Institutional Repositories: Keys to Success

Joan Giesecke
University of Nebraska-Lincoln, jgiesecke1@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/libraryscience

Part of the Library and Information Science Commons

Giesecke, Joan, "Institutional Repositories: Keys to Success" (2011). Faculty Publications, UNL Libraries. 255.
http://digitalcommons.unl.edu/libraryscience/255

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Institutional Repositories: Keys to Success

Joan Giesecke
Dean of Libraries, University of Nebraska-Lincoln

Abstract
Institutional repositories are a relatively new activity for higher education. They are defined most often as a set of services that are offered by an institution for the management and dissemination of digital materials created by the members of the institution or scholarly community. This article will describe the challenges institutions are facing in creating repositories, will explore the economics of managing repositories, and will offer a model for creating a successful set of services.

Keywords: economics, institutional repositories

“Why would anyone want to make their publications available through an institutional repository?” asked one of the deans when we explained our efforts to create an institutional repository at the University of Nebraska-Lincoln in 2005. Four years later over 40% of the faculty had publications in the institutional repository. Over 44,000 full-text documents, monographs, dissertations and theses are available in this open access repository. What had changed? How did we go from, “Why do this?” to “Why aren’t my publications in the repository?” To answer this question it helps to look at the history and economics of open access institutional repositories and what elements are needed to create a successful, cost-effective repository.

Overview of Institutional Repositories

Institutional repositories, online archives of locally produced scholarly works for the purpose of preserving and disseminating research, are a relatively new activity for higher education institutions. The development of repositories follows the development of the Internet and the World Wide Web. The first dis-
Discipline archive launched in 1991 was the physics repository now known as arXiv. The physics repository, begun by Paul Ginsberg, started as a server for articles on theoretical physics. As the Internet developed, the archive grew and expanded to include other areas of physics, math, and computer science. Other early repositories included EconPapers, CogPrints, and PubMed Central. In 1999, the framework for developing interoperable archives was developed and became the Open Archive Initiative. In 2001, the University of Southampton launched EPrints, a software package supporting open access deposit of research materials. Massachusetts Institute of Technology (MIT) followed in 2002 with the launch of DSpace to manage self-archiving of research articles by MIT faculty. Commercial products and additional open access systems followed as libraries and universities became more interested in the idea of creating repositories for faculty work (OARINZ, 2009).

Institutional repositories are also described as “a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members” (Lynch, 2003). Lynch further notes that essential services include “the management of technology changes and the migration of digital content from one set of technologies to the next as part of the organizational commitment to providing repository services” (Lynch, 2003). The repository then is a system for dissemination and stewardship of the intellectual life and scholarship of an institution. It becomes a new way for the institution to contribute to the broader world of scholarship. The repository is cast as a new way to do collection development, to expand this function from the identification and purchase of published materials, to the gathering and dissemination of the works of the faculty. The system is seen as collaboration among libraries, technologists, administrators, and faculty to enhance access to the scholarship of the institution. Lynch also included cautions for those developing institutional repositories. He noted that repository efforts would not succeed if institutions used them as a way to exercise control over faculty work, were overburdened with policies meant to limit submissions or serve as gatekeepers for admitting material, or if institutions implemented repositories because it was fashionable to have a repository but were not willing to make the administrative commitment needed to support the repository (Lynch, 2003).

In 2006, the Association of Research Libraries published a survey on the efforts of member libraries to develop institutional repositories. At that time, of those who responded to the survey, 43% had an institutional repository, 35% were planning a repository, and 22% were not planning one at that time. The authors noted that libraries were the driving force in the creation of institutional repositories with support from information technology units. Planning time ranged from six months to a year to begin a repository. Most institutions conducted a pilot project before making their repositories open to the public. Planning groups were used in most libraries to design the implementation pro-
cess. Staff chosen to work on the repository usually were chosen for their technology or metadata skills. Major challenges identified by respondents in developing repositories were problems in recruiting content, finding appropriate staff, and obtaining faculty buy-in for the program. Adequate funding, developing workflows, and resolving copyright issues were also discussed as problems (Bailey et al., 2006).

