The Role of Reference Librarians in Institutional Repositories

Charles W. Bailey, Jr.

Preprint: 4/6/05

What Is an Institutional Repository?

Since institutional repositories are a fairly recent development, it is not surprising that there are different views about what constitutes an institutional repository (IR). While these definitions vary, IRs are fairly easy to recognize. As Justice Potter Stewart once said about pornography (Silver, 2003), "I know it when I see it."

Clifford Lynch (Lynch, 2003) has defined and IR as follows:

> In my view, 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.

Notable in this definition is the emphasis on long-term preservation. Since preserving digital information, which can be prepared in a wide variety of formats (e.g., ASCII , HTML, or PDF), is not simple and the long-term costs of doing so are basically unknown, it is no small matter for a university to commit to preserving all these diverse and ever changing formats forever.

Mark Ware (Ware, 2004) adds OAI-compliance in his IR definition:

> An institutional repository (IR) is defined to be a web-based database (repository) of scholarly material which is institutionally defined (as opposed to a subject-based repository); cumulative and perpetual (a collection of record); open and interoperable (e.g. using OAI-compliant software); and thus collects, stores and disseminates (is part of the process of scholarly communication). In addition, most would include long-term preservation of digital materials as a key function of IRs.

Raym Crow's definition discusses the potential of IRs to change the scholarly communication system (Crow, 2002):

Institutional repositories—used in this paper to mean digital collections capturing and preserving the intellectual output of a single or multi-university community—provide a compelling response to two strategic issues facing academic institutions. Such repositories:

> o Provide a critical component in reforming the system of scholarly communication—a component that expands access to research, reasserts control over scholarship by the academy, increases competition and reduces the monopoly power of journals, and brings economic relief and heightened relevance to the institutions and libraries that support them; and

> o Have the potential to serve as tangible indicators of a university's quality and to demonstrate the scientific, societal, and economic relevance of its research activities, thus increasing the institution's visibility, status, and public value.

Bailey (Bailey, 2005) focuses on the diversity of digital materials that IRs can contain:

> An institutional repository includes a variety of materials produced by scholars from many units, such as e-prints, technical reports, theses and dissertations, data sets, and teaching materials. Some institutional repositories are also being used as electronic presses, publishing e-books and e-journals.

Using these definitions, we can make distinctions between IRs and other types of digital repositories:

1. Scholars' personal Websites often provide access to their e-prints (and possibly other digital publications); however, they are for a single individual, not an institution. Example: Dr. Carol Tenopir, http://web.utk.edu/~tenopir/eprints/index.html.

2. Academic department/unit archives provide access to the e-prints (and possibly other digital materials such as technical reports) of one department or other academic unit (e.g., school), but they do not provide access to a wide variety of types of materials for the entire institution. Example: Utrecht University, Department of Mathematics, http://www.math.uu.nl/publications/Preprints/.

3. Institutional e-print archives provide access to that institution's e-prints, but not to other types of digital materials produced by the institution. Example: Glasgow ePrints Service, http://eprints.gla.ac.uk/.

4. Disciplinary archives provide access to the global e-prints (and possibly other digital materials) for one (or more) scholarly disciplines, but they have materials from authors at many institutions and they are limited in their disciplinary focus. Example: arXiv.org, http://arxiv.org/.

While the above taxonomy is helpful, one must keep in mind that contemporary digital publishing, which is fueled by constant technical innovation, is slippery as a bucket full of eels. As scholars rapidly adopt the latest technological wizardry to meet their information dissemination needs, one must expect constant morphing of the systems they employ.

