

D-Lib Magazine

January/February 2013
Volume 19, Number 1/2

Institutional Repositories: Exploration of Costs and Value

C. Sean Burns

University of Missouri

csbc74@mail.missouri.edu or sean.csb@gmail.com

Amy Lana

University of Missouri

lanaa@missouri.edu

John M. Budd

University of Missouri

buddj@missouri.edu

doi:10.1045/january2013-burns

Abstract

Little is known about the costs academic libraries incur to implement and manage institutional repositories and the value these institutional repositories offer to their communities. To address this, the authors report the findings of their 29 question survey of academic libraries with institutional repositories. We received 49 usable responses. Thirty-four of these responses completed the entire survey. While meant to be exploratory, results are varied and depend on the context of the institution. This context includes, among other things, the size of the repositories and of the institutions, the software used to operate the repositories, such as open source or proprietary, and whether librarians mediate content archiving, or content producers directly deposit their own material. The highlights of our findings, based on median values, suggest that institutions that mediate submissions incur less expense than institutions that allow self-archiving, institutions that offer additional services incur greater annual operating costs than those who do not, and institutions that use open source applications have lower implementation costs but comparable annual operating costs with institutions that use proprietary solutions. Furthermore, changes in budgeting, from special initiative to absorption into the regular budget, suggest a trend in sustainable support for institutional repositories. Our findings are exploratory but suggest that future inquiry into costs and the value of institutional repositories should be directed at specific types of institutions, such as by Carnegie Classification category.

1 Introduction

Recognition of the open access (OA) movement varies by field, but scholars, both as readers and as contributors, are growing increasingly aware of open access publishing (OAP) as they encounter more and more OA and hybrid journals. Furthermore, interest among researchers in examining OA's impact on scholars has grown as awareness and adoption increase; for example, research exists that attempts to measure the adoption (Edgar & Willinsky, [2010](#)), the use (Norris, Oppenheim, & Rowland, [2008](#)), and the influence (Gargouri *et al.*, [2010](#)) of open access publishing.

In addition to open access publishing, the open access movement is also realized in open access archiving (OAA) (Chan, [2004](#)), usually expressed in the form of subject or institutional repositories. Some notable instances of OAA have received substantial attention, such as when Harvard University's Faculty of Arts and Sciences adopted a mandate to deposit their scholarship into their institutional repository (Albanese & Oder, [2008](#)), but scholars are generally less aware of the OAA front and adoption is at a much lower rate than it is for OAP (Kim, [2011](#)). Still, there is growing interest among librarians and scholars in examining OAA's impact, and we see this in studies of the use and influence of subject repositories, such as arXiv (e.g., Davis & Fromerth, [2007](#)), and various institutional repositories (e.g., Aguillo, Ortega, Fernández, & Utrilla, [2010](#)).

The serials crisis that began over thirty years ago (Greco, Wharton, Estelami, & Jones, [2006](#)), forcing librarians to confront dramatic increases in journal costs, has been a driving force for much of the open access movement. As serials costs have increased, and as library budgets have decreased or merely held steady, the impact on library collections has been drastic (Pascarelli, [1990](#)). It has been the hope that open access will help offset the financial burdens placed on academic libraries by eliminating some costs or by shifting some responsibility to journal publishers or to authors. Thus, although it is true the open access movement has strong ethical and civic components in seeking increased dissemination of scholarly and scientific knowledge, librarians also have had strong financial incentives to promote open access. However, it is not necessarily known whether open access models have helped to alleviate the financial burdens placed on academic library budgets (Waaaijers, Savenije, & Wesseling, [2008](#)).

A secondary solution to the crisis has been sought in the form of open access institutional repositories in which academic libraries house the digital intellectual output of a campus. Unfortunately, acceptance of this solution, as with OA in general, is a disruptive act in scientific and scholarly publishing and communication (see Guédon, [2008](#) for an excellent discussion on this). Although large publishing houses have recently begun to change their policies to allow submission of some articles to institutional repositories, usually after an embargo period (see Shotton, [2012](#) for a broad explanation of OA types), there continues to be a reluctance or lack of understanding on the part of many authors to submit content to a repository (Moore, [2011](#)). This combination of factors means that it is much too early to tell what kind of overall financial savings, if any, these entities have created. This should be a particular issue since building and managing an IR within an academic library will have a direct effect on the funds available to that library.

Despite a lack of evidence of positive financial impact that IRs may have on academic libraries, it seems that IRs should offer some kind value to the institution, the scholarly community, and the public. Ideally, this value should be framed by an IR's purpose and the commitment such purpose demands. Simply put, administrating an institutional repository requires a long-term commitment to safeguarding, preserving, and making accessible the intellectual content of one's institution. Without understanding the significance of this service, the value of such programs may be underestimated and, consequently, funds to ensure IR survival and growth may dwindle. This is further complicated by the fact that institutional repositories are not yet universally accepted by faculty as necessary or useful things. Encouraging buy-in requires a strong fiduciary relationship with the institution's community. Lynch ([2003](#)) argues that hosting an institutional repository is a "stewardship [that] is easy and inexpensive to claim; [but] it is expensive and difficult to honor, and perhaps it will prove to be all too easy to later abdicate. Institutions need to think seriously before launching institutional repository programs" (sect. 4, para. 12).

