

Blowing Your Own Horn - Institutional Repositories

“To *blow your own horn* is to be a braggart or "blowhard." This expression, arising in the American West about the middle of the 19th century, derives from an earlier expression, *blow your own trumpet*, dating back to at least 1576 and probably originating in medieval times, when heralds blew trumpets to announce the arrival of the king. Of course, any merchant or other commoner who wanted to announce his arrival had to blow his own horn.” (Liddiard, 2007)

Change is the norm in academic libraries today and academic libraries can use this impetus to reposition themselves to the center of the academic community by being the change agents in marketing academic accomplishments through institutional repositories. There are numerous signs for change in preserving academic content. Faculty members connect with researchers world-wide using email, creating ad hoc communities of scholars to work out solutions and problems together. In every discipline, content is created and shared digitally and online editions of journals are commonplace with back issues being digitized. *Google Scholar* and other search engines are giving high ranking to academic papers gleaned from institutional repositories (Association of Research Libraries, 2007). We now realize that the long term value of faculty writing is as digital assets and the ability to repurpose them for different uses, audiences, and situations (Yakel, 2004).

Conceptions of IR and Examples

Since the advent of IR (institutional repositories) in the 1990's with [ArXiv.org](http://arxiv.org) at Cornell featuring content from physics, math, computer science and quantitative biology, development contributions have been advancing rapidly. In 2000, the DSpace open source project was developed by MIT in cooperation with Hewlett-Packard and blazed the trail for more institutional development. Academic communities started to think about what the purpose of an IR was and what content was to be included. Some IRs are all-encompassing, including all the events of campus life from students plays to lectures and colloquia. Other institutions confine their collecting to documents that are they are legally mandated to retain while others collect only faculty, staff, and student writings. The larger institutions developed their IRs early on and smaller libraries became the “have nots,”

scrambling to develop consortial arrangements that capitalized on economy of scale and reduced technical system redundancy (Genoni, 2004). Standards for consortial custody are being drawn up by the Research Library Group (RLG) in a task force to address certification requirements, the certification process, and the certifying body to certify the consortium's ability to manage digital assets for the long term and consequences of failure for the digital repository (Yakel, 2004).

The pressure to institute IRs was increased by a shift in academic publishing methods and attitudes. Research conducted in 2005 indicated that journal subscription rates had risen an average of 8.5 percent since 1986 with library budgets rising only 5.6%. Articles found on the internet were cited 336% more than articles that were available only through subscription. Gunther Eysenbach of the University of Toronto also found that online articles were cited earlier and more often than those in traditional publishing cycles (Buehler & Boateng, 2005). Organizations like CreateChange now are beacons to an increasing number of authors who are becoming aware of copyright restrictions imposed by publisher agreements. Many universities encourage faculty to retain full control over their copyright when submitting scholarly work to publishers, to retain electronic rights, and grant rights to the university. Faculty are aided by librarians preparing standard letters of release for participating faculty to use to substitute for publishers' standard copyright releases (Westell, 2006). Faculty, nervous about violating copyright, are turning to services like the SHERPA/RoMEO Publisher Copyright Policies and Self-Archiving database to check on individual publishers. This particular service uses a color-coded system and is searchable by journal title. Another selling point is that links to articles and documents are more often kept intact in an IR than a general online service (Bell, Foster, & Gibbons, 2005). With all this support, though, Cornell faculty is still not moved toward new methods of archiving work, leaving 29% of the collections empty (Davis & Connolly, 2007).

The slow speed of the research cycle is in contrast to the rapid dissemination of research materials and their greater global visibility that has the possibility of raising scholars' recognition in academic circles and in-

creasing the prestige of the university. Often, some work and research ephemera do not fit into the traditional publication method but represent the process of discovery. The National Science Foundations' Blue-Ribbon Advisory Panel on Cyberinfrastructure states that "the traditional linear, batch processing approach to scholarly communication is changing to a process of continuous refinement as scholars write, review, annotate, and revise in near-real time using the internet" (Association of Research Libraries, 2007).

