Perceived Values and Benefits of Institutional Repositories: A Perspective of Digital Curation

Soo Young Rieh, Karen Markey, Elizabeth Yakel, Beth St. Jean, and Jihyun Kim
School of Information, University of Michigan
304 West Hall, 1085 South University Avenue, Ann Arbor, MI 48109-1107
{rieh, ylime, yakel, bstjean, jhkz}@umich.edu

Abstract

Institutional repositories (IRs) are increasingly deployed in academic institutions to manage a variety of digital content including educational, research, and archival materials. The proposed benefits of IRs have been identified in the literature including increased knowledge sharing, control over the digital assets of the university, and preservation. All of these benefits involve digital curation because they span the information life-cycle and involve achieving goals that are central to digital curation. In this paper, we report on an empirical study examining how library directors and others involved in IRs articulate their benefits. The purpose of this study is two fold: (1) to investigate whether IR administrators and staff agree on a set of values and benefits that IRs can offer and (2) to examine the extent to which IR administrators and staff understand the role of digital curation in the process of IR establishment. Our underlying theory is that how people conceptualize the benefits of IRs effects all other digital curation decisions. Therefore, how staff perceive the value and benefits is one of the fundamental questions to be addressed in IR development. Early identification of explicit benefits and value would assist IR staff in justifying the establishment of an IR to the larger university and provide a framework for IR development. The study is part of a larger research project Making Institutional Repositories a Collaborative Learning Environment (MIRACLE).

Introduction

Institutional repositories (IRs) are increasingly deployed in academic institutions to manage a variety of digital content including educational, research, and archival materials. The benefits of IRs have been touted by many authors and include increased knowledge sharing Yeates (2003), control over the digital assets of the university (Crow 2002a, 2002b), and preservation (Lynch 2003). Gibbons (2004) cites major benefits such as stewardship, efficiencies, showcasing an institution and wider distribution as compelling reasons for establishing an IR. All of these benefits involve digital curation because they span the information life-cycle and involve achieving goals that are central to digital curation, such as “interoperability with the future” and “communication across time” (Rusbridge et al. 2005).

In this paper, we report on an empirical study that examined the benefits and values of IRs as articulated by library directors and others involved in IRs in the United States. The purpose of this study is two fold: (1) to investigate whether IR administrators and staff agree on a set of values and benefits that IRs can offer; (2) to examine to what extent IR administrators and staff understand the role of digital curation in the process of IR establishment. Our underlying theory is that how people conceptualize the benefits of IRs effects all other digital curation decisions. Therefore, how staff perceive the value and benefits is one of the fundamental questions to be addressed in IR development. Early identification of explicit benefits and value would assist IR staff in justifying the establishment of an IR to the larger university and provide a framework for IR development. The study is part of a larger research project Making Institutional Repositories a Collaborative Learning Environment (MIRACLE) (http://miracle.si.umich.edu).

Literature Review

Over the past five years, the implementation of IRs has been growing rapidly and the publications on IRs have flourished accordingly. The authors of this paper recently published a report on a census of institutional repositories in U.S. academic institutions (Markey, Rieh, St. Jean, Kim, and Yakel, 2007) which provides one of the most comprehensive overviews of current IR practices including staffing, finances, planning, system selection, policies, benefits, and beneficiaries. Among other findings, this census demonstrates that the institutional repository movement is widespread and not just confined to research libraries. We found that academic libraries in large and small colleges and universities, liberal arts, medical, and other technical universities can all see potential benefits.

Gibbons (2004) presented compelling reasons for why an organization would want to establish an IR including providing an infrastructure for preservation of digital content, lowering the barrier to document distribution, creating a centralized digital showcase in which research, teaching, and scholarship can be highlighted, and facilitating wider distribution. Yeates (2003) also listed the benefits of IRs, such as: extending the range of knowledge sharing, existing investment in information and content management systems can be leveraged; and more flexible ways of scholarly communication are available. Academic institutions would also reap these benefits. IR proponents argue that they form the infrastructure for a new scholarly publishing paradigm that wrests control away from publishers and puts it back in the hands of the academy, increase visibility, prestige, and public value of contributors, maximize access to the results of publicly-funded research, and increase the number and diversity of scholarly materials that are collected and preserved by academic institutions (Crow 2002a, 2002b; Chan 2004). Given the number of previous studies on values and benefits of IR, it is time to learn more from IR staff about how they perceive these benefits based on their actual experiences with IR implementation and planning.

