C. Sean Burns, Amy Lana, and John M. Budd
University of Missouri, Columbia, MO

Institutional Repositories: Costs and Benefits

Abstract: There have been some exploratory studies of institutional repositories among academic libraries, but few have focused on costs. This research surveys U.S. academic libraries hosting IRs and asks what are the costs associated with implementing and maintaining them. The findings will benefit librarians and administrators with their IR related decision-making.

Résumé:
À la suite du projet initial DSpace du MIT, de nombreuses bibliothèques universitaires ont créé des dépôts numériques institutionnels. Ces dépôts jouent plusieurs rôles, y compris donner accès à de l'information et à des données qui seraient autrement inaccessibles à la communauté des chercheurs, du matériel de cours, des volumineux ensembles de données, des analyses de données détaillées, et plus encore. Avec une flexibilité accrue en matière de droit d'auteur, il est maintenant possible d'y verser les éditions pré-originales ou acceptées, augmentant ainsi leur utilité potentielle. Ce qui demeure largement inconnu sont les coûts associés à l'établissement et à la maintenance d'un tel dépôt, qui gère les opérations et dans quelle mesure l'information est-elle utilisée. Ce projet vise à répondre à ces questions.

Introduction

As librarians develop institutional repository (IR) initiatives a potential exists for greater access to scholarly information in many forms. The open access element renders objects deposited by faculty and researchers more accessible to people everywhere. However, there are a host of planning issues involved with implementing and managing an IR. Will the repository use open source or proprietary software, accept items from one campus or from a consortium of institutions, or offer services to facilitate deposits? We would also like a better understanding of staffing needs and start-up and ongoing costs. These questions must be asked to ensure the project's success. Unfortunately, little is known about these matters and librarians may have to discover answers as they initiate IR projects.

To address these uncertainties, we offer a preliminary summary of our survey of academic libraries with IRs. The 29 question survey addresses a number of factors that may influence total cost. These factors include personnel, software, and hardware as well as content, acquisitions, mandates, additional services, administrative responsibility, and usage.

Background

In 2002, the Massachusetts Institute of Technology (MIT) launched its DSpace IR, beginning a new storage and retrieval era for academic libraries (and some nonacademic partners). Since MIT was a groundbreaking creator of IRs it had something of a luxury in determining just what shape a repository would take. Among the first decisions made were to make DSpace an open access resource that would extend beyond the MIT campus and to use Dublin-Core to allow information harvesting. Another aspect of MIT's pioneering effort was its ability to create innovative services which could be marketed to other libraries. Without going into detail here, MIT developed a set of Core Services (offered to libraries at no charge) and Premium Services (offered for a fee) (see "MIT's DSpace Experience," n.d.).

Initially MIT's DSpace collected materials that could readily be made available via open access. The repository included a great variety of materials, including theses and dissertations, courseware, data sets, drafts of papers, etc. In recent years publishers have become more willing to allow authors to post works to repositories or to faculty members' own Web sites. MIT reports that the DSpace Open Access Articles collection contains approximately 3,000 scholarly works that the campus's faculty has made available. It also states that articles have been viewed more than 100,000 times since the collection was launched in October, 2009.

Literature Review

Little is known about the costs associated with implementing and managing IRs, but a number of studies have addressed associated issues. Lynch and Lippincott (2005) conducted a survey which focused on content type and the software used for IRs. Rieh, et al. (2007) gathered various descriptive data about IR development among academic libraries, which included types of funding sources. They found that the primary source of initial funding is a library's special initiative. This is followed by costs absorbed by the library's operating budget and then costs entered as line item in the budget.

Rieh, et al. (2007) asked whether academic librarians mediate submissions or allow self-archiving by content creators (see also Bevan, 2007; Giesecke, 2011). This question underlies different IR models, which may result in different operational costs. Such IR models may be complicated by the mandate status of an institution (Thomas and McDonald, 2007), whether additional services, such as digitization, are provided (Piorun and Palmer, 2008), and the use and acceptance of the IR (Davis and Connolly, 2007).

Method

This study summarizes preliminary results from a survey of academic libraries that have implemented IRs. The survey questions were framed to elicit responses of sufficient specificity so that conclusions could be drawn about the operation of IRs in general. Where pertinent, we asked respondents to provide data for the 2011 calendar year.

We identified 160 academic libraries with IRs through OpenDOAR. On February 7, 2012 we sent each of them an email requesting participation in a Qualtrics-hosted survey. We sent a reminder email two weeks later. This resulted in twenty-three finished surveys. On March 5, 2012, we sent a request for participation to the members of the SPARC-IR and the SPARC-SR mailing lists. As of March 23, 2012 we have thirty-eight total eligible responses. The survey will remain open until April 6, 2012. All findings are tentative and subject to revision as the responses are collected and thoroughly examined.

Findings

We asked respondents to tell us their Carnegie Classification. Of thirty-eight eligible responses, most can be classified as Research University with very high research activity (n = 17) and with high research activity (n = 7). This is followed by the Doctoral / Research University classification (n = 5) and the Master's Colleges and Universities levels: larger (n = 2), medium (n = 2), and smaller (n = 2) programs, and Other (n = 3). Most respondents report that there is no mandate at their institutions that require faculty to submit content to their IRs (n = 33).