Fast forward to 2010 and there are over 1,900 repositories in the world and over 300 at U.S. institutions according to the Registry of Open Access Repositories. These repositories differ widely from just a few thousand items to over 45,000 research documents. Repositories serve a variety of purposes. For the institution, the repository can raise the visibility of faculty research, help preserve the intellectual output of the institution, and, particularly for public institutions, can be an effective way to share research with their constituencies. For the producers of the research, the faculty, the repository is a way to disseminate their research within their peer community and to increase citations to their work. These two major stakeholders, the institution that supports the repository and the researchers who create the works, have therefore different viewpoints and needs for how a repository should be structured. From the library world, repositories may be seen as a way to address some of the economic challenges of obtaining access to scholarly works. As subscription costs increase at rates higher than inflation, and libraries face continuing budget reductions and challenges, open access repositories can help provide access to research findings. The challenge, however, is still how to identify articles that are available full-text from an institutional repository.¹

**Challenge of Populating Institutional Repositories**

One of the challenges for libraries managing institutional repositories has been convincing faculty to self-archive their work in the repository. A variety of research studies have been done examining this challenge. Opinion pieces on blogs and Web sites also propose solutions for how to obtain faculty cooperation in creating an institutional repository. Four basic models have emerged from these discussions. First is the, “If you build it, they will come” model from the phrase from the motion picture *Field of Dreams* (Gordon, 1989). This model has been shown to be unrealistic. Unlike the ghosts of ballplayers coming to the cornfield, faculty are not flocking to support institutional repositories. They do not come to play.

A second model is the approach Tom Sawyer took to get the fence painted. Try to make depositing articles seem fun and attractive. Unfortunately, this model also has not appealed to many faculty. Even those who do come to help “paint the fence” may make a mess of the project. This will be noted later in the research on faculty creation of metadata.
A third model is that of mandates. A number of prestigious and well known universities and colleges within universities have passed mandates requiring participation in the repository. Again, mandates may help but faculty do not always respond to mandates and deposit of articles remains slow even at institutions with such policies as noted in reviewing the statistics provided by the Registry on Open Access Repositories.

The fourth model is that of providing services. This model reflects Cliff Lynch’s description of a repository. Services include technical support, preservation services, and metadata services, among others. Unfortunately none of these models have proven to be universally effective in generating a large number of deposits by faculty. A number of research studies have been completed to try to determine what might be most effective as a way to encourage participation in an institutional repository.

Jingfeng Xia reports on research that studies whether scholars familiar with depositing articles in a subject based repository would be more enthusiastic or more likely to deposit articles in an institutional repository than scholars who did not have this experience (Xia, 2008). The study found that “faculty authors are indeed not enthusiastic about archiving their articles into their institutional repository even though they are familiar with self-archiving practice” (Xia, 2008). Further, the author found that for physicists in the study, “data collected from self-archiving in a subject repository and an institutional repository seem to not support the subject culture theory that scholar’s familiarity with self-archiving through practices in a SR will ensure enthusiasm about self-archiving in the repository of their own institution” (Xia, 2008). Some scholars who were enthusiastic about their institution’s repository did not participate in the subject repository while others exhibited the opposite behavior.

Morag Mackie, in describing the Daedalus project at the University of Glasgow, outlines the strategies they have used to populate the institutional repository including doing the archiving for the faculty. Strategies included holding meetings with faculty to discuss open access and self-archiving of publications, searching faculty Web sites for full text articles posted on the sites, checking copyright agreements for faculty before archiving, identifying journals that allow archiving, and contacting faculty with articles in those journals for permission to post their articles. Still, in 2004 when this study was published, faculty participation was relatively low and the strategies being used were somewhat time-consuming. Future strategies reported in the article included proposing a system to search a newly developed university-wide annual database report on faculty publications and posting the full text of articles, where possible, without having to contact each author individually (Mackie, 2004).

