

Institutional Repositories, Open Source Options, and Libraries

Megan Amaral

Introduction

Digital repositories are arriving at the forefront of software development for organizing collections of digital content. Libraries, museums, schools, open education organizations, universities and other research institutions have begun discovering the benefits of establishing and maintaining digital repositories. The rise of digital repositories is significantly increasing in large part due to the fact that several digital repository software products are available as open source software (OSS). Several headlining OSS communities have formed to create successful digital repository software that is available for free. The directory of open access digital repositories put together by OpenDOAR provides insight not only to the growing numbers of digital repositories, but also illustrates that the most widely used open source products are DSpace and Eprints (OpenDOAR, 2008). The Linux-distributed product Fedora is also a popular choice. DSpace, which was developed by MIT and Hewlett-Packard and is now managed by the DSpace Community, claims to have the most digital repository communities with over 250 institutions as of October 2008 (DSpace, 2008).

There are also, of course, proprietary companies creating digital repository software products. As opposed to the open source products, the implementation of these products comes with customer service, professional project managers, and suites of pre-designed customizations to choose from. Bepress (Berkeley Electronic Press) developed the repository platform Digital Commons, which is the most widely-used proprietary digital repository software product (Bepress, 2008; OpenDOAR 2008).

Universities and research institutions are establishing institutional digital repositories to collect, showcase, and disseminate their associated work. Choosing between OSS and proprietary software for these repositories rests on several considerations that each institution will have a unique perspective on. Libraries have an opportunity to remain visible in this migration to digital communication by being involved in the establishment and maintenance of institutional repositories and their contents.

Uses of Digital Repositories

Digital repositories can manifest themselves in several forms. Namely, organizations are constructing digital repositories for the purpose of creating digital archives, digital libraries, and institutional repositories. These types of repositories are similar in nature, each serving to preserve a collection of digital content that is of local interest to the community that has built it, yet nuances between them mark their individual characteristics. Digital archives are collections of digitized (or digitally-born) historical documents, while digital libraries are libraries comprised of materials that are in digital formats. The focus of this paper is on institutional repositories.

Institutional repositories are categorized as a digital repository sponsored by and containing the digital output of a specific institution, such as a research organization or a college or university. Governmental organizations and various research institutions, such as the American Museum of Natural History, have begun establishing institutional

repositories alongside many higher education institutions. Institutional repositories are more than digital archives in that they are perpetually being added to and the items are subject to revisions, making them a sort of living archive of collected documents.

Each institutional repository generally serves a particular community. Usually, the community will be a college or university, or a specialized research institution. In the institutional repositories of large universities, it is not unusual to see departmental communities within the main repository, but the repository as a whole is serving the community of the university.

Universities are generally driven to establish an institutional repository by a need and desire to collect and make accessible the research and digital output of the university. Digital output can include published papers, pre-publication papers, conference proceedings, educational resources, raw data, audio files, video files, and image files. Institutional repositories are designed to make this content accessible to members of the university community, and - if the university supports open access initiatives - to the world. What makes institutional repositories an ideal place to collect and store these things is that, like any form of digital repository, they are designed to make the content durable against the tests of time and changing technology. Submitted items are functionally preserved so that they will remain accessible regardless of how technology evolves in the future.

Another benefit of institutional repositories is that they serve as a showcase of the institution's work and thus increases the institution's marketing potential. In a 2002 SPARC (Scholarly Publishing and Academic Resource Coalition) paper, Raym Crow states that an "institutional repository concentrates the intellectual product created by a university's researchers, making it easier to demonstrate its scientific, social and financial value" (Crow, 2002). It becomes apparent, then, that an institution would benefit in several ways by creating an institutional repository.

Items that have been added to an institutional repository are not only captured and stored, but they are also cataloged and become searchable to both the institution and to the larger world communities - depending on the user restrictions an institution may choose to establish. Federated searching of multiple institutional repositories is an emerging development being powered by metadata standard. The metadata assigned to items within institutional repositories is imperative to each item's findability, yet each institution has the freedom to decide how their metadata is produced and what standards it will meet.

For example, the Open Archives Initiative (OAI) has developed the OAI-PMH (OAI Protocol for Metadata Harvesting), which is a repository metadata standard. The standard is based on Dublin Core, and most open source digital repository software such as DSpace, Fedora, and Eprints all support the OAI-PMH. The standard enables the metadata of digital repositories to be harvested by search engines such as Google and OAIster (<http://www.oaister.org/>).