Thirty-eight respondents answered our question about the size of their IR. Twenty-two report more than 5,000 items and sixteen participants have 4,000 or less. Librarians associated with the repository are primarily responsible for handling submissions (n = 33) rather than allow content creators to self-archive (n = 5). Seventeen offer additional services, such as digitization, copy editing, project and data management, copyright clearance, bulk import, and metadata creation.

The IRs host a variety of content and format types. The majority of this content includes journal articles, conference articles, and presentations but the IRs also include book chapters, books, theses, dissertations, data sets, tech reports, working papers, courseware, newsletters, grant proposals, plays, poetry, video, audio (including podcasts), photographs, maps, posters.

To help measure usage, fifteen respondents provided traffic data. The results show a wide range of visits (n = 15; *Min* = 1,000; *Max* = 5,500,000, *Mdn* = 65,930). We asked about in-site searches, which might be a more robust indicator of usage. These results are also wide ranging (n = 9; *Min* = 55; *Max* = 57,280; *Mdn* = 11,370). Another indicator of usage is the number of items retrieved or downloaded (n = 15; *Min* = 5,832; *Max* = 9,821,000; *Mdn* = 432,300). Since the

retrieved or downloaded indicator is much higher than the visit or in-site indicators, users may be retrieving content through third party discovery services without visiting the IR directly.

Respondents provided data on the number of full time equivalent (FTE) staff that work on the IR. Professional staff hold the most responsibility (*n* = 35; *Min* = 0; *Max* = 12; *Mdn* = 1.75) when compared to non- or para-professional or clerical staff (*n* = 29; *Min* = 0; *Max* = 9; *Mdn* = 0.5) and student workers (*n* = 34; *Min* = 0; *Max* = 7; *Mdn* = 0.5).

The cost (in U.S. dollars) to implement and operate IRs varies. The median reported cost to implement an IR (*n* = 21; *Min* = $0; *Max* = $300,000; *Mdn* = $20,000) is much lower than the median annual cost to operate an IR (*n* = 21; *Min* = $0; *Max* = $275,000; *Mdn* = $31,000). The primary expense is due to personnel (*n* = 17; *Min* = $100; *Max* = $235,200; *Mdn* = $70,000), followed by software (*n* = 10; *Min* = $2,500; *Max* = $40,000; *Mdn* = $23,000), and hardware (*n* = 6*; Min* = $500*; Max* = $50,000*; Mdn* = $5,500).

Perhaps the most striking cost differences are among the personnel, software, and hardware categories between open source and proprietary software IR solutions. We find that proprietary solutions appear to cost much more than open source solutions in all three categories. Table 1 reports the median value for these costs.

Table 1
Median Personnel, Software, and Hardware Costs by Software / Service Type

	Software / Service Type			
	Open Source		Proprietary	
	n	Costs	*n*	Costs
Personnel	9	$50,000	8	$98,750
Software	2	$17,000	8	$24,000
Hardware	5	$1,000	1	$17,000

Given that Rieh, et al. (2007) reported that funding for IRs primarily comes from special initiatives supported by the library, our data on funding sources seems to indicate that IRs are becoming better integrated into budgets. We find that sources of funding primarily come from costs absorbed in routine library operating costs (n = 25), followed by regular budget line item for the institution's library (n = 11), special initiative supported by the library (n = 4), special initiative supported by the institution's central administration (n = 4), grant awarded by an external source (n = 3), special initiative supported by the institution's archives (n = 1) and other (n = 1).

Conclusion

This preliminary summary of our survey highlights a wide range of answers to fairly standard questions. To reduce uncertainty and provide results of costs that can assist decision-making about IRs for different types of institutions, our next step will involve a more thorough examination in order to provide a clear description of the collected data.

References

Bevan, Simon J. (2007). Developing an institutional respository: Cranfield QUEprints -- a case study. *OCLC Systems & Services, 23*(2), 170-182. doi:10.1108/10650750710748478

Davis, P. M., & Connolly, M. J. L. (2007). Institutional repositories: Evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine 13*(3/4). Available at http://www.dlib.org/dlib/march07/davis/03davis.html.

Giesecke, Joan. (2011). Institutional repositories: Keys to success. J*ournal of Library Administration, 51*(5-6), 529-542. doi:10.1080/01930826.2011.589340.

Lynch, C. A., & Lippincott, J. K. (2005). Institutional repository deployment in the United States as of early 2005. *D-Lib Magazine 11*(9). Available at http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/september05/lynch/09lynch.html.

"MIT's DSpace Experience: A Case Study." *MIT Libraries*, n.d., http://libraries.mit.edu/dspacemit/info/case-study.pdf.

Piorun, M., & Palmer, L.A. (2008). Digitizating dissertations for an institutional repository: A process and cost analysis. *Journal of the Medical Library Association, 96*(3), 223-229. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2479051/

Rieh, S. Y., Markey, K., St. Jean, B., Yakel, E., & Kim, J. (2007). Census of institutional repositories in the U.S.: A comparison across institutions at different states of IR development. *D-Lib Magazine 13*(11/12). Available at http://www.dlib.org/dlib/november07/rieh/11rieh.html.

Thomas, C., & McDonald, R. H. (2007). Measuring and comparing participation patterns in digital repositories: Repositories by the numbers, part 1. *D-Lib Magazine 13*(9/10). Available at http://www.dlib.org/dlib/september07/mcdonald/09mcdonald.html.