documents be deposited into a collection. This requirement imposes an artificial construct and disrupts a person's organic organizational schema.

A second major reason found by the Rochester project for the difficulties in IR content recruitment is that most faculty have a more pressing need that overshadows the benefits of the IR. More often than not, the faculty interviewed are going through elaborate machinations to keep their works in progress safe and organized.

For instance, when collaborating with others, they express great difficulties trying to track the most recent version of a document. In response, the collaborators develop makeshift versioning systems, such as e-mailing the document to themselves to date-stamp the version.

Versioning also is an issue because the faculty members are usually working with their documents on more than one computer, across numerous environments. A professor may have a desktop in an office, another in the lab, a third at home, and a laptop carried to the library, café, and conferences.

If graduate students are assisting in the research, then several more computers are added to the mix. Digital files are being e-mailed, FTP'd to departmental and university servers, burned to CD, stored on floppies, Zip drives, and USB keys, and even sent to family members out of state for safekeeping.

These faculty members, who are quite typical of faculty at other universities, are in desperate need of an authoring system to assist with document versioning, collaborative authoring, and centralized document access from any computer at any location.

This need is so great that an IR for the storage, preservation, and distribution of finished works is just too abstract. As the work of this IMLS grant is revealing,

IRs will not succeed unless they are aligned with individuals' current work practices and other, more pressing needs of faculty are met.

These words of caution are not intended to dissuade you from establishing an institutional repository, but to warn you that success is not guaranteed by simply running the IR technology.

Understanding your users' needs and work practices ahead of time is key. Doing so should help with the decision of whether an IR is right for your organization and, if so, which IR system might be the best fit. Moreover, the current disconnect between the theory and practice of IRs should underscore the need to establish realistic goals, expectations, and timelines for the project.

No surefire methodology exists for determining whether an organization needs and will use an IR. Institutional repositories are still too new for all the dimensions of an IR-ready organization to be known. Exploring the following questions, however, is a step in the right direction:

- How well do the reasons for an IR discussed above resonate with the culture of your organization?
- Does your organization have centralized digital document storage already?
- What is the quantity of digital scholarly communication generated by your organization that does not find its way into formal publications?
- What is the level of computer sophistication in your organization?
- How many members of your organization have their own homepages?
- Is there centralized support of personal homepages?
- What types of content do they include on their homepages?
- What are peer institutions doing?

Through these questions you can begin to determine if the needs for document storage, preservation, and access are already being sufficiently met by other systems, both internal and external. Moreover, this exercise should provide you with a rough estimate of the potential size of your IR.