OSS Digital Repositories

Having the option and ability to customize software is fundamental to the OSS movement. This aspect of OSS is particularly important to institutional repository software because each repository is specific to the organizational needs and goals of a specific institution. Lynch (2003) rightly states that, "institutional repositories will

succeed precisely because they are responsive to the needs of campus communities, and advance the interests of campus communities and of scholarship broadly." As these needs change, it is highly beneficial that the repository can change quickly and easily as well. This is something that is not guaranteed with proprietary software for reasons associated with inflexible code and licenses.

Due to the customizable nature of OSS, local policies can be easily implemented to serve the special needs of a given institution. A repository can be customized with unique themes and user interfaces, and varying levels of access can be granted on the departmental and institutional levels. In some cases, a necessary customization that is possible with OSS might also include the need to integrate a repository that already exists within the institution. These are all very appealing qualities to institutions interested in establishing a repository.

Many would argue that the primary benefit of choosing an open source digital repository product over a proprietary product is that of cost. The OSS product is free, and thus provides a resource that creates the potential for institutions without the means of paying for this service to do it themselves. Some digital repositories, like DSpace, will be useable "out of the box," or simply in the form in which they are provided, without needing any additional customization. This would make for a digital repository resource without calling for money or great amounts of technical knowledge. But most institutions would undoubtedly prefer something a bit more customized, especially if there are local needs that must be addressed.

Customization of open source software beyond the "out of the box" model requires time and maintenance from people who have the skills to create the desired modifications. This will require some commitment of the institution's resources, such as staff time. It is also important to remember that just because the software is free, there will still be hardware and storage space requirements to be considered and purchased. Thus, the "free" cost of OSS can only go so far before there are other costs to consider. Fox (2006) recognizes the cost and investment of staff time that stems from the decision of choosing OSS over proprietary software. Yet he argues that, with open source repository software, "the long term investment is your staff, [so] why not also gain the flexibility of being able to modify software to meet your institutional needs?"

Concerns Regarding the Use of Open Source Repository Software

There are a few reasonable concerns regarding the choice of a digital repository built with OSS over one built with proprietary software. One major concern is that choosing an open source digital repository product is choosing to be self-reliant, as there is no direct customer service available for users of open source options, as there would be when proprietary software is used. This self-reliance also exemplifies the need for staff members that are familiar with programming to be available to work on the repository - particularly in the design and implementation stages. University information technology service departments are usually unable to take on such a project as their resources are generally taken up by other campus needs. Some universities may solve this problem by hiring a full time staff member just for these initial phases of the project due to the large dedication of time that is necessary.

Yet it is also possible to find support solutions through alternative routes. The open source movement in general has shown a strong tendency to create communities of

users and developers who communicate using wikis, listserves, and through the site where the software is acquired. Ideas, problems, and solutions are actively exchanged, and they are usually archived somewhere for searching later. Additionally, commercial customer service companies such as LibLime can also provide support for open source products such as staff training and software maintenance.

A different approach to reduce the risks associated with self-reliance in using open source repository software has been to develop a digital repository consortium among several institutions. The Texas Digital Library is a successful example of a digital repository consortium comprised of 16 universities that was built using DSpace. Australia's ARROW Project is a consortium of Australian institutions that used Fedora as a starting point to create a joint repository. ARROW also chose an innovative option that incorporates open source with proprietary products by partnering with the developer VTLS in order to provide user support and ongoing development on top of the Fedora software (Groenewegen & Treloar, 2006).

Other, more logistical concerns regarding the choice of open source repository software over proprietary options include the responsibilities of creating a back-up system, and the work associated with installing, storing, and maintaining any hardware that would be needed to run the repository.

It must be noted that in some cases, OSS may not be the most viable option for repository software. Some universities have published their experiences in implementing institutional repositories, and it is suggested by this collection of literature that sometimes the choice of a proprietary solution may make more sense than OSS. This tends to depend on the size and economic situation of the university, while also considering the demands of staff time and the technical and programming skills of current staff members (Kelly, 2007; Sutherland & Hopkins, 2006). The University of Wollongong, Australia, and Florida State University are two examples where proprietary software was chosen over open source options. These choices were both the results of a desire to establish the repository quickly and due to the costs of the "substantial local IT support" that would have been necessary to implement an open source repository software option (Organ & Mandl, 2007).