Methodology

Data were collected using two different research methods: web-administered questionnaires and telephone interviews. All of our instrumentation was pilot tested prior to deployment. The survey was done in two phases. In the first phase, we sent 2,147 emails to academic library directors or senior administrators asking them to categorize their stage in IR development as either: (1) no planning to date, (2) only planning to date, (3) both planning and pilot testing one or more IR systems, or (4) public implementation of an IR system. This results of the email indicated that 48 (10.8%) respondents had implemented (IMP) an IR, 70 (15.7%) were actively planning and pilot testing IRs (PPT), 92 (20.6%) were only planning (PO), and 236 (52.9%) institutions had done no IR planning to date (NP). In the second phase (April - June 2006), we then tailored the survey instrument, administered online through SurveyMonkey, to these initial responses. In the end, we received questionnaires from 446 library directors and administrators – a response rate of 20.8%. When we contacted library directors and senior library administrators via email to conduct the IR census, we asked them to pass our questionnaire to the staff member who was most familiar with the IRs. Interestingly, library directors tended to be the ones who filled out the survey (N=288, 73.7%). Other survey respondents included library staff (N=40, 10.2%), assistant-associate library directors (N=31, 7.9%), archivists (N=10, N=2.3%), and chief information officers (CIOs) (N=10, 2.0%). Responses were initially collected and cleaned in Excel, then these data were migrated to SPSS for analysis.

As a follow-up of national census of IR, we conducted telephone interviews with the IR staff. These were intended to elicit a more in-depth understanding of the motivations behind IR planning and implementation which will better explain our census findings. In order to do this we

again developed several semi-structured interview protocols based on how respondents categorized their stage in IR development.

One of the final questions on the census asked survey respondents whether they would be willing to volunteer for a follow-up interview and to provide their names and email addresses. As a result, 176 respondents agreed to be interviewed. We planned to complete approximately 40 interviews, so we created a purposive sample from the 176 volunteers. The factors we took into account were: IR stage of development (from no planning or only planning to implementation and planning and pilot testing), the size and Carnegie classifications of parent institution (from small colleges to research universities), and the and the position of respondents. Based on these criteria, we recruited 36 participants. For the phone interviews, we also recruited interviewees from various positions including: library staff (N=11), library directors (N=9), assistant-associate library directors (N=4), archivists or directors of archives (N=4), heads or directors in libraries (N=4), CIOs (N=3), associate deans for research (N=1). In the final group of 36 interviewees, 16 were from Research Universities, 8 from Master's Colleges and Universities, 1 from a Doctoral/Research University, and 11 from Baccalaureate Colleges (The Carnegie Foundation for the Advancement of Teaching, 2006). Interviewees were also geographically dispersed: 6 from New England, 6 from the Mid-Atlantic, 10 from the Midwest, 6 from the south, 3 from the Southwest, 2 from the Rocky Mountains, and 3 from the West Coast.

Phone interviews were conducted from October to December 2006 and each interview took about one hour or 1.5 hours. An initial coding scheme was developed for the interviews and they were ingested into N-Vivo for coding and analysis. We did inter-coder reliability testing prior to full coding of these data.

Results

Findings about IR Benefits from Census

Since benefits are a key aspect of IRs, all four versions of the questionnaire listed the same 16 anticipated benefits and asked respondents to rate the importance of these benefits according to a Likert scale. We found that respondents were uniformly positive about the benefits of IRs. Adding up the percentages of "very" and "somewhat important" ratings equaled or exceeded 67% for all but two of the 16 benefits enumerated on questionnaires. Table 1 presents the kinds of benefits listed in the questionnaires and which benefits were perceived as most important, moderately important, and least important.

As seen in Table 1, in general "Capturing the intellectual capital of your institutions" was perceived to be the most important benefit of IRs regardless of the stage of IR implementation. After that, greater disagreement emerged between respondents in implementing (IMP) institutions and those in institutions with no planning (NP), only planning (PO), and planning and pilot testing (PPT). For example, NP, PO, PPT respondents chose "Better services to you institution's learning community" as the most important benefit, IMPs ranked it fifth. The IMPs perceived providing "Better service to contributors" as a very important benefit; however, this benefit was not recognized as important by NPs and POs. On the other hand, "Longtime preservation of your institution's digital out" was perceived to be an important benefit across IRs at every stage; ranking 3rd by NPs and POs and 5th by PPTs and IMPs.

Respondents' positive ratings of benefits varied in a systematic way. IMP respondents' ratings were almost always more positive than PPT respondents' ratings. Likewise, PPT respondents' ratings were almost always more positive than PO respondents' ratings. Finally PO respondents' ratings were generally more positive than those of NP respondents. Even though

NPs were not as positive as respondents engaged in some aspect of IRs (i.e., POs, PPTs, and IMPs), they were still rated IR benefits positively.