If academic librarians seek to build and support institutional repositories for the long-run, and if they wish to know what kind of value the repository adds to the academic library and to the library's community, then it is necessary to understand the costs and the value they offer. In other words, how expensive is it to implement and operate an institutional repository? How might we examine an institutional repository's value? To answer these questions, this paper reports on a survey of academic libraries with institutional repositories. The survey was designed to acquire understanding about costs, including personnel, hardware and software, and to frame this within the context of the institution and of the repository.

2 Literature Review

Institutional Repositories

This paper focuses on institutional repositories, but, for the sake of clarity, we distinguish among other types of repositories that exist. Darby, Jones, Gilbert, & Lambert (2008) describe three types: institutional, subject, and funder repositories. While there are end-purpose, system, and some software based similarities among these three types of repositories, each is fundamentally different in scope, sources of authorship, and funding. Since the purpose of this paper is to explore the financial considerations of institutional repositories for academic libraries, we limit our discussion primarily to institutional repositories.

An institutional repository is a type of digital library that "capture[s] the original research and other intellectual property generated by an institution's constituent population active in many fields" (Crow, p. 16, 2002). IRs in the United States are generally maintained by academic libraries at higher education institutions and they offer their relevant academic communities an additional venue to participate in open access (Chan, 2004) through a process that may be called open access archiving.

IRs as concepts are very much influenced by IRs as software. The two predominant and original IR software platforms are Eprints and DSpace. In 2000, Eprints was developed by Stephen Harnad and team at the University of Southampton (Tansley & Harnad, 2000) and in 2002, DSpace was launched by the Massachusetts Institute of Technology (MIT) with support from Hewlett Packard (Barton & Walker, 2003). Each application, both open source software, represented fundamentally different views about the purpose of IRs with respect to open access and scholarly publishing. Chan (2004) notes that Eprints was designed to host traditional forms of scholarly publishing, including journal and conference articles, book chapters, and so forth, while DSpace was intended to host a much greater variety of material, such as the more formal instances of scholarly communication as well as various types of grey literature. In this respect, Eprints was developed specifically to allow researchers to make their work open access (Harnad, n.d.) while DSpace was originally designed to address greater preservation of and access to the work of the intellectual output of an institution. (MIT Libraries, n.d.)

Value

An institutional repository extends an academic library's capability to participate in the scholarly communication system, and this capability should be considered a source of value, especially in a world that continues to migrate to the digital. The source of the IR value proposition stems from academic librarians' belief that IRs are an integral part of the future of research libraries and will help maintain librarians' roles as essential actors in the network of scholarly communication (Heath, 2009; Mower & Chaufy, 2009; Carpenter, Graybill, Offord, & Piorun, 2011). The argument made is that academic librarians add value to the scholarly communication process when they establish institutional repositories by, among other things, enhancing the discoverability of work through robust metadata and providing a permanent URI for that work.

The value of the institutional repository may also be derived by its acceptance and by its use. Such value is dual-sided in that we may examine it from the perspective of those who contribute content (authors or producers) and of those who may retrieve that content (readers or consumers). For example, Xia (2008) examines faculty willingness to contribute to IRs by comparing self-archiving practices between subject and institutional repositories. His purpose is to learn if the experience and enthusiasm among those who contribute to subject repositories (SRs) could be extended to participation with academic library institutional repositories. Unfortunately, he finds little evidence that scholars depositing in subject repositories are interested in depositing in institutional repositories. The implication is that, in some cases, SRs and IRs may be competing for intellectual content.

One purpose of institutional repositories is the dissemination of research output. Thus, it stands to reason one should be able to determine some of an IR's value, in terms of its ability to disseminate, by whether the content of the IR is discoverable. Since bibliographic databases such as Scopus and Web of Science primarily index formal, published literature, discovering IR content may often be left to academic library websites, online public access catalogs (OPACs), or to scholarly search engines such as Google Scholar.

Google Scholar is perceived to be a useful and usable service (Chen, [2010](#); Cothran, [2011](#)) for information seekers, and it purports to help the scholarly community discover content, including that found residing in institutional repositories. As early as 2010, work by open source platform developers in cooperation with Google engineers helped address a problem with records described using the Dublin Core metadata element set (DuraSpace, [n.d.](#)) and Google has also provided additional guidelines to help ensure a repository's content is indexed (Google, [n.d.](#)). Although these actions suggest that recent data gathering on IR discovery statistics is improving, Arlitsch & O'Brien ([2012](#)) note that institutional repository software describing content using Dublin Core can be a barrier for Google Scholar's ability to index that content. This is because Dublin Core's ability to generate specific bibliographic elements has been considered inadequate. In Arlitsch & O'Brien's case, they find that the use of Dublin Core leads to low index ratios for their institutional repository and that switching to a more discoverable metadata schema increases the ratio.

Since the content held in a specific institutional repository reflects that repository's purpose, the quality of that content will likely affect the value of an IR if we deem such things to affect the reputation of the larger institution (Wacha & Wisner, [2011](#)). As noted, the philosophical differences between Eprints and DSpace, at least in the early days of IRs at the beginning of the century, reflect two opposing viewpoints about their purposes. As Chan ([2004](#)) observes, Clifford Lynch has argued that an institutional repository should capture a broad range of content type, and this may include a variety of grey literature; on the other hand, Stephen Harnad has argued that an institutional repository, as a tool of the open access movement, is meant to liberate scholarly research publications from pay walls, usually as part of the Green OA self-archiving model (e.g., see Harnad *et al.*, [2008](#) on the Green and Gold OA issue). How such differences persist today should be taken into consideration when determining the value an IR offers to its institutional community.