Brian Quinn (2010) approached the question of how to increase faculty participation in a repository by examining psychological resistance to digital repositories. He summarizes the literature on user behavior and digital repositories not-
ing that most studies found that faculty did not see value in depositing articles in a repository, did not feel they had the time to deposit articles, and were reluctant to learn and relearn yet another technology system that they would not use very often. Quinn suggests that looking at the psychology of resistance can help librarians develop more effective strategies for encouraging faculty to deposit articles in a repository. He concludes that mandates alone will not overcome psychological resistance to participation (Quinn, 2010).

In addition to encouraging faculty participation, another challenge for those managing institutional repositories is ensuring the accuracy of self-archived records. This challenge is particularly evident in the creation of metadata for the items in the repository. Although most archiving systems have instructions on how to create the basic metadata for a record, studies are finding that faculty are not necessarily very accurate in filling in these fields. For example, Mary Kurtz compared Dublin Core data in three institutional repositories that use DSpace software. She found that contributor generated metadata such as subject headings, description or abstract, publisher, and even language was either missing or inaccurate. This lack of consistency in the metadata complicates the process of searching the repository for information. She concluded that the review of metadata by librarians is a more effective process for ensuring accuracy in the record creation process than leaving metadata creation solely to the contributor (Kurtz, 2010).

Costs of Institutional Repositories

Establishing an institutional repository is not a cost free proposition. Estimates for the cost to an institution for establishing a repository range from over $130,000 per year to over $248,000 per year at MIT. Factors that impact costs include the number and type of staff, the type of technology chosen for the repository, the services provided, and the cost of preservation of data. One of the first decisions for an institution is to choose the type of hardware and software for the project. Open source software systems provide an institution with the ability to customize the program and develop a system that meets local needs. It does mean however that the institution will need programming and systems staff to run the system. Choosing a commercial software program can limit the amount of technical staff needed and may limit the amount of customization that can be done. Other technology costs include costs for digitizing content or the hardware and software needed for such services, costs for backup systems, and costs for digital storage (Blythe & Chachra, 2005; Houghton et al., 2010; McGovern & McKay, 2008).

Once the software platform is determined, the institution can determine the staffing needed to run the repository. Staffing will include those with direct re-
responsibility for the daily operation of the services and those who have new responsibilities added to their positions to support the service. The latter include staff who may take on marketing roles, staff who contribute metadata, and staff who provide training. Generally, repositories have 2–3 FTE staff assigned to the unit. Operational costs include costs for marketing materials such as brochures, supplies, costs for software upgrades, and hardware replacement costs. With these categories of costs in mind, an institution can determine the additional costs it will incur by adding an institutional repository to its set of services.

Libraries may also need to determine the cost to scan materials for the repository, particularly if the institution decides to digitize dissertations and theses for the repository. Institutions may decide to outsource this work or may see this as an opportunity to develop in-house skills. The University of Massachusetts Medical School chose to use digitizing dissertations as a way to gain experience and to begin to build an institutional repository. They provide an extensive breakdown of the costs and time needed to digitize 320 dissertations. They found that it cost approximately $.28 per page to digitize and process the dissertations. They also noted that it took approximately 170 minutes per title to do the work. Their report provides data others can use in determining how to approach the question of in-house production versus outsourcing digitization of documents for a repository (Piorun & Palmer, 2008).

Another set of costs for the institution to consider is the cost for self-archiving articles. These costs, borne by the authors of the content, are often missing in the reports on institutional repository costs. In a study of alternative scholarly publishing models published by the Joint Information Systems Committee in 2009, the authors estimated that each article archived by a faculty member cost the institution approximately $14.90. They also estimated the ongoing costs per year for a repository averages $159,000. If an institution’s faculty deposited 5,000 articles in a year, the institution would incur an additional $74,500 (Houghton et al., 2009).

Although open access institutional repositories are cited as an alternative to purchasing expensive scholarly journals, repositories come with their own costs that are additions rather than substitutions for current costs to the institution. To sustain repositories, institutions are reallocating funds or identifying new funding that can be used to develop and maintain the open access database.