There are compelling reasons to support development in IRs. They provide more direct access to scholarly materials by sharing in open access archival efforts rather than plodding through the system of serials' publishers and vendors. IRs preserve digital content, not just for published articles already peer-reviewed, but for work that never goes beyond an individual's desktop like student Capstone Projects and raw research data. IRs increase institutional visibility and prestige by showing the world the academic work of the learning community, an indicator of academic quality. Funding sources and potential new faculty, staff, and students are attracted to the quality of work disseminated through IRs. Authors published in an IR have higher output and visibility when publishing away from the crowded field in their discipline (Johnson, 2002).

The library as leader in the institutional repository movement is reflected in the library literature. Articles by Genoni, Chang and Smith, and Jenkins all reflect on the library as the leader in the academic community positioning and proclaiming these activities as the library's new role. Librarians and library staff members can encourage members to deposit material into the IR, provide advice about copyright and embargo policies for their materials, help convert material to HTML or PDF formats, deposit materials directly into the IR for authors who can't or do not wish to self-archive, and review the metadata of content that is self-archived to maintain the quality of the repository (Chang, 2003; Genoni, 2004; Johnson, 2002).

Institutional repositories share several key attributes. They are institutionally-defined, scholarly, cumulative and perpetual, scalable, open and interoperable. The collection could contain work generated from all con-

stituents on campus – faculty, staff, and students. IRs are often more diverse in content than the beginning subject-oriented IRs of aggregated works around a subject area. Some institutions start with their stellar faculty and don't worry that the collection is not representative. Agreements are hammered out guaranteeing that documents stay in the repository except for cases of libel, plagiarism, copyright infringement, or 'bad science.' Contributors have to be corralled to use accepted file formats to assure readability in the future (Johnson, 2002).

Early adopters of institutional repositories provide the benefits of experience in methodology, collection development, and technology. Examining the pioneering DSpace-driven MIT repository provides a clean standard interface for browsing by community, subject, author, and date. It has a search engine within each category or you can drill down into the offerings of a community or group. There is also a featured, list of recent submissions and a RSS feed of new additions. There is self-archiving protocol, explaining the gathering mechanisms for submissions (Massachusetts Institute of Technology, 2007). The Digital Library of the Commons at Indiana University features a comprehensive search engine with a keyword thesaurus always available. They are working on a 4,000-5,000 image collection of research and field trips and they feature a self-archiving mechanism complete with agreement document (Indiana University, 2007). Ball State University uses CONTENTdm software and displays featured collections on the main page of the repository. They offer a video about the repository that contains a great explanation about copyright, metadata, and the uses of digital objects in the classroom. IN their holdings there is video throughout using Windows Media Player. Items are located by browsing by category or by an advanced search engine and users may set up a My Favorites page. This project is supported by their library Friends group and an estate gift (Ball State University, 2007) The Knowledge Bank is a joint initiative of the University Libraries and the Office of the CIO at Ohio State University. Conceptualized in 2001 using DSpace, it was built as an institutional repository to preserve the history of all aspects of the university. The site has the ability to email notifications of new materials to interested users and extension of a simple record reveals the metadata page in full (The Ohio State University, 2007).

Through the list of repositories found at the Open DOARS website, smaller universities can be found as examples of how IRs can be scaled down. The Directory of Open Access Repositories can be found at <http://www.opendoar.org/countrylist.php>. With a total enrollment of 30,000 students, George Mason University in Fairfax, Virginia, has an explanation page documenting their process and a document outlining the reasons for contributing to a repository. This repository can be searched by title, subject, author's name, and academic community (George Mason University, 2007). Smaller libraries can build small scale and localized OAI-MHP-compliant (Open Archive Initiative Metadata Harvesting Protocol) repositories and fill it with faculty publications, theses, and dissertations or start with collections already hosted. Added value through user services makes the repository process at the Wendt Library at the University of Wisconsin-Madison more supportive, relieving librarians of some of the instruction for self-archivers. Using *BibApp* software to help authors at the ingestion stage, this software features user tags to supplement assigned subject headings, photographs of the authors, a citation analysis, a social network feed that connects authors to others with their interests, copyright instruction using SHERPA/ROMEO's color coding system, and it autogenerates all of this data to DSpace files. It is not publically online yet but holds promise for providing the social support self-archivers need (Larson & Vack, 2007).

