Nationwide Census of Institutional Repositories: Preliminary Findings

Karen Markey
School of Information
University of Michigan
Ann Arbor, MI 48109 USA
01–734–763–3581

karen.markey@umich.edu

Beth St. Jean
School of Information
University of Michigan
Ann Arbor, MI 48109 USA
01-734-647-8041

bstjean@umich.edu

Soo Young Rieh
School of Information
University of Michigan
Ann Arbor, MI 48109 USA
01-734-647-8040

rieh@umich.edu

Elizabeth Yakel
School of Information
University of Michigan
Ann Arbor, MI 48109 USA
01-734-763-3569

yakel@umich.edu

Jihyun Kim
School of Information
University of Michigan
Ann Arbor, MI 48109 USA
01-734-647-8041

jhkz@umiich.edu

Yong-Mi Kim
School of Information
University of Michigan
Ann Arbor, MI 48109 USA
01-734-647-8041

kimym@umich.edu

ABSTRACT

The MIRACLE (Making Institutional Repositories A Collaborative Learning Environment) Project is a multi-phased research project that is investigating the development of institutional repositories (IRs) in American colleges and universities to identify models and best practices in the administration, technical infrastructure, and access to repository collections. This paper features preliminary findings from the project's first phase, a nationwide census that will reveal the extent of college and university involvement with IRs in the United States. Preliminary findings address the types of investigative activities that institutions are conducting prior to making the decision to implement an IR, perceived benefits of IRs, staffing for IRs, methods of recruiting digital content, characteristics of operational IRs including their costs, and respondents' next steps on the road to implementing an IR.

Categories and Subject Descriptors

H.3.7 [**Information Storage and Retrieval**]: Digital Libraries – *collection, standards, systems issues, user issues.*

General Terms

Management, Measurement, Performance.

Keywords

Institutional repositories, surveys, users, contributors, costs, benefits.

1. INTRODUCTION

The development of institutional repositories (IRs) is a very new enterprise for many colleges and universities nationwide. Successful IRs could revolutionize scholarly publication in learning communities, opening up access to research much earlier in the discovery process and reaching entirely new audiences [1, 2]. This paper documents preliminary findings from a nationwide census that investigates the extent of college and university involvement with IRs. The census is the first of several investigative activities undertaken by the MIRACLE (Making Institutional Repositories A Collaborative Learning Environment) Project in the course of accomplishing its goal of identifying specific factors contributing to the success of institutional repositories and effective ways of accessing and using repositories.

2. IR SURVEYS TO DATE

Originally MIRACLE Project investigators had proposed to survey *operational* institutional repositories (IRs) in North America; however, we were concerned that we would be duplicating the efforts of Charles Bailey and his associates who are working on an ARL (Association of Research Libraries) SPEC Kit that reports on the results of a recent survey of Association members regarding their IR activity [3]. Other surveys target specific user groups such as CNI members in the United States [4], CNI members abroad [5], CARL-member libraries [6], and early adopters of IR technology worldwide [7]. Examining these surveys' results, MIRACLE project investigators decided not to limit their efforts to a particular user group, membership, affiliation, or restrict participation to institutions with an operational IR. Instead, we sought to cast our net broadly and fill a void. Conducting a *census* of American academic institutions about their involvement with IRs, MIRACLE Project investigators will not be excluding institutions that have not jumped on the IR bandwagon. Being more inclusive will not only increase our confidence that we will be able to identify the wide range of practices, policies, and operations in effect at institutions where decision-makers are contemplating, planning, pilot testing, or implementing IRs but also enable us to learn why some institutions have ruled out IRs entirely.

3. PROJECT DESIGN

The MIRACLE Project features these five investigative activities over a three-year period:

1. Nationwide census of institutional repositories, 9 months, October 2005–June 2006

2. Telephone interviews with IR staff, 7 months, June 2006–December 2006

3. Case studies of five model IRs, 8 months, December 2006–July 2007
4. Survey of IR users, 9 months, August 2007–April 2008
5. Experimental study of IR searching, 9 months, January 2008–September 2008

Because the nationwide census was ongoing during the writing of this paper, this report of census findings is preliminary and subject to change in subsequent documents. Census findings are foundational and will help project investigators plan and execute future activities.