Perhaps the best way to understand IRs is to use a few of them. Below are links to selected IRs:

- Boston College: http://escholarship.bc.edu/

- Caltech Collection of Open Digital Archives (CODA): http://library.caltech.edu/digital/

- espace@Curtin: http://espace.lis.curtin.edu.au/

- Glasgow DSpace Service: https://dspace.gla.ac.uk/index.jsp

- MIT: https://dspace.mit.edu/index.jsp

- Universiteit van Amsterdam: http://dare.uva.nl/en

- University of California: http://repositories.cdlib.org/escholarship/

- University of Rochester: https://dspace.lib.rochester.edu/index.jsp

The Relationship of Institutional Repositories to Open Access

The open access movement is a significant force for change in the scholarly publishing industry. While different definitions of open access exist, perhaps the most influential is that of the Budapest Open Access Initiative (BOAI) (Budapest Open Access Initiative, 2002):

The literature that should be freely accessible online is that which scholars give to the world without expectation of payment. Primarily, this category encompasses their peer-reviewed journal articles, but it also includes any unreviewed preprints that they might wish to put online for comment or to alert colleagues to important research findings. There are many degrees and kinds of wider and easier access to this literature. By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. . . .

To achieve open access to scholarly journal literature, we recommend two complementary strategies.

   I. Self-Archiving: First, scholars need the tools and assistance to deposit their refereed journal articles in open electronic archives, a practice commonly called, self-archiving. When these archives conform to standards created by the Open Archives Initiative, then search engines and other tools can treat the separate archives as one. Users then need not know which archives exist or where they are located in order to find and make use of their contents.

   II. Open-access Journals: Second, scholars need the means to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access. Because journal articles should be disseminated as widely as possible, these new journals will no longer invoke copyright to restrict access to and use of the material they publish. Instead they will use copyright and other tools to ensure permanent open access to all the articles they publish. Because price is a barrier to access, these new journals will not charge subscription or access fees, and will turn to other methods for covering their expenses.

A second key definition is the "Bethesda Statement on Open Access Publishing" (Brown et al., 2003), which requires that:

   1. The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible

purpose, subject to proper attribution of authorship, as well as the right to make small numbers of printed copies for their personal use.

2. A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in a suitable standard electronic format is deposited immediately upon initial publication in at least one online repository that is supported by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving (for the biomedical sciences, PubMed Central is such a repository).

The first thing to note about these definitions is that open access is not the same thing as free access. This is a common misconception. For example, an electronic journal can be freely available, but, if its articles are not also available with minimal use restrictions (e.g., proper author attribution), it is not an open access journal. Consequently, free e-journals that just have conventional copyright statements—even those that allow liberal educational or noncommercial copying—are not open access journals. Open access journals typically use the Creative Commons Attribution license (http://creativecommons.org/licenses/by/2.0/) in addition to a copyright statement to clearly lay out in both layperson's and legal terms what uses are permitted. In addition to copying (and similar rights), this license allows anyone to make derivative works and to make commercial use of the material without permission. This means that a commercial publisher can republish material from an open access publisher without permission or payment of fees. Of course, an open access publisher can choose to write its own license agreement instead of using the Creative Commons one as long as it embodies open access principles.

The second thing to note is that there are two major open access strategies: self-archiving of e-prints and open-access journals. As indicated previously, e-prints can be archived in a variety of different ways, including in institutional repositories. Does this mean that every e-print is available under a Creative Commons or similar license? The answer is no. Some e-prints may be available under such terms, some may have conventional copyright statements (including those where the author has transferred rights to the publisher and the copyright statement is the publisher's), and some may have no copyright statement at all (under US copyright law such works may still be under copyright depending on when they were published). Consequently, institutional repositories (as well as other digital repositories) are not typically pure "open access" repositories. Rather, they contain digital materials that have a mixed bag of copyright or license terms, and, generally, there is free and unrestricted access to these materials.

With the exceptions that they have Creative Commons (or similar licenses) and that they are usually e-only journals to keep production costs low, open access

journals are typically very similar to conventional journals: they have editors and editorial boards, they publish scholarly articles, and they use a peer-review process.

How many open access journals are there? According to the *Directory of Open Access Journals* there are over 1,514 open access journals as of April 4, 2005, with 43 of those journals having been added in the last 30 days. An impressive number, but it is important to note the wording of the *DOAJ* selection criteria (Directory of Open Access Journals, 2005):

> We define open access journals as journals that use a funding model that does not charge readers or their institutions for access. From the BOAI, Budapest Open Access Initiative, definition of "open access" we take the right of "users to read, download, copy, distribute, print, search, or link to the full texts of these articles" as mandatory for a journal to be included in the directory. The journal should offer open access to their content without delay. Free user registration online is accepted.