Old Wine in Different Bottles, A Different Viewpoint

Institutional repositories, as has been noted, developed out of a collection centric viewpoint which assumes that the creation of repositories is an extension of the traditional roles of collecting and organizing published or completed projects. Original designers of these systems assumed faculty would have a digital copy of a completed manuscript that they could deposit in the repository. Logically,
this assumption makes sense when talking about preprints, not yet published articles, and working papers. In some ways this is the model of the physics archive. However, when institutions want to expand beyond these working documents to the published content of the faculty of their institution or the faculty in the discipline, they encounter a number of problems. One problem that needs to be addressed is the challenge of disentangling copyright permissions. Understandably, faculty are often unaware of whether or not they can post an article in an institutional repository. Some publishers allow deposit of the final, published version of the article. Some allow the deposit of the author’s version, pre-copyediting, and final publication. Some allow no deposit at all. And some articles are in the public domain but the author may not realize this loophole. For example, articles authored by federal employees cannot be copyrighted. This means that any article that is co-authored with a federal employee can be posted. Many faculty may be unaware of this wonderful opportunity for repository deposit.

A second concern is that faculty may try to self-archive files that are badly scanned, or over scanned so that the files are very large, or may leave out co-authors and abstracts. That is, faculty are not necessarily posting a file that meets quality standards. These files then need to be corrected and improved if an institution wants to ensure a quality repository that others will use.

A third concern is that faculty may not know how to describe their work in a way that will increase the chances of the article being discovered by search engines such as Google. Providing appropriate key words and including a meaningful abstract can increase the chances of users identifying and then citing or using the faculty research.

One way to address these concerns is to recast the institutional repository as a publishing venture rather than as a set of services or to put the old wine (faculty research) into the “publishing bottle” instead of the “library collection bottle.” When a library begins to think as a publisher, the approach to creating a repository changes. The manager of the repository needs to have a publishing background rather than a technology or library science background. The manager needs to understand how publishers solicit works, to know how manuscripts become finished products, to understand the world of copyright and permission, and to understand how the world of online dissemination functions. With these skills, a repository manager working with librarians and technology staff can create a robust set of services that will encourage faculty to become supporters and advocates for the repository.

The University of Nebraska–Lincoln Story or Repository as Publisher

The University of Nebraska–Lincoln Libraries were very fortunate that when the English faculty were looking for a software package to publish an online jour-
nal, they came to the library for advice. The Libraries introduced them to BePress software. In that presentation, the libraries also saw how the software could be an effective option for creating an institutional repository. The libraries and the campus did not have the computer programming staff available to work with an open-source program, so a commercial package that cost less per year than a programmer was a cost effective decision. In addition, the universities’ dissertations which had been deposited with ProQuest/UMI since the early 1950s, could be batch loaded into the repository, giving the libraries a core set of materials to start a repository. While the English department chose not to create a journal at that time, the libraries did decide that contracting for this software would permit the libraries to develop an institutional repository, contract for the digitization of all dissertations from 1902 to the present, and perhaps support other journal efforts on campus. The next fortunate event for the libraries is that the libraries were able to hire the former director of the University of Nebraska Press as the Coordinator for Scholarly Communication. While this position was envisioned as continuing the various programs the libraries provided to the campus on scholarly communication issues, with the hiring of a skilled publisher into the position, the libraries changed direction and turned over the development of an institutional repository to the coordinator.

Why does having a publisher as the manager of a repository make a difference? First, we tried the same things everyone else tried to encourage faculty to participate. The coordinator held demonstrations, went to department meetings, and talked about how faculty research could be disseminated more widely than seen in traditional journals. These activities resulted in a little interest but did not capture the imagination of the faculty. Realizing that self-archiving was unlikely to be successful; the coordinator developed a system for mediating deposit through the libraries (Royster, 2009). Faculty were asked to send the libraries their vitae. The coordinator then reviewed the list of publications, identified those that could be posted from the publisher’s site, identified those that would need to be scanned and then OCR’d, and those that could not be deposited. Using work study students to do the scanning and uploading, the coordinator was able to begin to build a set of faculty articles. The unit did copyediting and typesetting when needed to make sure the full-text of the article met quality standards. Although the copyediting work can be time consuming, it is less problematic than trying to sort out well-intentioned but not well done faculty submissions. Further, only about 10% of the deposits fit this category, but is it an important part of the services. Metadata is created in the uploading process and is verified for accuracy. Abstracts are included or, if needed, are created for each work so that search engines will be more likely to discover the faculty member’s work.