Establishing Institutional Repositories

What is particularly useful to institutions interested in establishing a repository is that there is an increasing amount of available literature about how to do so, and many institutions are publishing recounts of their individual processes of creating digital repositories. Of particular use for getting started with institutional repositories are the publications produced by SPARC, and Susan Gibbon's write-up in July 2004's *Library Technology Reports* published by the ALA.

What nearly all of the literature suggests is that developing an institutional repository takes a major investment of time and energy. Aside from selecting software and implementing it, it is necessary to have the support of the institution as a whole in order for the project to be a success. In a report published by the Association of Research Libraries, Lynch (2003) states that "an effective institutional repository...represents a collaboration among librarians, information technologists, archives and records managers, faculty, and university administrators and policy makers." This massive collaboration is necessary because there needs to be coordination among the institutional

departments, and the repository will need to be effectively marketed to members of the university community in order to acquire content.

The foundation for acquiring content in an institutional repository lies in building trust with the faculty and other potential contributors. Two main reasons support this: first, if contributors are going to spend their time participating and uploading content, they will want to be sure that their efforts are worth the investment. Secondly, they may come to rely on the repository as the place to store their work, and they cannot lose their work due to project abandonment (Lynch, 2003). In order to be convinced that the institutional repository is trustworthy, a university-wide marketing campaign endorsed by head administrators will illustrate the seriousness of the project.

It is also important to recognize the link between institutional repositories and the open access movement and how this influences potential repository contributors. Efforts are being made to encourage researchers to publish their papers in institutional repositories either as a supplement to, or instead of, traditional scholarly communication methods. It is well known that traditional publication sometimes produces prohibitive costs for journal subscriptions, but increased readership is inevitable if articles are made freely available by a open access initiatives and institutional repositories. This increased readership should be marketed as a benefit to potential contributors due to personal publicity interests as well as altruistic scholarship. Indeed, a recent survey of institutional repository contributors found that motivation to contribute was driven by such altruism to make research freely accessible (Kim, 2007).

Yet in order for institutional repositories to be legitimized as a suitable publication medium, certain guarantees to the repository contributors must be assured (Lynch, 2003). Institutional repositories must be effective and permanent mediums to disseminate information in order to succeed. And these guarantees are made by the preservability features that are inherent to digital repositories, as well as by the wider readership that open access repositories are capable of producing. It is up to the hosting institution to maintain the repositories integrity beyond these points.

Role of Libraries in Establishing Institutional Repositories

It seems logical that libraries would be strong advocates for establishing an institutional repository. The library's intrinsic duty to an institution of collecting and preserving information should apply as much to the digital output of an institution as it might to the paper collections within the physical library walls. John (2005) states that by establishing an institutional repository, "the library is responding to a variety of concerns: long term access, open access, and improved re-use of intellectual property." These are activities that already correlate with the mission and goals of libraries.

Additionally, librarians would make excellent managers of institutional repositories by nature. The organizational architecture of the repository, the creation and/or management of metadata, and the understanding and communicating of various licensing policies coincide with what is already traditional library work. Gibbons (2004) points out that the core features of institutional repositories – material collection, preservation, distribution, and metadata application – are tasks that only librarians can claim expertise in across the board.

Institutional digital repository management will also provide libraries with an opportunity to remain at the forefront of their institution's scholarly communication

ventures. "Libraries taking part in the process will undergo a metamorphosis: from paper-based thinking to the digital paradigm, from importers of global knowledge to exporters of local knowledge, from suppliers of visible collection to invisible partners in academic processes" (Waaijers, 2005). Institutional repository management would place libraries in a highly visible role that serves to facilitate the central hub of the institution's scholarly communication. It is therefore important that libraries take advantage of this opportunity while institutional repositories are increasing in popularity and thus secure their place in this new way of collecting and organizing an institution's digital output.

When undertaking the responsibility of organizing the establishment of institutional repositories, libraries have the opportunity to investigate OSS repository options when considering repository options. Fox (2006) encourages library participation in the general open source movement based on the motivation to "be innovative, solve unique problems, add to a growing body of knowledge, and use resources to their best advantage." If the conditions exist to allow for taking on the potential challenges of selecting open source repository software, some OSS proponents would argue that it should be done. Libraries and the open movement mesh together logically. Providing access to resources and information at little to no charge to users is historically an essential role that libraries have played within their communities. Libraries also "have a natural synergy with the open source movement," based on an extensive history of managing various kinds of licensing issues (Krishnamurthy, 2007).