What's missing? A key part of the BOAI definition:

> or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. . . .

Since the *DOAJ* omits the controversial right to use materials with minimal restrictions, it's hard to know exactly how many open access journals there really are. Casual review of the copyright statements of a few e-journals included in the *DOAJ* quickly turns up examples of free, but not open access, journals (e.g., *Bulletin of the American Society for Information Science and Technology*) as well as true open access journals (e.g., *Biomedical Digital Libraries*).

Of course, the existence of over 1,500 *free* journals is not to be lightly dismissed. Fifteen years ago when the first fledgling free e-journals on the Internet began to emerge in earnest, this would have seemed a staggering, almost inconceivable number. At the time, few conventional publishers would have believed this to be possible. They were barely aware of new free e-journals like *EJournal*, *Electronic Journal of Communication*, *New Horizons in Adult Education*, *Postmodern Culture*, *Psycoloquy*, and *The Public-Access Computer Systems Review*, which were typically allowing authors to retain their copyrights and permitting liberal educational copying.

As open access quickly matures, there appears to be more give about what can be considered to fall under its umbrella. Peter Suber, the prolific author of many lucid and insightful works on open access, has recently said (Suber, 2005):

> There is some flexibility about which permission barriers to remove. For example, some OA providers permit commercial re-use and some do not. Some permit derivative works and some do not. But all of the major public definitions of OA agree that merely removing price barriers, or limiting permissible uses to "fair use" ("fair dealing" in the UK), is not enough.

While there are a few major open access journal publishers (e.g., BioMed Central, the Public Library of Science, and SciELO), many journals are published by scholars and a wide variety of organizations, such as universities (and their subunits), research institutes, libraries, and professional associations.

While not yet common, some e-journals are being published in the context of institutional repositories (e.g., e-journals in the University of California's eScholarship repository), and there is no reason that open access journals could not be published in the same way.

The third thing to note about the open access definitions is the emphasis in the "Bethesda Statement" on permanent archiving. As seen in the IR definitions, IRs are usually conceived of as being permanent repositories, and they may have a better chance of being able deliver on the promise of permanent archiving than other archiving strategies because the institutions that sponsor IRs are less likely to go away and, presumably, these institutions know what they are getting into when they make a commitment to have an IR. The active role that many academic libraries are playing in IRs also bodes well for their permanence, since a historic role of libraries has been the preservation of information. Consequently, there is a good fit between IRs and the "Bethesda Statement."

For open access advocates, the establishment of IRs also nicely aligns with a growing movement to mandate self-archiving. As Harnad et al. (Harnad et al., 2004) indicate:

> We believe the most promising way to achieve the goal of Open Access is for institutions to introduce policies requiring that published articles be self-archived. It is they and their researchers who will benefit from maximizing research impact and eliminating the costs of lost impact. This should motivate authors and their institutions to create and fill more archives—100 universities worldwide already have.

The final thing to note about the open access definitions is the mention of the Open Archives Initiative's Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) in the BOAI definition. Although OAI-PMH seems somewhat daunting on the surface, the concept is really quite simple and

elegant. Digital objects (e.g., preprints) in repositories are described by metadata (e.g., author, title, and subject). OAI-PMH allows external systems to retrieve (or "harvest") this metadata so that it can be used for searching and other purposes. Perhaps the best known OAI-PMH search system is the cleverly named OAIster (http://oaister.umdl.umich.edu/o/oaister/), which has 5,272,686 metadata records from 458 institutions at the time this paper was written.

What can we conclude from this whirlwind tour of open access?