Costs

Although librarians' initiatives to develop institutional repositories increases the potential for greater access to scholarly information and may offer other kinds of value, a host of planning issues involved with implementing and managing IRs need to be considered by library administrators and planners. A number of questions that could be asked in the early stages include some of the following. Will the repository use open source software or license a proprietary system, accept items from one campus or from a consortium of institutions, or offer services to facilitate deposits? Librarians may 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 the costs associated with implementing and managing IRs, but a number of studies have addressed related issues. Lynch & Lippincott ([2005](#)) conducted a survey which focused on content type and the software used for IRs, and since decisions about software choice will have a financial impact, such studies aid in understanding a potential IR's maintenance needs. Sources of funding will also influence what can be accomplished since how much funding is available and where the funding is sourced will be a factor of how much can be accomplished and under what time constraint. For example, Rieh, Markey, St. Jean, Yakel, & Kim ([2007](#)) gathered various descriptive data about IR development among academic libraries, which included types of funding sources. The authors found that the primary source of starter funds is a library's special initiative, followed by costs absorbed by the library's operating budget, and then costs entered as a line item in the budget. They refer to other considerations that affect expenses: fixed costs, economies of scale, and whether consortial and multi-institutional repositories are part of an initiative (see also Lynch & Lippincott, [2005](#)).

Labor or personnel issues are also a factor. 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 reflects different IR submission models, which may result in different operational costs, especially in terms of personnel. Additionally, IR costs may be complicated by the mandate status of an institution (Thomas & McDonald, [2007](#)), whether additional services are provided, such as digitization (Piorun & Palmer, [2008](#)) and copyright management (Li & Banach, [2011](#)), by the use and acceptance of the IR (Prosser, [2003](#); Davis & Connolly, [2007](#); Bevan, [2007](#); Cullen & Chawner, [2009](#)), software type chosen (Giesecke, [2011](#)), and sustainability in staff and funding (Li & Banach, [2011](#)). Long term, collective strategic plans may take into consideration the potential impact on traditional publishing models (Crow, [2002](#); Lynch, [2003](#)).

Few details are available, but research suggests costs of various services may be high. Piourun & Palmer (2008) report on the costs associated with digitizing dissertations for an IR and find that the cost to scan and digitize 320 documents totaled to \$23,562 in labor, which includes 906 labor hours for both temporary help and librarians, and \$0.27 per page. Bevan (2007) estimates that a mediated service, where librarians are responsible for depositing items in the IR, costs much more to operate than an institutional repository where content creators self-archive their work. Giesecke (2011) writes that "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" (section "Costs of Institutional Repositories", para. 1). Giesecke further writes that

Another set of costs for the institution to consider is the cost for self-archiving articles.... 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 (para. 4).

In an attempt to address some of these issues, the authors conducted a study comprised of a survey of libraries that have implemented IRs. The questions were framed to elicit responses of sufficient specificity that conclusions or leads can be drawn about the operation of repositories. For example, the survey instrument included questions about the size and nature of the institutions. A research university may well have needs and costs that are substantially different from those of a master's degree granting institution. It may be expected that, on a campus where the faculty publish sizable numbers of journal articles, the repository would be populated with many articles. Since this survey inquired into the kinds of items deposited in institutional repositories, there is the opportunity to examine both the potential for repositories and, to an extent, the realization of the potential.

3 Methods

We developed a 29 question survey guided by research from the literature review and framed to elicit responses about the costs, usage, and contexts of IRs at academic libraries. This meant that our survey addressed a number of factors that may influence total cost and help determine value. These factors include personnel, software, and hardware as well as content, acquisitions, mandates, additional services, administrative responsibility, and usage. Where pertinent, we asked respondents to provide data for the 2011 calendar year.

To recruit participants, we used [OpenDOAR](#), the Directory of Open Access Repositories, to identify potential candidates for our survey. Although OpenDOAR lists open access repositories of all types (subject, institutional) and is a world-wide directory, we initially sought candidates only from US higher education institutions with institutional repositories. This was a self-imposed restriction justified only for simplicity and measurement consistency in cost calculations. We identified 160 US academic libraries with IRs through the OpenDOAR registry. On February 7, 2012 we sent each of them an email requesting participation in a Qualtrics-hosted survey and sent a reminder email two weeks later. These solicitations resulted in 23 finished surveys for a 14.38% response rate.

Due to the low response rate, on March 5, 2012, we sought additional participation from members of the SPARC-IR and the SPARC-SR mailing lists. While we restricted responses to those from higher education institutions with institutional repositories, the invitations broadened our list of potential candidates to include those from international institutions. The mailing list invitations resulted in a total of 57 responses to our survey and a 72% completion rate ($n = 41$) of that total. Since the total potential pool of candidates is unknown, a final response rate is difficult to estimate. We filtered out respondents who categorized their repository as non-institutional (e.g., departmental, dissertation and thesis only, etc.), and this resulted in a

sample size of 49 with a 69% completion rate ($n = 34$) and a 31% partial completion rate ($n = 15$). The survey was closed on April 7, 2012.

Some of the data analysis was completed using the Qualtrics survey software, but most of the work was done by downloading a comma separated value file from the Qualtrics site and using The R Project for Statistical Computing (R Development Core Team, [2011](#)), which is free and open source statistical software.