Problems and Issues

The inertia of the fundamental structure of scholarly communication through the traditional publishing paradigm makes change slow because of the revenue, resulting profits, and intellectual property that publishers want to protect. With the shift to electronic publishing, faculty researchers can enjoy increased citations using standard indexing and search mechanisms from open access harvesting. The use of multiple search engines and discovery tools enables outside users to find the material giving the author enhanced professional visibility. With more visibility, there is less emphasis on the quantity of work produced resulting in quality scholarship. Librarians understand the benefits of open access sharing but faculty do not understand, or often care about, the

jargon about persistent URLs and metadata schemas. Faculty need individualized information and direct support. Academic librarians use the established library liaisons assignments or the system already devised to connect to departments and faculty members. Faculty are concerned that there is little incentive, it is too much trouble, and that the articles they have are already published. They are also concerned that there be a variety of levels of sharing for different types of users other than academic affiliates. To aid librarians working with faculty in soliciting academic writing, the librarians at the University of Rochester (NY) have developed a crib sheet, found at <http://docushare.lib.rochester.edu/docushare/dsweb/View/Collection-331>, to use when answering faculty questions (Bell et al., 2005).

When an institution makes a commitment to a repository, monies must follow that decision. There will be an increased staff load that will increase as the collection grows. It may require a revamping of position descriptions and some traditional tasks reexamined. Librarians must argue aggressively that this venture will raise the university profile, aiding both marketing and recruiting efforts for staff and students. The librarian's skills in relevant metadata, content organization skills, awareness of digital preservation, and self-archiving methods are needed for a successful repository. Librarians know how to support faculty in their archiving in the event self-archiving is too daunting or doesn't work (Devakos, 2006).

The role of the academic librarian will change to include increasing comfort with digital resources and recognizing the value of gray literature and how it is manifested in various disciplines. A list has been written by the librarians at the University of Rochester of what gray literature may look like by content area. It is found at: <http://docushare.lib.rochester.edu/docushare/dsweb/GetRendition/Document-16896/html> (Bell et al., 2005). Librarians did not expect to become scholarly publishing change agents. This work is tricky business because faculty are depending on the traditional system for advancement and their stakes are high. The reference librarians' relationships with patrons will not be that of providing information but will change to a sales pitch potential. This proactive shift may cause some discomfort for some librarians who are used to patrons coming to

them. The library's level of relevance and visibility to faculty will increase at a time when brick and mortar is competing with patrons' presence online as librarians move from the tradition of collector to the role of collaborator (Bell et al., 2005; Phillips, Carr, & Teal, 2005).

Software Considerations

Institutional repository development has spawned much research and consequently has developed several open source software products. While open source software may come with lack of customer support, online communities provide help for adopters and share code to solve problems or supply needs for repository builders. Open source software has an impressive list of early adopters, incorporates good functionality and features, guarantees the preservation of many common file formats, and supports Dublin Core metadata and controlled vocabulary. To eliminate the decay of online citations, the use of URLs-persistent network identifiers are used. Open source software also supports a workflow largely controlled by self-archivers that gives submitters a feeling of ownership and frees librarians from many of the tasks of archiving (Falk, 2003).