4. CENSUS METHODOLOGY

Project investigators are recruiting the library directors of four-year colleges and universities in the United States to participate in the census. We have purchased mailing lists of library director names and addresses from Information Today's American Library Directory Online and Thompson-Peterson's. After subtracting duplicates and community colleges, our sample is 2,147 email addresses. Using the results of a comprehensive literature search for inspiration for questions and answer categories, project investigators are drafting survey instruments and programming them on SurveyMonkey for distribution via the World-Wide Web. Project staff are sending email messages to library directors asking them to participate in the census by first characterizing the extent of their involvement with IRs as follows: (1) implementation of an IR (IMP), (2) planning & pilot testing an IR software package (PPT), (3) planning only (PO), or (4) no planning to date (NP). In response to their answer to this question, project staff send them a link to one of four survey instruments. Many of the same questions are listed across two, three, or all four survey instruments so that comparisons can be made based on the extent of institutions' involvement with IRs. Some directors complete questionnaires on their own and others delegate the task to decision-makers at their institution who are knowledgeable about their institution's plans for IRs.

Section 5 features respondents' answers to census questions received through mid-May 2006 that pertain to these topics:

• Investigative activities for decision-making
• Benefits of IRs
• IR staffing
• Recruiting IR content
• Respondents' IR(s)
• IR costs
• Respondents' next steps

5. PRELIMINARY CENSUS FINDINGS

5.1 Census Respondents

To date, project investigators have been successful sending email messages to library directors at 2,117 college and university libraries to invite them to participate in the nationwide census. We are researching names from 30 institutions that have bounced our email messages back to us as "undeliverable." Of the 273 respondents to date, 28 (10%) have characterized their IR involvement as IMP, 42 (15%) as PPT, 65 (24%) as PO, and 138 (51%) as NP. All of these respondents have completed questionnaires pertaining to one of these four levels of IR involvement.

5.2 Investigative Activities for Decision-Making

Three questionnaires (IMP, PPT, & PO) ask respondents to rate the importance of 12 investigative activities in influencing decisions about IR implementation. Because most respondents give high ratings to listed activities, we weight their ratings so that we can rank order them from top to bottom. Tables 1 and 2 list the top three and bottom three ranked activities per questionnaire type, respectively. Ranks lower than three are enclosed in parentheses.

Table 1. Top-ranked Investigative Activities

Activities	IMP	PPT	PO
Demonstrating operational IRs to my institution's decision-makers	1	(6)	(4)
Learning from reports of other institutions' IR planning, pilot testing and implementation activities to date	2	1	1
Learning about successful implementations at a wide range of institutions	3	3	(5)
Learning about successful implementations at comparable institutions	(4)	2	2
Learning about available expertise and assistance from a library consortium, network, group of libraries, etc.	(9)	(4)	3

All three respondent types are almost in agreement about the highest-ranked investigative activities. The most useful investigative activities are learning from reports of other institutions' IR planning and learning about successful IR implementations at comparable and a wide range of institutions. The response category "Learning about available expertise and assistance from a library consortium, network, group of libraries, etc.," is not as helpful to decision-makers at IMP institutions as it is to those at PPT or PO institutions.

Although bottom-ranked activities are in a different order (Table 2), the activities are the same. Decision-makers at the institutions who participated in this census want to get involved with IRs now rather than wait for a critical mass of IR implementation to happen.

Table 2. Bottom-ranked Investigative Activities

Activities	IMP	PPT	PO
Waiting for a critical mass of IR implementation at comparable institutions to happen	12	12	10
Waiting for a critical mass of IR implementation generally to happen	11	11	11
Identifying better digital preservation techniques	10	10	12

5.3 Benefits of IRs

All four questionnaires (IMP, PPT, PO, & NP) ask decision-makers to rate the anticipated importance of 16 benefits of IRs during the IR planning process. Because most respondents give high ratings to listed benefits, we weight their ratings so that we can rank order them from top to bottom. Tables 3 and 4 list the top two and bottom two ranked activities per questionnaire type. Ranks beyond the top or bottom two are enclosed in parentheses. T's indicate a ranked benefit that tied another benefit's weight.

