INSTITUTIONAL REPOSITORIES AND THE INSTITUTION’S REPOSITORY: WHAT IS THE ROLE OF UNIVERSITY ARCHIVES WITH AN INSTITUTION’S ON-LINE DIGITAL REPOSITORY?

BY DOUGLAS BICKNESE

ABSTRACT: Many college and university libraries have recently begun implementing or experimenting with “institutional repositories” and other on-line digital repositories that allow for the centralized management and sharing of locally produced electronic publications and records. This article reviews some of the literature advocating that libraries implement on-line digital repositories, and notes that the role of the campus archives is generally overlooked. The article then examines how institutional repositories could benefit college and university archives and argues that archivists need to be involved in their planning and design.

What is an On-line Digital Repository?

On-line digital repositories, which in some cases may also be called “institutional repositories,” “digital institutional repositories,” or “consortial digital repositories” have recently become a subject of interest among academic librarians and scholars in general. College and university archivists need to be aware of the recent developments in on-line digital repositories for two reasons. First, such knowledge will help archivists with the management of electronic records that may already be in their custody. Second, having a current knowledge of on-line digital repositories will enable the archivist to participate in institutional policy decisions that will inevitably have an impact on the archives itself. On-line digital repositories offer archivists the opportunity of affirming or reaffirming their role as a manager of the campus’ records and information.

On-line digital repositories of various scales have been in operation with varying degrees of success since the early 1990s. Simply put, on-line digital repositories are a way for an institution or consortia to bring together and preserve the intellectual
products of a laboratory, a department, a university, or even an entire discipline of study in an on-line environment. Most colleges and universities have numerous faculty members who have put information related to their research on a personal Web page. Likewise, many departments and research centers have published white papers, preliminary findings, minutes, reports, and other documents on departmental Web pages. Some disciplines have also taken the initiative to share preprints and other research data prior to their publication in peer-reviewed journals.

In the past several years, open source platforms such as EPrints, DSpace, Fedora, and other programs have been released. These allow digital resources to be organized, described, and most importantly, shared through metadata harvesting projects such as the Open Archives Initiative (OAI). These open source programs are not proprietary in nature, which reduces the risk of software obsolescence. Projects such as OAI will allow the contents of on-line digital repositories at numerous institutions to be searched in the same way that on-line library catalogs can be. The platforms can be run locally with few resources and are designed with interoperability in mind. DSpace and EPrints, the two most popular platforms, can be downloaded free of charge from the DSpace Federation (http://www.dspace.org/) and EPrints.org (www.eprints.org), although some customization by a person familiar with the programming languages of Java or Perl will be required with both programs.2

The collecting foci of on-line digital repositories vary but they can generally be divided into two models. One model focuses on collecting material from an entire discipline, such as high-energy physics. The other model focuses on collecting material from a single institution, such as an individual college or university, but will accept material from any of its departments, centers, or labs. This second model is sometimes referred to as an institutional repository.

The Interest in Institutional Repositories

The discussions appearing in scholarly and library literature have generally stressed two reasons to develop on-line digital repositories. The first is to provide better access to scholarly research in an on-line environment, especially to "grey literature", that does not appear in peer-reviewed journals. The other reason is to revolutionize the scholarly publishing paradigm. Archivists should be familiar with both of these motivations as they try to make a case for a seat at the planning table.

Scholarly Research

One of the most successful examples of an on-line digital repository is the discipline-based arXiv.org repository developed by Paul Ginsparg in 1991. ArXiv.org allows physicists to post digital copies of manuscripts related to high-energy physics prior to publication. Since its creation, arXiv.org has become a major source for current research on physics, computer sciences, astronomy, and many mathematical specialties. According to the Public Library of Science, "Today, more than half of all research articles in physics are posted to [arXiv.org] prior to their publication in conventional journals. In many fields, these "eprints" are the defacto publications of
The SARS outbreak of 2003 is cited as an example supporting the argument that open access to current research is important. Scientists posted their findings on SARS on-line as soon as possible, allowing the public health community to combat the epidemic in a faster and more coordinated method than would have been possible in a print environment.