4 Results

Overall, the survey resulted in a number of interesting and insightful responses; however, since many responses are incomplete, we should highlight the following two consequences. First, we sought answers to some specific statistics about budgets and uses of the IRs, and many of our respondents may not have had easy access to such information while some may have. As a result, there seem to be some guesses or estimations in the responses. Second, where the data is inconsistent, the findings suggest something about the state of institutional repository use among institutions of higher education in the US and a lack of consistent reporting methods for accounting and budget considerations.

We asked respondents to tell us the basic [Carnegie Classification](#) of their higher education institution. Of 49 eligible responses, most can be classified as RU/VH: Research Universities with very high research activity ($n = 23$) and RU/H: Research Universities with high research activity ($n = 9$). This is followed by DRU: Doctoral / Research Universities classification ($n = 6$) and the Master's Colleges and Universities levels: Master's/L with larger programs ($n = 2$), Master's/M with medium programs ($n = 2$), and Master's/S with smaller programs ($n = 3$), as well as Other ($n = 4$). Most respondents report that there is no mandate at their institutions to require faculty to submit content to their IRs ($n = 36$), but a handful respondents do have a mandate ($n = 5$).

Thirty-four of 45 respondents accept items for their IR from their campus only. Six respondents accept items from a university system and one accepts from a consortium. Four respondents collect from other entities, and these include, per write-in, regional community, campus community and external partners, university community, ejournal authors and editors. We also asked how many universities or institutions may deposit in the IRs. Four respondents wrote that two institutions may deposit into their IRs and two respondents wrote that more than five institutions may deposit into their IRs.

Forty-three respondents gave us the age of their IR and forty-six respondents provided data on software licensing. As of January 1, 2012, the minimum age of an IR was 153 days old. The median age was 4.59 years and the mean age was 4.39 years. The oldest IR was 8.50 years old. Calculations are based on a year equaling 365.24 days. Although we did not ask for specific software titles or services, twenty-nine respondents noted they use an open source solution for their IR and seventeen respondents use a proprietary solution.

Forty-five respondents answered our question about the size of their IR. Twenty-eight report more than 5,000 items and 17 participants have 4,000 or less with no one reporting in the 4,000 range. Librarians associated with the repository are primarily responsible for handling submissions ($n = 39$) rather than allowing content creators to self-archive ($n = 6$). Twenty institutions offer additional services, such as digitization, copy editing, project and data management, copyright clearance, bulk import, and metadata creation while 25 offer no additional services.

We asked our respondents if they collect web traffic data and, if so, about end-user usage of their IRs, including total visits, in-site searches, which may be a conservative indicator of usage (Chen, Kochtanek, Burns, & Shaw, [2010](#)), and total items retrieved or downloaded. Forty respondents said they collect web traffic data and four said they do not. Twelve to eighteen of those who do collect this data provided some numbers of their various usage statistics; Table 1 summarizes this data. Overall, the results indicate wide-ranged and skewed distributions of end-user engagement of IRs.

Table 1: End-User Usage Statistics

<i>End-User Usage Type</i>	<i>n</i>	<i>Min</i>	<i>Mdn</i>	<i>M</i>	<i>Max</i>
Visits	18	1,000	65,372	1,132,796	5,500,000
In-site Searches	12	55	7,854	83,578	848,204
Items Retrieved	17	1,234	432,332	1,822,348	9,821,234

Since the *Items Retrieved* indicator, by all measures, in Table 1 is much higher than the visit or in-site search indicators, we suspect this means that users may be retrieving content through third party discovery services without visiting the IR directly. Arlitsch & O'Brien's (2012) description of Google Scholar, which provides a direct link to the source document and not to the IR's page for the document, may explain this finding. If so, web log analytics may under report IR use. Work by the University of Edinburgh, which uses DSpace for its digital library, in cooperation with Google Analytics may help resolve this problem. By manipulating a piece of the application programming interface (API), programmers created a redirect so a user's click on a PDF link in a Google search results list will take the user to the file's item page in the IR, rather than directly to the file (Knowles, 2012).

The IRs host a great variety of content and format types. Table 2 lists the types of content we specifically asked about and the number or count of respondents, along with the question response rate, who said they hold that type of content. For comparison, we also provide the number of respondents who see downloads or retrievals of the types of content.

Table 2: Content Type, Types of Content Held and Retrieved by Total Number of Respondents with Response Rate

<i>Content Type</i>	<i>Count of IRs, Items Held</i>	<i>Survey Question Response Rate</i>	<i>Count of IRs, Items Retrieved</i>	<i>Survey Question Response Rate</i>
Courseware (including syllabi)	14	31%	13	43%
Data sets	23	51%	13	43%
Other	25	56%	15	50%
Books	29	64%	18	60%
Book chapters	35	78%	18	60%
Tech reports, working papers	39	87%	22	73%
Conference articles	40	89%	21	70%
Presentations (e.g., PPT slides)	41	91%	24	80%
Theses and dissertations	43	43%	27	90%
Journal articles	44	44%	24	80%

We asked respondents to write-in other types of content held, and we saw a diversity of material including

newsletters, video and audio, images, grant proposals, university ephemera, plays, poetry, software code, patents, brochures, podcasts (including lectures), webcasts, maps, archival material, reports, oral histories, whole journals, videos of lectures, and honors theses (including undergraduate). These findings generally agree with the findings from Lynch & Lippincott (2005) except that our survey had a much higher number of respondents who collect conference and journal articles. Ten respondents also noted that a number of these other types of items have also been downloaded or retrieved.