Using the chart found on the DigitalKoans blog (<http://www.opendoar.org/countrylist.php>) regarding usage of open source software, DSpace has a large share of the market now (Bailey, 2007). DSpace uses the Lucene search engine to handle text and document metadata (Falk, 2003). ePrints, developed by UK's University of Southampton, excels in pre-press works and is used by the Digital Library of the Commons at Indiana University and CalTech (Tennant, 2002). It is not surprise that commercial vendors have been active in software development. SirsiDynix has Hyperion Digital Media Archive and Horizon Digital Library (Sirsi's product before the merger.) Innovative has Metasource and Symposia and ExLibris has DigiTool. VTLS embraces open source FEDORA to create VITAL. This product is used by Oxford University Library Services (Pace, 2006). FEDORA is a general-purpose management layer for digital objects and containers that aggregate mime-typed datastreams to ingest, manage, and provide basic delivery of objects using a simple user interface. It can

be locally customized (Johnson, 2002). Secondary publishers and middleware companies are also active in the field. ProQuest created Digital Commons, powered by Berkeley Electronic Press. BePress is used by the Escholarship Repository of the California Digital Library for the University of California system (Pace, 2006). Grokker software, often employed in OPACs, is used by Stanford University Library to create visual displays of up to 2,000 items to create visual maps of the collection and its subtopics (Falk, 2003).

When choosing software for an IR, the back up strategy is crucial because a repository is billed as a venture for perpetuity. An adopter has to consider how the software handles metadata and that the data conforms to standard library practices. The ingestion process, or how items are moved from the author's desktop to the database, needs to be examined for ease and efficiency. Searching, navigation through the system, and the software's ability to handle a variety of file formats are aspects to be considered. The software must be flexible to respond to the institution's need to organize the materials as a hierarchical, subject-based, entity grouped, or format based arrangement to suit their needs as a learning community.

Dealing with metadata is a challenge for institutional repositories as it is with other online resources. A major question is how to provide search terms that make sense to the user of the repository who has to learn the idiosyncrasies of many different IRs while conducting research. To use controlled vocabulary, allow user tagging, or some combination of the two is a decision to be made. Some IRs provide a thesaurus or pull-down menu of subject headings of acceptable terms at the point of ingestion to aid self-archivers. If the controlled vocabulary process becomes too laborious, authors will leave the metadata blank.

Issues of copyright and culture play heavily in open access archives. Even though reform to scholarly publishing is promoted by moving to open access archives, one hundred years of a scholarly publishing plan that gives control to publishers is a long, known history. Faculty have concerns about the equivalence between IR and journal publishing and have a fear of disrupting relationships with publishers (Jenkins, Breakstone, &

Hixson, 2005). Faculty are troubled about who owns the material they put in the IR and that adding materials in the IR can actually impede other, more traditional publishing opportunities (Buehler & Boateng, 2005). The institution has to be able to respond to a publisher's request that an article be removed when it is published in their journal (Tennant, 2002). Faculty members may be averse to having the university seemingly endorsing their work. A mistrust of technology and doubts about the ability to preserve the work into perpetuity cause some authors to pause. To solicit faculty members to contribute to an institutional repository is to ask for risk-taking in a risk-averse setting.

Discipline related issues abound. Some disciplines enjoy pre-existing forms for sharing scholarly work. Some disciplines are performance-based or 3-dimensionally expressed and would require multimedia files to capture their work. Some disciplines have only a few publishers controlling the journals in the field. Some disciplines' professional organizations are failing to recognize or make progress in changing the process of scholarly publishing. Some authors are simply not comfortable sharing pre-press versions of their work without the traditional vetting of the material (Jenkins et al., 2005).

Promotion is an intense activity in mounting an institutional repository. Often, faculty do not have a clear sense of what an IR is, what it contains, and its governing structure. The library can lead by example by entering librarians' scholarly writings, presentations, instructional materials, statistics, and reports to test out the process. Retired faculty can be solicited to build the repository as well. When starting to contact faculty members start with already-published faculty and campus authors. Other promotional ideas are to highlight a research of the day on the library website with links to the repository, conduct seminars on campus, present at department meetings, and know the titles of "green" journals of which some allow self-archiving. The DSpace Federation offers reproducible promotional materials and ideas (Dill & Palmer, 2005). Word-of-mouth and targeted presentations were effective strategies at the University of Toronto when they implemented their DSpace repository in 2003 (Devakos, 2006). In a California Digital Library study of 2004 reported by Margret Bran-

schofsky, faculty need to hear about IRs seven times before they begin to think of it as a do-able, worthy technology (Buehler & Boateng, 2005).