The recent development of open source on-line digital repository platforms has made it possible for individual colleges and universities to establish such repositories at the institutional level as well. What distinguishes these from discipline-based repositories is that the former include material from all of the disciplines within an institution. Clifford Lynch, director of the Coalition for Networked Information, states that "fully realized [institutional-focused on-line digital repositories] ... will contain the intellectual works of faculty and students—both research and teaching materials—and also documentation of the activities of the institution itself in the form of records of events and performances and of the ongoing intellectual life of the institution. It will also house experimental and observational data captured by members of the institution that support their scholarly activities."

The institutional and the discipline-focused models overlap to some extent and local policies will need to be developed to decide if a university with a high-energy physics department will want to collect material in an institutional-focused program if its faculty are already depositing material in a discipline-based on-line digital repository that is maintained outside of their university. Some colleges and universities may want to do both as a way to promote the campus. The Scholarly Publishing and Academic Resources Coalition (SPARC) and other proponents of creating digital repositories at the institutional level believe that capturing a university's "intellectual capital" will serve as a meaningful indicator of that university's academic quality. According to a 2002 SPARC paper, "an [on-line digital repository] concentrates the intellectual product created by a university's researchers, making it easier to demonstrate its scientific, social and financial value." If publicized correctly, some have advocated, the increased visibility could even increase public and private funding as well. On-line digital repositories are also promoted as a way to provide better access to the "grey literature," the technical reports, theses, white papers, etc., that are created at each college and university. While librarians have had a challenging time acquiring and cataloging these non-commercial publications, these sources undoubtedly fill a substantial number of the acid free boxes that line the shelves in most college and university archives. Archivists should be ready to provide their knowledge about working with grey literature and provide information on both past and current use and users of this material.

Finally, SPARC and other advocates of on-line digital repositories view them as a way to revolutionize the way scholarly publishing is done in order combat the rising cost of serials. These advocates encourage authors to avoid publishing in journals with restrictive copyright policies and to deposit their final manuscripts and supporting data into open source on-line digital repositories. Institutional-based on-line digital repositories are sometimes promoted as a tool in this open source struggle against restrictive and expensive journal publishers. Archivists should be aware of this argument, as it may serve as a catalyst for encouraging the direction an on-line digital repository at their institution may take. Many issues in this new publishing paradigm are still being
resolved, such as the evolving process of peer review in relation to electronic publishing and issues involving copyright. However, in the long run, it will probably be the larger discipline-based on-line digital repositories such as arXive.org and PubMed Central that will have a greater role to play, rather than the smaller ones operating out of individual institutions.

Preservation and Storage

Most advocates of on-line digital repositories are well aware of the significant challenges involved in preserving electronic records and data for an extended period of time. However, they believe that by putting the electronic documents into an on-line digital repository, the information will be under the stewardship of an organization committed to preserving the data. Most on-line digital repository programs can be customized to limit what formats can be deposited into their holdings and can also require that specific metadata be created about the digital objects being deposited (such as what software and hardware may be required to view the digital object, as well as who created it and why). These metadata should make it easier to identify files in need of migration. Although some archivists may question how well institutions will keep up with these difficult migration issues, more information is likely to be preserved through this method than by relying on individual faculty members who are running servers out of their office. The two major programs currently used in on-line digital repositories (DSpace and EPrints) allow for a great deal of local customization. Most open source on-line digital repository programs can accept virtually any type of electronic record of any size, the only limits typically being the boundaries of the http protocol. Local administrators can set policies either limiting the format of material that can be submitted or allowing any type of file to be submitted. They can also adjust the metadata requirement to either the bare minimum required for Dublin Core, or mandate that numerous screens of metadata be created for each digital item in the on-line digital repository, or find a level in-between.