Respondents provided data on the number of full time equivalent (FTE) staff that work on the IR. Professional staff, including librarians, IT professionals, and administrators, hold the most responsibility ($n = 38$; $Min = 0$; $Mdn = 1.875$; $M = 2.82$; $Max = 12$) when compared to non- or para-professional or clerical staff ($n = 31$; $Min = 0$; $Mdn = 1$; $M = 1$; $Max = 9$) and student workers ($n = 37$; $Min = 0$; $Mdn = 0.5$; $M = 0.81$; $Max = 7$). These numbers suggest that the development and maintenance of IRs is still considered a high-level task, not yet routine enough to fall into the purview of the paraprofessional, unlike a growing number of other jobs in libraries (Zhu, 2012).

We asked whether the work to maintain the institutional repository resided in a single library department / unit or whether the responsibility was spread across multiple departments / units. Nineteen respondents confirmed that the responsibility is maintained by a single department / unit and 20 respondents noted that the work is spread out. The most frequent single departments responsible for an IR include Scholarly Communication, Technical Services, Collections Management / Development, Electronic Resources, and Digital Scholarship / Services / Resources. When responsibility for the IR is spread out, respondents described the additional departments / units as Digital Library, Library IT Special Collections and Archives, Publishing / Scholarly Communication and Information Technology, Cataloging and Metadata Services, Resource Management, Library Technology or Library Information Technology Services, Digital Initiatives, Theses and Dissertation Office, Preservation — Conservation, President's Office, Digital Resources and Publishing, Digital Repositories Services Archives, Reserves, Campus IT, Bibliographic and Metadata Services Team, and Operations Team. Respondents also made note of individual participants such as liaisons from schools with OA policies, subject liaisons, various academic units, graduate students, and technical services staff.

The implementation / start-up costs and the annual operating costs of the IRs formed a major section of the survey. There is a rather substantial variance in most of the reported values and we report the descriptive statistics in Table 3. A number of respondents entered zero for these questions, and we exclude those from the calculations (see Appendix A for the calculations with zeros included). We propose that the nonzero calculations may be more indicative of the sampled values. Note that the annual data for personnel, software, and hardware will not equal the amount in the *Annual* row. This table represents three separate survey questions about total implementation costs, total annual costs, and total annual by category (personnel, software, and hardware), and not all respondents provided a breakdown.

Table 3: Implementation and Annual Reported Costs Plus Annual Reported Costs for Personnel, Software, and Hardware

Costs	<i>n</i>	<i>Min</i>	<i>Mdn</i>	<i>M</i>	<i>Max</i>
Implementation	17	\$1,200	\$25,000	\$52,100	\$300,000
Annual	20	\$500	\$31,500	\$77,300	\$275,000
Personnel, Annual	17	\$100	\$70,000	\$86,186	\$235,200
Software, Annual	10	\$2,500	\$23,000	\$22,350	\$40,000
Hardware, Annual	6	\$500	\$5,500	\$13,250	\$50,000

Table 4 provides the median implementation and annual operating costs by various categories. The categories include software type, the [Carnegie Foundation's Basic Classification categories](#) for doctorate-granting universities and master's colleges and universities, repository and institution size, whether a mandate exists at the institution, whether additional library services are offered, and whether librarians are primarily responsible for depositing content in the IR or authors/producers hold that responsibility (collection development). We report the median calculations since they are a more robust measure of central tendency. Calculations only include nonzero entries (see [Appendix B](#) for the calculations with zeros included). All numbers have been rounded to the nearest dollar amount.

Table 4: Median Costs by Implementation and Annual Operating Expenditures

	<i>Implementation Costs</i>		<i>Annual Operating Costs</i>	
	<i>n</i>	<i>Mdn</i>	<i>n</i>	<i>Mdn</i>
Software Type				
Open Source	8	\$20,500	9	\$31,000
Proprietary	9	\$30,000	11	\$32,000
Carnegie Foundation's Basic Classification Categories				
RU/VH	4	\$60,000	5	\$116,000
RU/H	5	\$21,000	6	\$31,500
DRU	3	\$100,000	3	\$100,000
Master's/L	— —	— —	— —	— —
Master's/M	2	\$17,500	2	\$17,500
Master's/S	— —	— —	1	\$14,000
Other	3	\$13,500	3	\$17,000
Size of Repository				
<1,000	1	\$40,000	2	\$27,000
1,001 to 2,000	3	\$13,500	3	\$13,500
2,001 to 3,000	2	\$20,000	2	\$24,500
3,001 to 4,000	3	\$25,000	3	\$31,000
4,001 to 5,000	— —	— —	— —	— —
> 5,000	8	\$35,000	10	\$150,000
Size of Student Body				
< 5,000 students	2	\$11,750	3	\$14,000
5,001 to 10,000	3	\$10,000	3	\$7,000
10,001 to 15,000	2	\$17,500	2	\$31,500
15,001 to 20,000	1	\$21,000	1	\$21,000

> 20,000 students	9	\$40,000	11	\$150,000
Mandate Status				
Mandate, Yes	1	\$20,000	1	\$16,000
Mandate, No	15	\$25,000	18	\$31,500
Additional Services Offered				
Yes	8	\$19,250	10	\$65,500
No	9	\$30,000	10	\$26,500
Mediated (e.g., by librarians) or Self Depositing (e.g., by authors)				
Mediated	14	\$27,500	16	\$31,500
Self Depositing	3	\$21,000	4	\$85,500