Table 1. IR Benefits

Top-ranked benefits (1 to 7)	**NP**	**PO**	**PPT**	**IMP**
Capturing the intellectual capital of your institution	2	2	2	1
Better service to contributors	(8)†	6	3	2
Exposing your institution's intellectual output to researchers around the world who would not otherwise have access to it through traditional channels	(9)	(9)	(7)	3
An increase in your library's role as a viable partner in the research enterprise	6	5	4	4
Longtime preservation of your institution's digital output	3	3	5	5T*
Better services to your institution's learning community	1	1	1	5T
A solution to the problem of preserving your institution's intellectual output	5	4	6	7
Middle-ranked benefits (8 to 14)	**NP**	**PO**	**PPT**	**IMP**
An increase in the accessibility to knowledge assets such as numeric, video, audio, and multimedia formats	(7)	8	8	8
A boost to your institution's prestige	14	13	10	9
Maintaining control over your institution's intellectual property	(4)	(7)	9	10
Contributing to the reform of the entire enterprise of scholarly communication and publishing	13	14	12	11
New services to learning communities beyond your institution	10	10	11	12
A reduction in the amount of time between discovery and dissemination of research findings	12	11	13	13
Providing maximal access to the results of publicly funded research	10	11	14	14
Bottom-ranked benefits (15 to 16)	**NP**	**PO**	**PPT**	**IMP**
An increase in citation counts to your institution's intellectual output	15	15	15	15
Reducing user dependence on your library's print collection	16	16	16	16

Note: NP=no planning, PO= planning only, PPT=planning and pilot testing, IMP=implementation

† Parentheses indicate NP, PO, and PPT benefits that deviated from IMP top, middle, or bottom ranks.

* T's indicate a ranked benefit that tied another benefit's weight.

These data beg the question—why should IMP respondents be more positive about IR benefits? We speculate that IMP respondents, having experienced the IR implementation effort from beginning to end, are more confident about IR benefits and express this confidence by giving benefits high ratings. Or, having invested much time and effort into IR implementation, IMP respondents want so much for the IR to succeed that they give it the highest ratings.

In addition to the extent of IR implementation, we found that the type of institution also made a more difference in the perceptions of IR benefits. It was noted that baccalaureate and master's institutions especially valued the ability of the IR to increase their institution's prestige. This benefit was less important to research universities. These universities felt especially strongly that IRs increased their library's role as a viable partner in the research enterprise.

Questionnaires also asked IMP respondents to examine IR benefits a second time, reassessing whether benefits were more or less important now that they had implemented an IR. The results are presented in Table 2. When respondents noted a change, the change was an increase in importance. The library's role as a viable research partner made the biggest jump, with almost 50% of IMP respondents rating this benefit as increasing in importance. Overall, 11 of 16 IR benefits register a 30%-or-more increase in importance between planning and implementation. This result reinforces our idea about the multifaceted nature of IR benefits.

Table 2. Increases in Benefits' Importance for Institutions with Operational IRs

Benefit	**% Increase**
An increase in your library's role as a viable partner in the research enterprise	48.7
Longtime preservation of your institution's digital output	35.0
An increase in the accessibility to knowledge assets such as numeric, video, audio, and multimedia datasets	35.0
Better service to contributors	34.2
Better services to your institution's learning community	34.2
A solution to the problem of preserving your institution's intellectual output	32.5
Exposing your institution's intellectual output to researchers around the world who would not otherwise have access to it through traditional channels	32.5
New services to learning communities beyond your institution	32.5
A boost to your institution's prestige	31.7
Capturing the intellectual capital of your institution	30.0
Maintaining control over your institution's intellectual property	30.0

Findings about IR Benefits from Phone Interviews

Interview participants also discussed the values of IRs and their comments fall into four categories: (1) Digital collection building; (2) Access to digital collections; (3) Use of IR materials; (4) Long-term preservation. These topics directly relate to the core issues of digital curation (Beagrie, 2006). Analyses of interview transcripts revealed that more specific themes emerged around these four categories: the value of unique collections, the importance of centralizing access, the efficiencies of sharing digital materials, and the IR as an intellectual preservation venue. Each topic and relevant theme will be discussed below.

(1) Unique Digital Collection Building

Analysis of interview transcripts indicates that the interview participants considered 'capturing the intellectual capital of the institution' as one of the important benefits of IR, which is consistent from the findings from the survey. They often mentioned the importance of providing "faculty and students a place to put their research and their output" (IMP12). Interviewees identified three distinct foci for collections in IRs: faculty e-prints, student work, and archival

primary source material. Each of these collection types is valued differently depending on the institution.