Another strategy for encouraging faculty participation was to find a faculty written article that could be easily deposited and then contact the author for
permission. Once a faculty member had one article in the repository and began to receive monthly statistics on downloads and use, the faculty member was more inspired to ask the Libraries to deposit the rest of his or her materials. Emeriti faculty were another group that proved to be interested in the repository. After a presentation to the emeriti faculty organization, the coordinator was contacted by many of the attendees to learn how to have their work deposited in the repository. Perhaps because these faculty are thinking more about the legacy they will leave, emeriti faculty became one of the earlier groups to join the system. A wonderful outcome of this process is that some faculty discovered that their work was being used by a new group of researchers and they found themselves being consulted on some of their original research. For some faculty, the repository created a second career for them and brought them new recognition in their own fields.

A successful marketing strategy developed by the coordinator and based on his knowledge of authors and author behavior was to appeal to author self-interest. Instead of describing the importance of open access efforts, or of contributing to changes in the world of scholarly communication, he marketed the repository as a way for faculty get concrete evidence—that is, statistics—on use of their research. He showed faculty that citations to their work would tend to increase as their work became more visible through Google and Google Scholar. He also created a great marketing slogan: Participation is a two step process. Step one: send us your vita. Step two: there is no step two. This simple statement resonated with faculty as they were not being asked to learn new skills or learn new technology. All they had to do was send an e-mail. The process worked.

The monthly download statistics were also crucial to convincing faculty of the worth of the repository. Faculty participating in the repository began to tell their colleagues about the statistics they were getting. Departments began to talk about who had the most articles in the repository. Faculty began to compete with each other for most downloads. Faculty sold the repository to each other. By creating a “buzz” around the publishing work, the coordinator was able to change the viewpoint from why participate to how to participate (see Figure 1).

As a publisher, the coordinator also saw opportunities others might have missed. Early in the process of creating the repository, the coordinator met with faculty in the parasitology lab and noted a manuscript lying on the shelves. Upon inquiry, he learned that the manuscript was a dictionary of invertebrate zoology that the faculty had complied but had been unable to find a commercial publisher to accept the work. The coordinator offered to publish the reference work in the repository. The faculty mentioned they had considered an online publication but had been unable to determine how best to complete the work. The Institutional Repository manager took on the project and took 2 weeks to create an online version of the Dictionary of Invertebrate Zoology. This reference work, translated to a Portable Document Format (PDF) document from a Micro-
soft Word document, quickly became one of the most downloaded items in the repository even though it was a very large file. About 6 weeks after publication, recognizing the importance of this work, the coordinator prepared a better version using Adobe InDesign and PDF-ed with Adobe Acrobat to reduce the file size to 3.4 mbytes. Later, the libraries added a print on demand version for those who wanted a print copy. As a publisher, the coordinator saw the potential for this reference work and had the skills to turn a 15 year project into a very successful published work (see Figure 2).

A second original monograph that came to the attention of the coordinator was that of a book on the Hopi Indians that covered a series of papers presented at a conference in the 1980s. The authors had been unable to find a publisher as the book included over 70 colored prints that would make the book very expensive to produce in print. The coordinator was able to take the manuscript and turn it into an e-book, with a print on demand option. By publishing an e-book that had a limited commercial market but an important scholarly market, the coordinator was able to show faculty yet another purpose for the institutional repository. Another set of e-books that the coordinator was able to publish were those from an emeriti faculty member who studies the birds of Nebraska. The author was making his field guides available by printing them himself. The coordinator was able to take the manuscripts and create quality publications. Based on

**Figure 1.** Usage of University of Nebraska–Lincoln’s (UNL) Digital Commons, July 2005–November 2010.
the success of this first set of works, the author brought additional works, which do not have a commercial market but are important scholarship, to the repository. The author noted that he is more productive now and is again inspired to continue his research now that he knows it will be published and made available to researchers throughout the world. By taking on these different projects, the Libraries built support among faculty for the idea of the repository.