Content-related issues focus around collection development and evaluation of the institutional repository. The institution must decide under whose domain this project is managed. The principal actors are usually the library in tandem with the instructional technology department. Tasks must be drawn, timelines developed, and workflow managed. IRs demand policies and mechanisms to outline how to manage the content and control document versions. As in all collection development processes, priorities for acquisition need to be established as well as the relationship of acquired writings to the collection as a whole. Evaluation of an institutional repository, though, may seem backwards from the traditional way of measuring use within the institution. In an IR, who is using the content and what categories are being accessed from users *outside* the institution is the more valuable data. The institution will have to decide whether to include pre-refereed drafts along with papers that are published in a peer-reviewed journal. Items will have to endure an embargo queue until a reviewer/librarian can assess the item for inclusion if the IR is to be restricted. Some IRs use a qualitative review process while others just verify the authors' relationship to the institution (Johnson, 2002).

Scholarly writings involve pre-work. "Gray literature" is preserved in some institutional repositories as evidence of an informal scholarly communication process that is valued. In the traditional publishing process, pre-work is discarded and dropped out of the process. Gray literature can be manifested as experimental data, pre- and post-prints of published materials, out-of-print materials, conference papers and presentations, working papers, journals, student work, finding aids for collections, web sites, audio files, images, slideshows, raw data from experiments, papers in support of grant applications, conference proceedings, surveys, technical reports, and newsletters (Genoni, 2004)(Jenkins). Again, librarians cannot collect what they do not understand. Librari-

ans working with institutional repositories must be familiar with the purpose of gray literature in the scholarly process.

Pushing the Past to the Future

Preservation strategies for institutional repositories will continually be honed. Harvard's Digital Repository Service offers three levels of preservation service that corresponds to an author's expectations for the future of the item. The choices made are an agreement of understanding regarding the integrity and ease of various document formats to survive migration. They have a systematic program for adding and maintaining projects in the repository to perpetuity. This repository, maintained by the IT department, is apart from the Harvard Digital Archives of professors' papers and theses managed by the library found at: <http://hul.harvard.edu/huarc/> (Harvard University, 2007a, 2007b). There are three strategies for preservation. A not very successful or practical one is to maintain obsolete technologies to read back document formats. Emulation is another way to read back documents but has not been practiced in this field. Migrating digital assets through successive hardware and software platforms is the most reliable strategy. To maximize success in migration, data must be saved in its purest ASCII form, or bit-stream formats. An IR has to prove the claim of preservation and organizations like SPARC help with guidelines for preservation (SPARC International, 2007).

Institutions making decisions to have institution repositories must have a clear vision and integrate it into institutional planning. Some universities have begun projects on a "pay as you go" model but an ongoing funding model costs about \$200,000 a year to assure sustainability. Opening the repository to cross repository searching tools – OAI harvesting – will assure the repositories' function as a promotion tool. Cornell University created an incentive to faculty authors to contribute to the repository. They compete for grants and have the results placed in the repository (Westell). Now, that's the way to toot your own horn *if* the culture adapts to new venues of promotion and tenure. Uncertainty will mute this horn until a new tune is compelling enough.