DSpace, which is currently one of the more popular on-line digital repository programs, organizes material on its servers into communities, subcommunities, and collections. According to the DSpace Federation, “DSpace Communities might be departments, labs, research centers, schools, or some other administrative unit within an institution. Communities determine their own content guidelines and decide who has access to the community’s contributions. An administrator on the DSpace team, usually the DSpace User Support Manager, works with the head of a community to set up workflows for content to be approved, edited, tagged with metadata, etc.” The collections then house the individual files for a particular community, although multiple communities can share the same collection. Eprints and many other on-line digital repository programs can be set up in a similar manner, although some local programming will be required. A community could be established for a college, with subcommunities for departments, centers, etc. Screen shots showing how the Massachusetts Institute of Technology used DSpace in this manner appear in this article. This hierarchical arrangement system based on provenance should be familiar to most college and university archivists, who have been arranging and describing material in a similar manner for decades, using record groups, subgroups, series, and subseries.
Example of a DSpace implementation. Note the ability to browse by title, date, author, or community.

An example of DSpace showing the use of subcommunities. A subcommunity can have its own subcommunity as well.
At the next level, collections within a subcommunity in DSpace.

An example of an Eprints implementation. Although the EPrints interface default is to allow browsing by subject headings, this page shows that it is possible to modify the interface to include organizational information as well.
Where Does the University Archives Fit In?

The role of the university archives in an institution’s on-line digital repository is not discussed at great length in the literature advocating the adoption of such repositories. Therefore, it is possible that local champions of institutional-based on-line digital repositories may not think to include the university archives in planning for such programs. Clifford Lynch points out that archives should be included in discussions on on-line digital repositories and recognizes that an effective program “represents a collaboration among librarians, information technologists, archives and records managers, faculty, and university administrators and policymakers.” Lynch goes on to make many strong arguments for institutions to establish on-line digital repositories, although he does not specifically address the role of archivists and records managers further. The role of the university archives is also briefly discussed in the SPARC position paper on institutional-based on-line digital repositories:

Depending on the university, a[n on-line digital 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 [on-line digital 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 to store. Still the potential overlap of roles of the two repository types merits consideration at institutions that support both.

By identifying a supposed dichotomy between an institution’s on-line digital repository and the university archives, SPARC’s statement implies that most archivists would either not be interested in the ability to preserve the entire intellectual output of their institution or would view such an effort as a threat to the archives’ continued importance. Nevertheless, many university archives have tried with varying degrees of success to collect the published and unpublished intellectual output of their faculty. The greatest challenge to collecting all of the publications of a university’s faculty is probably physical space, followed by difficulties in obtaining the commitment of the institution to follow through on agreements to deposit such material in the university archives. Even Leonard Rapport, in his classic article advocating deaccessioning, stated, “If storage, preservation, and servicing of records cost nothing … I would advocate saving a record copy of every document.” But since these efforts do have a cost, archivists have had to be increasingly selective about what they accept, especially with faculty papers. Being able to overcome the problem of space to provide better documentation of an institution’s intellectual output should be of interest to any college and university archivist.
Trying to preserve the entire intellectual output of an institution is a noble goal for an on-line digital repository, although realistically some appraisal of its contents will eventually be required. This is where archivists need to be ready to offer their expertise. Institutions have the ability to set the bar for admitting material at a number of levels. The institution could rely entirely on the judgment of faculty and allow them to submit anything, or they could set the bar very high and allow only designated (and presumably trained and screened) individuals to be able to submit digital files. Because of this, the types and extent of material found in an on-line digital repository could vary substantially from one college or university to another, or even within a single institution itself. Archivists’ experience in selecting records of enduring value should be of interest to committees trying to decide what is appropriate to allow into the on-line digital repository.

Once material has been deposited, decisions will need to be made about if and when it should be removed from the on-line digital repository. Archivists could take an approach along the lines of that advocated by Hillary Jenkinson and conclude that if the creating faculty member has deemed the file important enough, there is little need to question its presence in the digital repository. However, unless there is a rigorous selection process for what is deposited in the on-line digital repository, it is very likely that some appraisal of deposited material may be necessary over time. Local committees at each individual institution will need to decide what is appropriate to maintain in their on-line digital repository, and archivists should be ready to provide their expertise in appraising faculty papers, office records, data sets and other electronic information to help decide what should be retained in their institution’s on-line digital repository.