First, open access is a rapidly evolving concept. The vision is clear; the implementation of the vision is, of necessity, less pure. Ideally, all e-prints would be under license terms such as the Creative Commons Attribution license and all "open access" journals would be too. Then it would be possible to "use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself." As we have seen, this is not the case.   Does it matter? Yes and no. Clearly, the most critical thing is for this information to be freely available. The ability to reuse it with minimal restrictions is highly desirable, but not absolutely essential at this stage. It is a complex issue for publishers and authors, and it can emerge more slowly without significantly diminishing the significant impact of the open access movement. Open access has a lot on its plate: convincing scholars to self-archive, fostering the development of digital repositories for them to deposit documents in, adopting old funding models (i.e., author charges) and inventing new ones to support open access journals, and convincing conventional publishers to try these models. True believers will experiment with Creative Commons licenses or similar licenses; if they succeed, others will follow. Open access advocate Stevan Harnad (Harnad, 2003) has argued that the free vs. open access distinction is "both spurious and a retardant on progress toward free/open access" and that open access should simply be defined as "free, immediate, permanent access to refereed-article full-texts online."

Moreover, as the heat is turned up, conventional publishers are trying to adopt the open access concept, without necessarily fully embracing it. This leads to more free information. Although these efforts may not meet the current open access doctrinal purity test, they may, in the long run, help create new open access models and contribute to the concept's evolution.

Second, open access and institutional repositories are not synonyms. Universities and other institutions can have complex motives for establishing IRs; providing free access to all (or some) IR materials is often one of them. This point can lead to some differences of perception between librarians and some open access advocates about IR support requirements and operational costs: open access advocates may focus on technical support costs of IRs, while librarians may also be concerned with additional costs, such as staff and user training and support, IR advocacy and promotion, metadata creation and maintenance (including depositing items for busy faculty), and long-term digital

preservation. Consequently, some open access advocates can see IRs as cheap to support and quick to implement, while librarians can take a more cautious approach that takes in consideration other costs and the library maxim that it is easier to establish a new service than to stop offering one. Gibbons (Gibbons, 2004) outlines some representative annual IR costs: (1) $285,000 at MIT, (2) $100,000 (Canadian) at Queens University (staffing only), (3) $200,000 at the University of Rochester, and (4) between 2,280 and 3,190 staff hours at the University of Oregon.

Third, IRs are best seen as an enabling technology for open access and as their best hope (barring unforeseen circumstances) for establishing permanent repositories.   Faculty retire, and their publication pages vanish. As IRs become more prevalent, departmental/unit archives and institutional e-print repositories may fold as digital material migrates to the IRs.   Funding agencies may decide to stop supporting disciplinary archives with generous grants, or the individuals or organizations that offer them may lose interest. Once established as part of the institutional mission, IRs will persist and, while it is not impossible that they would cease operation, institutional inertia favors their continuation.

For more detailed information about open access, see Peter Suber's (Suber, 2005) excellent "Open Access Overview: Focusing on Open Access to Peer-Reviewed Research Articles and Their Preprints."

Reference Librarians and Institutional Repositories

 Reference librarians can play a significant role in planning, establishing, and supporting IRs. Here is a partial list of some of the possible activities that they may engage in, which are suggested by the articles in this special *RSR* issue:

1. Helping to create sensible IR policies and procedures and to provide feedback about how they work in practice.

2. Assisting in designing the IR user interface so that it is clear, easy to use, and effective.

3. Helping to identify current self-archiving activity on campus to aid the content recruitment effort.

4. Acting as change agents by promoting the IR to faculty and graduate students in their subject areas.

5. Informing faculty and graduate students about Creative Commons licensing options and publisher e-print policies.

6. Depositing digital materials for faculty in their subject areas if such assistance is desired.

7. Participating in the creation of IR metadata, such as local controlled vocabularies (e.g., subject categories for IR documents).

8. Preparing Web-based and paper documents that explain and promote the IR and advocate scholarly publishing reform.

9. Training users in IR deposit and searching procedures.

10. Assisting local and remote users with IR utilization, answering questions about IR policies and procedures, and using the IR to answer reference questions.