Faculty e-prints including working papers, preprints, journal articles, and conference presentation was most valued by large research universities. To the research universities, exposing the institution's research output to the world was valued highly as IMP2 described: "it's really nice to have them there for people outside the institution to see because these are kind of hidden resources that just sort of didapper into the archives even when they are catalogued and I think it's really important to share information."

The second type of digital collection is student work and includes doctoral dissertations, master's and undergraduate theses, student portfolios, and other student projects. Theses and dissertations are low-hanging fruit for many colleges and universities. It is a natural progression for a library to move from collecting paper dissertations to collecting digital ones. We found that Carnegie-classified master's and baccalaureate institutions highly valued master's and undergraduate theses. For institutions focusing on teaching, all student work provided potential content for the IR. "A lot of student research will never be published because college student research you know they work very hard at it but it often is not as comprehensive as would need to be to show up in a publication somewhere. And so therefore this is a modest form of publication for college students who wouldn't get that opportunity otherwise and so in that sense we look at it as something that's potentially positive because it's a way for students to get their work out and around beyond the boundaries themselves and their professors" (NP4). By and large, interviewees did not mention FERPA **(**Family Educational Rights and Privacy Act) or the legal and privacy issues involved in selecting student work as a primary focus of IR collection development.

The third type of collection is digitally reborn materials from the institutions archives or special collections. These include such materials as newspapers, photographs, manuscripts, and maps. Archival and manuscript materials were considered to be a "unique collection" (IMP20), and valued highly in both large and small institutions. In PPT13's institution, the special collections project was indeed the impetus of IR "because they were very proud of their special collections here, very deep in a few areas and they weren't getting used. And the director was kind of forward-looking and he thought the trend was to make your special collections viewable on the Internet and it was mostly for promotion of the library's collection."

(2) Centralizing access to Digital Collections

IR staff and administrators perceive making digital objects accessible any time anywhere as quite important. According to IMP20, both end users of and contributors to the IR view the most "exciting thing" about it as "providing access to those collections that no one would ever know exists." PPT13 pointed out that what is especially valuable is "a single repository" where "faculty will put their digital objects that they don't know what to do with and don't know how to deliver them." She continued: "We want one centralized location that if they're willing to put in the work then it will be there and could even get preserved. I think it's centralizing all of this."

(3) Use and Sharing of IR Materials

Once IRs have built digital collections and made them accessible centrally, researchers use the materials in a number of different ways. Several interview participants articulated how IR materials are or can be utilized by their learning community and even by general public. The expectation is that it will "give much more targets of retrievability of their materials" (OP6). In particular, IMP7 noted: "serendipitous discovery across disciplines that was not possible with all these materials either not being exposed to the web or being exposed from all different areas on

campus – their personal web presences." In addition, more frequent citations were mentioned as a positive outcome of IR use. PPT8 recalled a study which found people who contribute to IRs get cited three to five times more frequently than those who rely on the traditional publishing venues. She believed that IR would help faculty understand the importance of sharing resources and would also assist them to actually share those resources in a different way. "Your colleagues see it, now they can use it and they don't have to ask you because you've made it freely available. So I think we'll see more and more of that kind of usage" (PPT8).

(4) Institutional Repository as an Intellectual Preservation Venue

Compared to the responses about collection, access, and use, our interview participants did not seem to be so confident about the IR as a solution to the problem of preservation. They seemed to agree that preservation "is one of the biggest things" (IMP8) and that "IRs will be a part of that answer" (IMP13). But, that answer is not yet a reality. Interviewees raised concerns or reservations about IRs as a preservation venue for several reasons. First, interviewees pointed to the preservation issues associated with digital materials in general. IMP12 articulated the strongest comments: "You're making the promise that you're going to archive it and keep it forever. There's always a risk for that probably more so with digital material. There's a risk of failure or really not getting anybody to buy in." To some respondents still in the planning and pilot testing phase, preservation was not a priority in IR development. For instance, PPT13 said: "It [preservation] hasn't been one of the focuses. It's [the IR] mostly been for institutional consumption and curriculum use for the courses. We haven't had much feedback from faculty and anybody else on whether they think this is really an intellectual repository. It's kind of practical repository right now." PPT6 echoed these concerns: "I don't think we have yet figured out a really good way to ensure protection of digital information. There are too many variables." IMP13's comments are worthy of noting because she emphasized the critical role of the library in the process of IR development with respect to preservation issues. IMP13, an assistant director of library information technology, said: "Preservation has only been a library activity. ITS [Information Technology Services] doesn't seem to have any interest in that and they don't promise the faculty that there's any preservation component to this at all and so they're not offering preservation of these objects." He further stressed that he needed "To make sure that you get the library involved in this project and not just let ITS do it on their own because they're not going to worry about some of the things, like metadata and preservation. So, I think that the library has to stay in it."