Archivists need to have a key role in these committees to share their expertise in collecting faculty papers and in working with researchers who use faculty papers. They need to explain how their experience in appraising faculty papers will help with both recruiting faculty contributors and identifying specific material that will be worth the institution’s investment.

The SPARC position paper also simplifies the mission of most university archives, which as Helen Samuel articulated in *Varsity Letters*, can be very comprehensive in its coverage. Samuel’s functional approach advocated the documentation of how the university confers credentials, conveys knowledge, fosters socialization, conducts research, sustains itself, provides public service, and promotes culture both to the campus and the surrounding community. Any archivist attempting the institutional functional analysis advocated by Samuels would be interested in better documenting both the traditional faculty publications as well as the research that supports those publications.

The interests and expertise of college and university archivists overlap with those of advocates for on-line digital repositories to a much greater degree than most people in the institution may be aware. While the library literature has extensive use and citation studies, few of these look specifically at how archival records are used, let alone at what specific types of records are being used. If the university archives has a reputation (rightly or wrongly) of being a place where material is kept mainly for its intrinsic value and not because of its informational value, the archivists’ input may not be solicited. Archivists need to provide information not only on the long-term
preservation of electronic records deposited in the on-line digital repository, but also
to explain how their experience in areas such as selection, appraisal, and description
will be of value to the institution's on-line digital repository program. This is an op-
portunity for archivists to share their knowledge about the users of both faculty papers
and institutional records as well as about the types of records those users find most
useful. This knowledge can also be helpful in targeting faculty and departments to
recruit for contributing to the institution's on-line digital repository as well.  

What About Nonscholarly Material?

While SPARC, Lynch, and others focus on the ability of on-line digital repositories
to share scholarly information, and potentially to change the scholarly publishing
paradigm, the value of an on-line digital repository in documenting the history of an
institution is generally overlooked. Archivists should take every opportunity to push
the institution to expand the focus of its on-line digital repository to include as much
truly institutional material as possible. If the institution really wants to document itself,
it needs to go beyond simply capturing scholarly output, and also find ways to capture
and share its nonscholarly material as well.

Every campus has dozens, if not hundreds of regular reports, newsletters, magazines,
and other official and unofficial publications created by its administrators, faculty, and
students. Archivists need to explain how valuable these resources are in identifying
such matters as when the curriculum changed, when key events took place, the date of
a particular faculty member's hiring or retirement, or the recipient of a particular award
of excellence. Every academic archivist is also aware that while the transition to a
"paperless" office may be far from complete, it is beginning to take root. Major reports,
publications, departmental newsletters, and key announcements increasingly appear
only in electronic format. These records are now distributed by Web pages, Listservs,
and E-mail. Additionally, it is increasingly easier for end users to create digital audio
files, digital video files, and other nontextual records. These may include interviews,
oral histories, digital images, speeches, presentations, as well as more ephemeral
material. These digital audiovisual files are often scattered across the campus on both
streaming servers as well as on removable media such as tapes, CDs, and DVDs.

In many campuses, these important electronic records are simply maintained on the
Web space or desktops of the creating office, college, or department. One could take
a noncustodial approach and rely on the individual offices or departments to maintain
these Web-based documents. However, archivists certainly have encountered situa-
tions where university personnel have destroyed valuable departmental records while
getting rid of their office files. The commitment of the office of origin to preserving its
older records can vary substantially across the campus, and the risk of older records
not being migrated or simply being deleted to avoid having to acquire new server space
is very real. Indeed, even if the creating office has the intent to preserve its on-line
files, human error or technological failures may cause the loss of key documents. The
accidental deletion of the entire contents of the Web site for the prominent Pittsburgh
Functional Requirements for Record-keeping Project should be enough warning to most
archivists.24 Due to the high degree of autonomy present at many colleges and universities, information kept on locally maintained servers may have backup procedures that vary significantly, if they exist at all. If the campus archives desires physical control of any of its institution’s electronic publications at a central location, the archives must either print out a copy or set up a complicated procedure that captures and maintains the publication on a mirror server under the archives’ control.