Interesting differences in the data pertain to the size of the institution (in terms of number of students), whether additional services are offered, and whether librarians or other professionals are responsible for collection development (depositing in the IR) or whether authors or producers hold that responsibility. In terms of institution size (*Size of Student Body*), implementation and annual operating costs seem to be much greater for larger institutions than for smaller. In agreement with the previous studies mentioned, the median annual operating costs seem to be much higher when additional services are offered, although initial costs seem to be much lower for IRs that offer additional services. Interestingly, the reported median cost to operate IRs is much lower for IRs where librarians or other professionals are responsible for collection development (mediated depositing) than when authors or producers hold that responsibility. This may suggest a certain level of efficiency or simply may be a matter of the unequal number of respondents who answered. However, if it is generally true, then it is a reverse of that estimated by Bevan (2007).

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 more integrated into sustainable budget categories. 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 ($n = 1$).

Tables 5a, 5b, and 5c represent usage statistics by start-up and annual costs of the IR. It notes only the statistics for those respondents who contributed to all three respective categories, such that, if respondent A gave total visits AND start-up costs AND annual costs, then we provide the calculations for the set of all such respondents. We repeat this for in-site searches and items downloaded or retrieved.

Table 5a: Total Visits by Implementation Costs and Annual Costs

<i>n</i> = 8	<i>Visits</i>	<i>Implementation Costs</i>	<i>Annual Costs</i>
<i>Mdn</i>	27,500	\$25,500	\$35,500
<i>M</i>	1,140,760	\$32,150	\$79,188

Table 5b: Total In-site Searches by Implementation Costs and Annual Costs

<i>n</i> = 3	<i>In-site Searches</i>	<i>Implementation Costs</i>	<i>Annual Costs</i>
<i>Mdn</i>	4,337	\$20,000	\$16,000
<i>M</i>	5,254	\$14,067	\$12,500

Table 5c: Total Items Retrieved / Downloaded by Implementation Costs and Annual Costs

<i>n</i> = 7	<i>Retrieved / Downloaded</i>	<i>Implementation Costs</i>	<i>Annual Costs</i>
<i>Mdn</i>	215,875	\$30,000	\$32,000
<i>M</i>	963,179	\$27,286	\$78,714

5 Discussion

The most important lesson learned from this survey is that not all institutional repositories are alike. This has made it difficult to conduct a blanket survey. Future explorations of the costs and values of institutional repositories should be targeted to certain population types, such as by Carnegie classification category. Such targeting might offer a better understanding of IRs at various types and sizes of institutions. We feel it necessary to point out a caution: a number of smaller institutions responded to the survey. This suggests a need to understand the perceived costs and value of IRs at smaller institutions. Overall, the diversity of institutions with IRs should imply a healthy trend of their usage.

A surprising find was the near equal median expenses between annual operating costs for institutions that use open source software and institutions that use proprietary solutions. We did not receive enough responses to provide a breakdown of personnel, software, and hardware annual costs by software type, but this should be explored in future studies. One respondent noted extreme unhappiness with a proprietary solution. No additional comments about software solutions were provided by other respondents.

Given Bevan's (2007) estimation that self-depositing (or self-archiving) IR models cost less than IR models that depend on librarian or professional mediated depositing and our finding that reports the opposite, further inquiry in the roles and costs involved with mediated models should be pursued. If it can be confirmed that mediated depositing by librarians or other professionals results in a more efficient work flow, higher quality metadata, or some other advantage and at the same time reduces costs, then this would lend support to the argument T.S. Plutchak (2012) offers about the importance of the professional librarian in the future of academic libraries.

6 Conclusion

No one can predict the future of institutional repositories at this time and it remains to be seen if individual institutional management of a repository is the most efficient and effective means of operation. A question that should be asked of the users of repositories is whether their needs are met by the dispersed model of repositories that exists at the present time, or if some kind of unification (or at least unified search and retrieval capability) would be more useful. The latter sort can be exemplified by the [arXiv](#) project at Cornell University, which provides access to more than 700,000 papers. The results of this cost examination can assist decision making on many campuses—for librarians and for institutional administrators. No one wants institutional repositories to be reduced to a solution in search of a problem; the objective of the project is to meet the genuine needs manifested in campus communities.