Interviewees identified other values and benefits that are consistent with the findings from the census. For instance, the IR will help "people know about us not only for our research but also in terms of what we've been doing and IR development and we're pretty well respected" (IMP6). This relates to increasing an institutions prestige, a category used in the census.

Other comments in the interviews were closely related to 'an increase in your library's role as a viable partner in the research enterprise' used in the survey questionnaires. A few IMP interviewees noticed changes taking place in the institutions which they attributed to the IR. IMP16, for instance, stated that the IR provided the opportunity to begin interesting discussions between the library and the departments and the faculty on numerous issues such as: "How will your materials be accessed in the future? What do you want to have open access? How do you see scholars communicating with themselves in the future?" He said that "just starting these conversations at the discipline level has just been fantastic whether or not they participate in the repository or not" (IMP16).

Conclusion

The results of the survey and interviews demonstrate that IRs require a digital curation perspective in order to achieve the major benefits college and university libraries envision for them. IRs involve the management of digital assets through a large part of the information life cycle, particularly regarding (1) digital collection building, (2) access to digital collections, (3) use of IR materials, (4) long-term preservation. Our findings indicate that IR policies and practices for digital curation in terms of collection, access, and use vary depending on the size, Carnegie classification, and IR development stage (from no planning or only planning to planning and pilot testing and implementation). Although survey respondents ranked preservation between 3rd and 5th in importance, the interviews suggest that this is perceived as the hardest benefit to realize across all types of institutions. To date, much of IR literature views IRs as a new venue for collection development that extends a library's current purpose and goals. Our findings suggest that a broader digital curation perspective is needed with respect to IRs given the diverse collections they hope to attract and maintain long-term.

Acknowledgment

This work was supported by the Institute of Museum and Library Services (IMLS) through its National Leadership Grant Program (grant number LG-06-05-0126).

References

Beagrie, N. (2006). Digital Curation for Science, Digital Libraries, and Individuals. *The International Journal of Digital Curation, 1*(1), 3-16. Available: http://www.ijdc.net/./ijdc/article/view/6/5.

The Carnegie Foundation for the Advancement of Teaching. (2006). Carnegie Classifications: Lookup & Listings. Available from: http://www.carnegiefoundation.org/classifications/index.asp?key=807.

Chan, L. (2004). Supporting and Enhancing Scholarship in the Digital Age: The Role of Open-Access Institutional Repositories. *Canadian Journal of Communication, 29*, 277-300. Available: http://eprints.rclis.org/archive/00002590/01/Chan_CJC_IR.pdf.

Crow, R. (2002a). *The Case for Institutional Repositories: A SPARC Position Paper*. Available: http://www.arl.org/sparc/IR/ir.html.

Crow, R. (2002b). *SPARC Institutional Repository Checklist & Resource Guide*. Available: http://www.arl.org/sparc/IR/IR_Guide_v1.pdf.

Gibbons, S. (2004). Establishing an Institutional Repository. *Library Technology Reports, 40*(4), July-August 2004.

Lynch, C. A. (2003, February). Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report, 226*, 1-7. Available: http://www.arl.org/newsltr/226/ir.html.

Markey, K. Rieh, S. Y., St. Jean, B., Kim, J., & Yakel, E. (2007, February). *Census of Institutional Repositories in the United States: MIRACLE Project Research Findings.* Washington, D.C.: Council on Library and Information Resources. Available: http://www.clir.org/pubs/reports/pub140/pub140.pdf.

Rusbridge, C., Burnhill, P., Ross, S., Buneman, P., Giaretta, D., Lyon, L., & Atkinson, M. (2005). The Digital Curation Centre: A Vision for Digital Curation. From Local to Global Data Interoperability -- Challenges and Technologies. Mass Storage and Systems Technology Committee of the IEEE Computer Society. June 20-24, 2005. Sardinia, Italy. Available: http://www.dcc.ac.uk/docs/DCC_Sardinia_paper_final.pdf

Yeates, R. (2003). Over the Horizon: Institutional Repositories. *VINE: The Journal of Information and Knowledge Management Systems, 33*(2), 96-99.