The current surge of interest in on-line digital repository programs presents the campus archivist with an opportunity to access a cheap and relatively simple way to consolidate both the official nonscholarly as well as scholarly publications of an institution at a central virtual location. If departments are required to submit key documents such as reports, minutes, newsletters, etc., to an on-line digital repository, this would help ensure that the material will be preserved over time and that it will be available for searching or browsing from a single location. Once a publication is in an on-line digital repository, it will be easier for the archives to work with the data stewards to ensure that the material remains accessible, reliable, and authentic. The files will all have at least a minimal amount of metadata, will be organized by provenance using the “communities” feature of programs such as DSpace, and will be in formats that the institution has decided to support. This will make it easier for the both archivists and data stewards to prioritize migration projects as software becomes obsolete.

**What About Electronic Records Already in the Archives?**

One other potential use for on-line digital repository programs is as a storage site for electronic records that arrive with hybrid collections of both paper and electronic documents. By using the digital repository’s ability to arrange material by communities and subcommunities, the archivist could easily create a community for the university archives, a subcommunity for electronic collections, and named collections containing the specific electronic files that have been deposited in the archives on removable media such as floppy disks or CDs. By being present at the design and implementation of the on-line digital repository, the archivist could also ensure that the system contains sufficient metadata to document the provenance of the collection. Of course, this should be recognized as only an interim solution aimed at addressing the problem of format obsolescence. Nevertheless, by collecting the correct metadata, it will then be easier for the archivist to work regularly with the information technology personnel to make better-informed decisions on what files, if any, need to be migrated to new software formats. Indeed, it may even be possible for the archivist to create reports documenting the number of files stored and their various electronic formats in order to help with planning long-range electronic records preservation.
Conclusion

Because of the flexibility of on-line digital repository programs, policy decisions on what is appropriate for each such repository will be made locally. College and university archivists bring an important perspective to a committee planning an on-line digital repository for their institution. Archivists should be ready to provide input based on their experience working with similar tangible material in the campus's university archives, and if possible assume a leadership role. Archivists' expertise in selection, appraisal, and preservation, as well as their knowledge of the institution's organization and history, bring an important perspective to any institution planning on capturing its intellectual output in an on-line digital repository. Finally, archivists should advocate expanding any developing on-line digital repositories to include both scholarly and nonscholarly material that will help document and preserve the institution's history.

ABOUT THE AUTHOR: Douglas Bicknese has been serving as the university archivist at the University of Illinois at Chicago since 1998. He has an M.I.L.S. degree from the University of Michigan and an M.A. in history from the University of Illinois at Chicago.

NOTES


3. Grey literature is defined by Vilma Alberani, Paola De Castro Pietrangeli, and Anna Maria Rossi Mazza as "non-conventional, fugitive, and sometimes ephemeral publications." Archivists who collect faculty papers should be very familiar with grey literature, as Alberani, Pietrangeli, and Mazza explain that it may include, but is not limited to the following types of materials: "... reports (pre-prints, preliminary progress and advanced reports, technical reports, statistical reports, memoranda, state-of-the art reports, market research reports, etc.), theses, conference proceedings, technical specifications and standards, non-commercial translations, bibliographies, technical and commercial documentation, and official documents not published commercially (primarily government reports and documents)." Collecting and cataloging these nonconventional publications has been a challenge for librarians and increased their interest in ideas such as on-line digital repositories.


5. For an example of this, see “Scientific Publishing Picks up Speed,” Canadian Medical Association Journal 168:13 (2003), http://www.cmaj.ca/cgi/content/full/168/13/1637 (19 September 2004).


7. SPARC, the Scholarly Publishing and Academic Resources Coalition, was developed by the Association for Research Libraries and is an alliance of academic and research libraries and organizations working to address issues in the scholarly publishing system. See their Web page, at http://www.arl.org/sparc/, for more information.