Conclusion

Institutional repositories provide organizations with an opportunity to create a central location that collects and preserves their digital output. The opportunity to share and distribute this output is hugely significant, and would serve to benefit the repository's contributing authors and the institution itself. If part of the purpose of an institutional repository is to make its contents freely available, then the global community also has something to be gained. When an institution collects and shares its output, the members of the institution benefit while also making the world academically richer by allowing scholarly communication to flow more freely.

The digital repository software products that are available as open source are proving to be effective in fulfilling these informational harvesting functions. Libraries are ideal candidates for initiating and managing an institution's migration to a digital repository system, and thus fulfill their duty to the institution as stewards of local information.

References

- Bepress. (2008). Showcase your institution's research and scholarship on the Web with Digital Commons. Retrieved October 25, 2008, from <http://www.bepress.com/ir/>.
- Crow, R. (2002). The case for institutional repositories: A SPARC position paper. Retrieved November 4, 2008, from <http://www.arl.org/sc/models/repositories/index.shtml>.
- DSpace. (2008). Largest community of users and developers worldwide. Retrieved October 20, 2008, from http://www.dspace.org/index.php?option=com_content&task=view&id=305&Itemid=142
- Fox, R. (2006). The digital library in the bazaar. *OCLC Systems and Services*, 22(2), 100-106. Retrieved November 4, 2008, from Research Library database.
- Gibbons, S. (2004). Benefits of an institutional repository. *Library Technology Reports*, 40(4). Retrieved October 15, 2008, from Academic Search Complete database.
- Groenewegen, D, and Andrew, T. (2008). The ARROW project: A consortial institutional repository solution, combining open source and proprietary software. *OCLC Systems and Services*, 24(1), 30-39. Retrieved November 4, 2008, from Research Library database.
- John, N. (2005). Digital repositories: Not quite at your fingertips. *Libri*, 55(4), 169-235. Retrieved October 30, 2008, from <http://www.librijournal.org/2005-4toc.html>.
- Kelly, J.C. (2007). Creating an institutional repository at a challenged institution. *OCLC Systems and Services*, 23(2). Retrieved October 12, 2008, from Emerald database. doi: 10.1108/10650750710748441
- Kim, J. (2007). Motivating and impeding factors affecting faculty contribution to institutional repositories. *Journal of Digital Information*, 8(2). Retrieved November 28, 2008, from <http://journals.tdl.org/jodi/article/view/193/177>
- Krishnamurthy, M. (2008). Open access, open source, and digital libraries: A current trend in university libraries around the world. *Program: electronic library and information services*, 42(1), 48-55. Retrieved October 12, 2008, from Emerald database. doi: 10.1108/00330330810851582
- Lynch, C.A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *ARL: A Bimonthly Report*, no.226. Retrieved November 4, 2008, from <http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>

- Open Archives Forum. (2008). OAI for beginners - the Open Archives Forum tutorial. Retrieved October 8, 2008, from <http://www.oaforum.org/tutorial/english/page1.htm#section2>.
- OpenDOAR. (2008). Usage of open access repository software-worldwide. Retrieved October 30, 2008, from <http://www.opendoar.org/onechart.php?cID=&ctID=&rtID=&clID=&lID=&potID=&rSoftWareName=&search=&groupby=r.rSoftWareName&orderby=Tally%20DESC&charttype=pie&width=600&height=300&caption=Usage%20of%20Open%20Access%20Repository%20Software%20-%20Worldwide>
- Organ, M., and Mandl, H. (2007). Outsourcing open access. *OCLC Systems & Services*, 23(4). Retrieved November 28, 2008 from Emerald database. doi: 10.1108/10650750710831493
- SPARC. (2008). Repository Resources. Retrieved November 4, 2008, from <http://www.arl.org/sparc/repositories/>.
- Sutherland, M. and Hopkins, P. (2006). Open source or off-the-shelf?: Establishing and institutional repository for a small institution. Retrieved November 4, 2008, from http://epublications.bond.edu.au/library_pubs/11/.
- Waijers, L. (2005). From libraries to 'libratories.' *First Monday*, 10(5). Retrieved November 4, 2008, from <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1303/1223>