7 References

- [1] Aguillo, I.F., Ortega, J.L., Fernández, M., & Utrilla, A.M. (2010). Indicators for a webometric ranking of open access repositories. *Scientometrics*, 82(3), 477-486. <http://dx.doi.org/10.1007/s11192-010-0183-y>
- [2] Albanese, A., & Oder, N. (2008). [Harvard faculty unanimously agree to establish open access repository](#). *Library Journal*.
- [3] Arlitsch, K., & O'Brien, P.S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech*, 30(1), 60-81. <http://dx.doi.org/10.1108/07378831211213210>
- [4] Barton, M.R., & Walker, J.H. (2003). Building a business plan for DSpace, MIT Libraries digital institutional repository. *Journal of Digital Information*, 4(2). <http://hdl.handle.net/1721.1/26700>
- [5] Bevan, S.J. (2007). Developing an institutional repository: Cranfield QUEprints — a case study. *OCLC Systems & Services*, 23(2), 170-182. <http://dx.doi.org/10.1108/10650750710748478>
- [6] Carpenter, M., Graybill, J., Offord Jr., J., & Piorun, M. (2011). Envisioning the library's role in scholarly communication in the year 2025. *Portal: Libraries and the Academy*, 11(2), 659-681.
- [7] Chan, L. (2004). [Supporting and enhancing scholarship in the digital age: The role of open-access institutional repositories](#). *Canadian Journal of Communication*, 29(3).
- [8] Chen, H., Kochtanek, T., Burns, C., & Shaw, R. (2010). Analyzing users' retrieval behaviors and image queries of a photojournalism image database. *The Canadian Journal of Information and Library Science*, 34(3), 249-268. <http://dx.doi.org/10.1353/ils.2010.0003>
- [9] Chen, X. (2010). Google Scholar's dramatic coverage improvement five years after debut. *Serials Review*, 36(4), 221-226. <http://dx.doi.org/10.1016/j.serrev.2010.08.002>
- [10] Cothran, T. (2011). Google Scholar acceptance and use among graduate students: A quantitative study. *Library & Information Science Research*, 33(4), 293-301. <http://dx.doi.org/10.1016/j.lisr.2011.02.001>
- [11] Crow, R. (2002). [The case for institutional repositories: A SPARC position paper \(Release 1.0\)](#).
- [12] Cullen, R., & Chawner, B. (2009). [Institutional repositories and the role of academic libraries in scholarly communication](#). Paper presented at the Asia-Pacific Conference on Library & Information Education & Practice.
- [13] Darby, R.M., Jones, C.M., Gilbert, L.D., & Lambert, S.C. (2009). Increasing the productivity of interactions between subject and institutional repositories. *New Review of Information Networking*, 14(2), 117-135. <http://dx.doi.org/10.1080/13614570903359381>
- [14] 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). <http://dx.doi.org/10.1045/march2007-davis>
- [15] Davis, P.M., & Fromerth, M.J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*, 71(2), 203-215. <http://dx.doi.org/10.1007/s11192-007-1661-8>
- [16] DuraSpace. [Google Scholar Metadata Mappings](#). n.d.
- [17] Edgar, B.D., & Willinsky, J. (2010). [A survey of scholarly journals using open journal systems](#). *Scholarly and Research Communication*, 1(2).
- [18] Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-Selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE*, 5(10):e13636+. <http://dx.doi.org/10.1371/journal.pone.0013636>
- [19] Giesecke, J. (2011). Institutional repositories: Keys to success. *Journal of Library Administration*, 51(5-

6), 529-542. <http://dx.doi.org/10.1080/01930826.2011.589340>

[20] Google. [Webmaster Tools](#). n.d.

[21] Greco, A.N., Wharton, R.M., Estelami, H., & Jones, R.F. (2006). The state of scholarly publishing: 1981-2000. *Journal of Scholarly Publishing*, 37(3), 155-214.

[22] Guédon, J-C. (2008). Mixing and matching the green and gold roads to open access—Take 2. *Serials Review*, 34(1), 41-51. <http://dx.doi.org/10.1016/j.serrev.2007.12.008>

[23] Harnad, S. [What is Open Access?](#). n.d.

[24] Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Hajjem, C., & Hilf, E.R. (2008). The access/impact problem and the green and gold roads to open access: An update. *Serials Review*, 34(1), 36-40. <http://dx.doi.org/10.1016/j.serrev.2007.12.005>

[25] Heath, F.M. (2009). The University of Texas: Looking forward: Research libraries in the 21st century. *Journal of Library Administration*, 49(3), 311-324. <http://dx.doi.org/10.1080/01930820902785413>

[26] Kim, J. (2011). Motivations of faculty self-archiving in institutional repositories. *The Journal of Academic Librarianship*, 37(3), 246-254. <http://dx.doi.org/10.1016/j.acalib.2011.02.017>

[27] Knowles, C. (2012). [Surfacing Google Analytics in DSpace](#). n.d.

[28] Li, Y., & Banach, M. (2011). Institutional repositories and digital preservation: Assessing current practices at research libraries. *D-Lib Magazine*, 17(5/6). <http://dx.doi.org/10.1045/may2011-yuanli>

[29] Lynch, C.A. (2003). [Institutional repositories: Essential structure for scholarship in the digital age](#). *ARL Bimonthly Report*, No 226.

[30] Lynch, C.A., & Lippincott, J.K. (2005). Institutional repository deployment in the United States as of early 2005. *D-Lib Magazine* 11(9). <http://dx.doi.org/10.1045/september2005-lynch>

[31] MIT Libraries. [MIT's DSpace Experience: A Case Study](#). n.d.

[32] Moore, G. (2011). [Survey of University of Toronto Faculty Awareness, Attitudes, and Practices Regarding Scholarly Communication: A Preliminary Report](#). Toronto: University of Toronto.

[33] Mower, A., & Chauffy, L. (2009). Do something no one has imagined: The 2008 SPARC Digital Repositories meeting. *College & Research Libraries News*, 70(3), 158-160.