References

- Association of Research Libraries. (2007). CreateChange: Digital Scholarship. Retrieved March 7, 2007, from <http://www.createchange.org/digitalscholarship.html>
- Bailey, C. W. (2007). DigitalKoans: Open Access Repository Software Use By Country. Retrieved March 30, 2007, from <http://www.escholarlypub.com/digitalkoans/2007/02/27/open-access-repository-software-use-by-country/>
- Ball State University. (2007). Digital Media Repository. Retrieved March 30, 2007, from <http://libx.bsu.edu/>
- Bell, S., Foster, N. F., & Gibbons, S. (2005). Reference Librarians and the Success of Institutional Repositories. *Reference Services Review*, 33(3), 283-290.
- Buehler, M. A., & Boateng, A. (2005). The Evolving Impact of Institutional Repositories on Reference Librarians. *Reference Services Review*, 33(3), 291-300.
- Chang, S.-H. (2003). Institutional Repositories: the Library's New Role. *OCLC Systems and Services*, 19(3), 77-79.
- Davis, P. M., & Connolly, M. J. L. (2007). Institutional repositories: evaluating the reasons for non-use of Cornell University's Installation of DSpace. *D-Lib Magazine*, 13(3/4).
- Devakos, R. (2006). Towards User Responsive Institutional Repositories: a Case Study. *Library Hi Tech News*, 24(2), 173-182.
- Dill, E., & Palmer, K. L. (2005). What's the Big IDeA? Considerations for Implementing an Institutional Repository. *Library Hi Tech News*, 22(6), 11-14.
- Falk, H. (2003). Digital Archive Developments. *The Electronic Library*, 21(4), 375-379.
- Genoni, P. (2004). Content in Institutional Repositories: a Collection Management Issue. *Library Management*, 25(6/7), 300-306.

- George Mason University. (2007). Mason Archival Repository Service. Retrieved March 30, 2007, from <http://mars.gmu.edu/dspace/about>
- Harvard University. (2007a). Papers of Tenured Faculty and Senior Guidelines: Collected Guidelines. Retrieved March 16, 2007, from http://hul.harvard.edu/huarc/donating_01.shtml
- Harvard University. (2007b). System Development and Upgrade Projects for FY2007. Retrieved March 31, 2007, from <http://hul.harvard.edu/ois/projects/>
- Indiana University. (2007). The Digital Library of the Commons. Retrieved March 30, 2007, from <http://dlc.dlib.indiana.edu/>
- Jenkins, B., Breakstone, E., & Hixson, C. (2005). Content In, Content Out: the Dual Roles of the Reference Librarian in Institutional Repositories. *Reference Services Review*, 33(3), 312-324.
- Johnson, R. K. (2002). Institutional repositories: partnering with faculty to enhance scholarly communication [Electronic Version]. *D-Lib Magazine*, 8. Retrieved March 7, 2007 from <http://dlib.org/dlib/november02/johnson/11johnson.html>.
- Larson, E., & Vack, N. (2007). *The BibApp*. Paper presented at the code4lib Conference. from <http://code4lib.org/2007/larson>.
- Liddiard, S. D. (2007). Writer's Block: Origins-Instruments of Expression: Bells, Drums, and a Horn. Retrieved April 1, 2007, from <http://www.writersblock.ca/fall2000/origins.htm>
- Massachusetts Institute of Technology. (2007). DSpace at MIT. Retrieved March 28, 2007, from <http://dspace.mit.edu/>
- Pace, A. K. (2006). Technically Speaking. *American Libraries*, 37(8), 72-73.
- Phillips, H., Carr, R., & Teal, J. (2005). Leading Roles for Reference Librarians in Institutional Repositories. *Reference Services Review*, 33(3), 301-311.

SPARC International. (2007). The SPARC Institutional Repository Checklist and Resource Guide. Retrieved March 31, 2007, from http://www.arl.org/sparc/bm~doc/IR_Guide_&_Checklist_v1.pdf

Tennant, R. (2002). Institutional Repositories. *Library Journal*, 127(15), 28.

The Ohio State University. (2007). The Knowledge Bank at OSU. Retrieved March 31, 2007, from <https://kb.osu.edu/dspace/index.jsp>

Westell, M. (2006). Institutional Repositories: Proposed Indicators of Success. *Library Hi Tech News*, 24(2), 211-226.

Yakel, E. (2004). Digital Assets for the Next Milllenium. *OCLC Systems and Services: International Digital Library Perspectives*, 20(3), 102-105.