9. A full discussion of peer-review and copyright issues associated with on-line digital repositories is beyond the scope of this article. Some currently advocate for an author’s right to “self-archive” material in on-line digital repositories, while others promote more aggressive strategies to circumvent restrictive journal policies by putting the preprints and the corrigenda containing the peer-review comments on-line along with information about how the author used the corrigenda to turn the preprint into the published version. For more information, see the SPARC Web site, http://www.arl.org/sparc (24 October 2004), as well as the Budapest Open Access Initiative, http://www.soros.org/openaccess/software/ (24 October 2004), and Steven Harnad’s “For Whom the Gate Tolls: How and Why to Free the Refereed Research Literature Online through Author/Institution Self-Archiving, Now,” http://www.ecc.soton.ac.uk/~harnad/Tp/resolution.htm#Harnad/Oppenheim (24 October 2004).


11. A recent article by Joe Staub pointed out that storage requirements for electronic files is growing 100 percent annually, while the cost/performance of computer storage media is growing only 35 percent annually. See Joe Staub, “The Digital Tsunami,” Information Management Journal (January–February 2004): 42–43. David Stephens and Roderick Wallace indicated in a 2003 publication that while database sizes are approaching a petabyte (one thousand terabytes), data transfer technologies are not keeping up. Given current transfer rates, it could take a decade to move a petabyte of data onto new media for preservation. See David Stephens and Roderick Wallace, Electronic Records Retention: New Strategies for Data Life Cycle Management (Lenexa: ARMA International, 2003): 64–69.

12. Dublin Core is a metadata standard established by the Dublin Core Metadata Initiative (DCMI). It is commonly used by many organizations to facilitate access to Web-based digital objects. See the Dublin Core Metadata Initiative for more information, at http://dublincore.org/ (23 October 2004).


19. Recent examples of how college and university archivists have struggled with this can be found in Tom Hvry, Diane Kaplan, and Christine Weideman, “‘Though This Be Madness, Yet There Is Method in ‘t’: Assessing the Value of Faculty Papers and Defining a Collecting Policy,” American Archivist (Spring/Summer 2002): 56–70, and Tara Zachary Laver, “In a Class by Themselves: Faculty Papers at Research University Libraries and Manuscript Repositories,” American Archivist 66:1 (2003): 159–196.

21. Many universities are having trouble getting faculty to submit information to their on-line digital repositories and are trying to recruit faculty to help build up a critical mass of material. See Nancy Fried Foster and Susan Gibbons, “Understanding Faculty to Improve Content Recruitment for Institutional Repositories,” *D-Lib Magazine* 11:1 (2005), http://www.dlib.org/dlib/january05/foster/01foster.html (10 May 2005).


23. Mark Greene has argued that with business records, internal published sources are used much more frequently than unpublished sources. See Mark Greene, “‘The Surest Proof’: A Utilitarian Approach to Appraisal,” *Archivaria* 45 (1998): 127–169. It is possible that, excluding photograph collections and biographical files for faculty and alumni, the same situation exists for college and university archives.

24. Visitors to the official page for the Functional Requirements for Evidence in Recordkeeping, at www2.sis.pitt.edu/~rcox/FunReqs.htm, learn that “Due to a technical glitch at the School the Web site with the working files of this project was destroyed.”
ALTER YOUR CAREER PATH OR ENRICH AND STRENGTHEN YOUR SKILLS!

CAS CERTIFICATE OF ADVANCED STUDY
ARCHIVES AND RECORDS ADMINISTRATION

UWM's CAS in Archives and Records Administration provides advanced coursework for working professionals who hold a MLIS or related Master's degree. Students take 15 credits of coursework to develop their specialty in Archives and Records Administration.

ON-LINE COURSES
The Program may be completed at UW Milwaukee or entirely on-line while working and living anywhere with internet access. Regardless of location, students pay in-state tuition plus a technology fee charged for on-line classes.

UNIVERSITY OF WISCONSIN
UWM MILWAUKEE

SCHOOL OF INFORMATION STUDIES

FOR MORE INFORMATION
Go to our website at: www.uwm.edu/Dept/SOIS/academics/cas.htm
Or contact the Archives Program Coordinator: Amy Cooper Cary
Ph: 414-229-6929   Email: amycary@uwm.edu