[34] Norris, M., Oppenheim, C., & Rowland, F. (2008). Finding open access articles using Google, Google Scholar, OAIster and OpenDOAR. *Online Information Review*, 32(6), 709-715. <http://dx.doi.org/10.1108/14684520810923881>

[35] Pascarelli, A.M. (1990). Coping strategies for libraries facing the serials crisis. *Serials Review*, 16(1), 75. [http://dx.doi.org/10.1016/0098-7913\(90\)90045-Z](http://dx.doi.org/10.1016/0098-7913(90)90045-Z)

[36] Piorun, M., & Palmer, L.A. (2008). Digitizing dissertations for an institutional repository: A process and cost analysis. *Journal of the Medical Library Association*, 96(3), 223-229. <http://dx.doi.org/10.3163/1536-5050.96.3.008>

[37] Plutchak, T.S. (2012). Breaking the barriers of time and space: The dawning of the great age of librarians. *Journal of the Medical Library Association*, 100(1), 10-19. <http://dx.doi.org/10.3163/1536-5050.100.1.004>

[38] Prosser, D. (2003). [Institutional repositories and Open Access: The future of scholarly communication](#). *Information Services & Use*, 23(2003), 167-170.

[39] R Development Core Team. (2011). [R: A language and environment for statistical computing](#). R

Foundation for Statistical Computing, Vienna, Austria.

[40] 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).

<http://dx.doi.org/10.1045/november2007-rieh>

[41] Shotton, D. (2012). The five stars of online journal articles — a framework for article evaluation. *D-Lib Magazine*, 18(1/2). <http://dx.doi.org/10.1045/january2012-shotton>

[42] Tansley, R., & Harnad, S. (2000). [Eprints.org software for creating institutional and individual open archives](#). *D-Lib Magazine*, 6(10).

[43] 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).

<http://dx.doi.org/10.1045/september2007-mcdonald>

[44] Waaijers, L., Savenije, B., & Wesseling, M. (2008). [Copyright angst, lust for prestige and cost control: What institutions can do to ease open access](#). *Ariadne*, 57.

[45] Wacha, M., & Wisner, M. (2011). Measuring value in open access repositories. *The Serials Librarian*, 61(3-4), 377-388. <http://dx.doi.org/10.1080/0361526X.2011.580423>

[46] Xia, J. (2008). A comparison of subject and institutional repositories in self-archiving practices. *The Journal of Academic Librarianship*, 34(6), 489-495. <http://dx.doi.org/10.1016/j.acalib.2008.09.016>

[47] Zhu, L. (2012). The role of paraprofessionals in technical services in academic libraries. *Library Resources & Technical Services*, 56(3), 127-154.

Appendix A

Table 3 with Zero Values Included. Implementation and Annual Reported Costs Plus Annual Reported Costs for Personnel, Software, and Hardware

Costs	n	Min	Mdn	M	Max
Implementation	21	\$0	\$20,000	\$42,176	\$300,000
Annual	21	\$0	\$31,000	\$73,619	\$275,000
Personnel, Annual	33	\$0	\$0	\$44,000	\$235,200
Software, Annual	34	\$0	\$0	\$6,574	\$40,000
Hardware, Annual	34	\$0	\$0	\$2,338	\$50,000

Appendix B

Table 4 with Zero Values Included. Median Costs by Implementation and Annual Operating Expenditures, Zero Values Included

		Implementation Costs	Annual Operating Costs
	n	Mdn	Mdn
Software Type			

Open Source	10	\$15,000	\$26,000
Proprietary	11	\$25,000	\$32,000
Carnegie Classification			
RU/VH	6	\$25,000	\$78,000
RU/H	6	\$13,000	\$31,500
DRU	3	\$100,000	\$100,000
Master's/L	—	—	—
Master's/M	2	\$17,500	\$17,500
Master's/S	1	\$0	\$14,000
Other	3	\$13,500	\$17,000
Size of Repository			
< 1,000	2	\$20,000	\$27,000
1,001 to 2,000	3	\$13,500	\$13,500
2,001 to 3,000	2	\$20,000	\$24,500
3,001 to 4,000	3	\$25,000	\$31,000
4,001 to 5,000	—	—	—
> 5,000	11	\$21,000	\$150,000
Size of Student Body			
< 5,000 students	3	\$10,000	\$14,000
5,001 to 10,000	3	\$10,000	\$7,000
10,001 to 15,000	2	\$17,500	\$31,500
15,001 to 20,000	1	\$21,000	\$21,000
> 20,000 students	12	\$35,000	\$133,000
Mandate Status			
Mandate, Yes	1	\$20,000	\$16,000
Mandate, No	19	\$13,500	\$31,000
Additional Services Offered			
Yes	10	\$11,750	\$65,500
No	11	\$21,000	\$21,000
Mediated (e.g., by librarians) or Self Depositing (e.g., by authors)			
Mediated	17	\$13,500	\$31,000
Self Depositing	4	\$20,500	\$85,500

About the Authors



C. Sean Burns is a PhD Candidate at the University of Missouri's School of Information Science and Learning Technologies, where he also received his M.A. in Library Science. His dissertation uses social bookmarks of bibliographic references to examine the impact on academic library collections. He is a 2012 recipient of a Beta Phi Mu Eugene Garfield Doctoral Dissertation Fellowship.



Amy Lana is a MA candidate in the School of Information Science and Learning Technologies at the University of Missouri and the head of the monographs acquisitions unit at the MU Libraries. Her research interests include collection development, institutional repository management and scholarly communication. Amy also holds an MA in Classical Studies from Loyola University, Chicago.



John M. Budd is a professor in the School of Information Science & Learning Technologies of the University of Missouri. He has also been on the faculties of Louisiana State University and the University of Arizona. He has published widely on issues related to academic libraries.

Copyright © 2013 C. Sean Burns, Amy Lana, John M. Budd

PRINTER-FRIENDLY FORMAT

[Return to Article](#)