As faculty began to see results from the repository, they started to think of other ways to use the system. The Architecture College needed a way to showcase the final projects from their students. The projects naturally included images as well as text. They could include a variety of multimedia. The repository was a great solution for archiving the students’ work, and making the projects available
through the Internet so students could reference them as they applied for jobs. The College of Agriculture started an open access journal on undergraduate research using the repository software. The extension unit used the repository as a way to maintain an historical file of their publications. Instead of limiting the types of publications that can be deposited, or trying to train faculty to be self-publishers as well as doing self-deposit, the UNL Libraries have looked broadly at how to best publish, and disseminate the intellectual output of the faculty and the intellectual heritage of the university.

**Future for Institutional Repositories**

At times, as libraries struggle to determine how to develop and obtain content for an institutional repository, it appears that institutional repositories are a solution in search of a problem. Surveys of faculty, such as that done at Oklahoma University, find that faculty do not necessarily see advantages to participating in an institutional repository (Brown & Abbas, 2010). Research on faculty interest in self-archiving articles shows that faculty have limited interest in self-archiving systems. Even faculty who participate in subject based repositories do not necessarily participate in their own institution’s institutional repository. Efforts to persuade faculty to voluntarily deposit their articles on work funded by the National Institutes of Health were not very successful. Only when the government mandated deposit of articles and noted that failure to comply would jeopardize future funding did faculty and institutions become more concerned about how to meet this mandate. Current evidence then would make one ask why should institutions reallocate resources to create and maintain an institutional repository.

Recently, Mike Furlough, from Pennsylvania State University, wrote about repositories from a user perspective (Furlough, 2009). He notes that repositories are services that can help an institution manage content and preserve the intellectual work of the campus. Repositories are the means by which an institution can distribute information in an open access mode that makes that research accessible throughout the world. The repository then is a tool to manage the image of the university by showcasing the research work of the faculty and students. For libraries, repositories were seen as a way to combat increasing serial prices and collect and manage the work of their own constituencies. The reality though is that unless these is a strong emphasis on outreach and showing impact of faculty work it becomes more difficult to justify the staffing and costs associated with creating an institutional repository. For some universities then it may be more fruitful to work with other institutions or disciplinary repositories to manage selected works from the campus, rather than start their own repository.

For those institutions where a digital repository program fits well with the overall strategy for collection development, stewardship, and access to research
for the university’s constituencies, managing an institutional repository becomes an important strategy. It is, however, a strategy that needs to be integrated into the overall work of the library and the campus. For libraries that offer services to publish open access journals, or manage born digital materials, or are looking at how to provide data curation services, an overall plan that integrates these various functions into a coherent strategy is crucial. If the repository and other publishing efforts function as silos, separate from each other and from the other activities in the library, they will not be as successful. Repositories though, that are seen as part of the collection process and are part of the system for managing digital content, become a part of the core services of the library. For example, with consistent standards for metadata creation and ensuring the accuracy of metadata records it is possible to harvest repository data into a library catalog using the new discovery tools. Integrating the repository records into the catalog enables users to identify not only the purchased books and journals available in the collection but also the faculty and student authored research. The repository then becomes a part of the overall libraries collection of books, journals, images, digital files, and research articles (Park & Tosaka, 2010).

For repositories to survive and become more than just holding places for local, institutional research, systems need to be developed that will help link or network individual repositories. Exchanging information between institutional, subject, and funder repositories can lead to a systematic view of an integrated network of research (Darby, Jones, Gilbert, & Lambert, 2008). The changes needed in approach, standards, and workflows to make repositories successful will likely be evolutionary rather than revolutionary as repositories reach critical mass and institutions determine how best to integrate a repository into their services.

Notes

References