While all these roles are important, roles three to six are especially critical in the early days of an IR. As Gibbons (Gibbons, 2004) notes:

> Unless you can quickly prove the value of an IR, the organization's long-term commitment to the project may begin to wane. The best way to demonstrate the enduring value of the IR and to ensure its long-term survival is to quickly populate it.

Nixon (Nixon, 2002) provides an additional perspective on this issue:

> The challenge, ultimately will not be the technical implementation of an e-prints service but rather the cultural change necessary for it to become embedded and commonplace in the activities of the institution.

Reference librarians are a library's eyes and ears. They understand user needs and perceptions. They know what's working and what's not. When they act as subject selectors, they are the library's primary liaison with faculty in their subject areas and its most visible representatives. They know how to help, inform, persuade, and teach users. For an IR to succeed, it is essential that they be involved in its planning, implementation, and operation.

References

1. Bailey, Charles W., Jr. (2005), *Open Access Bibliography: Liberating Scholarly Literature with E-Prints and Open Access Journals*, Association of Research Libraries, Washington, DC, pp. xviii, available at: http://www.escholarlypub.com/oab/oab.htm.

2. Brown, Patrick O. et al. (20 June 2003), "Bethesda Statement on Open Access Publishing," available at: http://www.earlham.edu/~peters/fos/bethesda.htm.

3. Budapest Open Access Initiative (14 February 2002), "Budapest Open Access Initiative," available at: http://www.soros.org/openaccess/read.shtml.

4. Crow, Raym (2002), *The Case for Institutional Repositories: A SPARC Position Paper*, The Scholarly Publishing and Academic Resources Coalition, Washington, DC, available at: http://www.arl.org/sparc/IR/ir.html.

5. Directory of Open Access Journals (2005), "Questions and Answers," available at: http://www.doaj.org/articles/questions#selectioncriteria.

6. Gibbons, Susan (2004), "Establishing an Institutional Repository," *Library Technology Reports* 40, no. 4, pp. 54, 56.

7. Harnad, Stevan (14 December 2003), "Re: Free Access vs. Open Access," OAI-eprints, available at http://lists.openlib.org/pipermail/oai-eprints/2003-December/000127.html.

7. Harnad, Stevan, Tim Brody, François Vallières, Les Carr, Steve Hitchcock, Yves Gingras, Charles Oppenheim, Heinrich Stamerjohanns, and Eberhard R. Hilf (17 May 2004), "The Green and the Gold Roads to Open Access," *Nature Web Focus*, available at: http://www.nature.com/nature/focus/accessdebate/21.html.

8. Lynch, Clifford A. (2003), "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age," *ARL: A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC*, no. 226, available at: http://www.arl.org/newsltr/226/ir.html.

9. Nixon, William J. (2002), "The Evolution of an Institutional E-Prints Archive at the University Of Glasgow," *Ariadne*, no. 32, available at: http://www.ariadne.ac.uk/issue32/eprint-archives/.

10. Silver, Judith (2003), "Movie Day at the Supreme Court or 'I Know It When I See It': A History of the Definition of Obscenity," *FindLaw*, available at: http://library.findlaw.com/2003/May/15/132747.html.

11. Suber, Peter (2005). "Open Access Overview: Focusing on Open Access to Peer-Reviewed Research Articles and Their Preprints," available at: http://www.earlham.edu/~peters/fos/overview.htm.

12. Ware, Mark (2004), *Pathfinder Research on Web-based Repositories*, Publisher and Library/Learning Solutions, London, pp. 3, available at: http://www.palsgroup.org.uk/palsweb/palsweb.nsf/0/8c43ce800a9c67cd80256e370051e88a?OpenDocument.

Citation for Published Version

Charles W. Bailey, Jr., "The Role of Reference Librarians in Institutional Repositories," *Reference Services Review* 33, no. 3 (2005): 259-267. http://dx.doi.org/10.1108/00907320510611294.

Bailey, Charles W., Jr. "The Role of Reference Librarians in Institutional Repositories." *Reference Services Review* 33, no. 3 (2005): 259-267. http://dx.doi.org/10.1108/00907320510611294.