Table 3. Top-ranked Benefits

Benefits	IMP	PPT	PO	NP
Capturing the intellectual capital of your institution	1	1T	2	(3)
Better service to contributors	2	1T	1	2
Longtime preservation of your institution's digital output	(5)	(3)	(3T)	1

Table 4. Bottom-ranked Benefits

Benefits	IMP	PPT	PO	NP
Reducing user dependence on your library's print collection	16	16	16	16
Providing maximal access to the results of publicly funded research	15	(14)	(11)	(10)
An increase in citation counts to your institution's intellectual output	(13)	15	15	15

IR decision-makers give top rankings to benefits that have a direct impact on their institution's learning community such as capturing the institution's intellectual capital, providing better service to IR contributors, and preserving the institution's digital output. Because benefits connected with the library, publicly funded research, and citation counts may indirectly affect one's local institution, decision-makers may have been reluctant to rate them as highly as other benefits.

Questionnaires for implementers (IMP) ask them whether certain benefits *increased* or *decreased* in importance following their implementation of an IR. These two benefits increased in importance—"an increase in your library's role as a viable partner in the research enterprise" and "longtime preservation of your institution's digital output."

5.4 IR Staffing

Several questions address staffing for IR planning, pilot testing, and implementation. For example, one question asks respondents what percentage of the responsibility for an operational IR has been given (IMP) or should be given (PPT and PO) to various campus units. Respondents type a percentage from 0% to 100% into a dialog box. Project staff average the percentages to make sense of the responses. Figure 1 summarizes results. Generally, PPT and PO decision-makers envision the library sharing operational responsibility for an IR. Decision-makers from institutions with full-fledged operational IRs choose responses that show library staff bearing the burden of responsibility for the IR.

Project investigators will follow up this finding in subsequent project activities such as phone interviews and case studies to determine the reasons why most of the responsibility for the operational IR eventually falls into the hands of library staff.
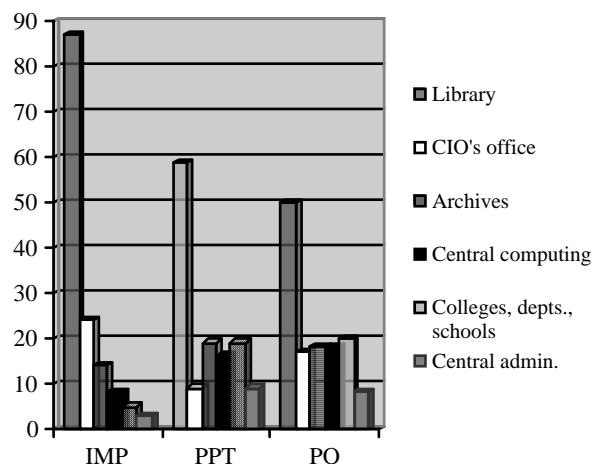

Figure 1. Percentage of Responsibility for the IR

5.5 Recruiting IR Content

PPT and PO questionnaires ask decision-makers what nine methods of recruiting digital content would be the most successful at their institution and the IMP questionnaire asks them how they would assess their success with these nine methods of recruiting such content. Because respondents give high ratings to listed methods, project staff weight their ratings so they can be rank ordered from top to bottom. Tables 5 and 6 list the top four and bottom two methods per questionnaire type, respectively. Table 6 uses parentheses to cite methods ranking higher than at the very bottom.

Table 5. Top-ranked Methods

Methods	IMP	PPT	PO
Staff responsible for the IR working one-on-one with early adopters	1	1	1
Presentations by staff responsible for the IR at departmental and faculty meetings	2	4	4
Personal visits by staff responsible for the IR to faculty and administrators	3	2	2
Word-of-mouth from early adopters to their colleagues in the faculty and staff ranks	4	3	3

All three respondent types are in agreement about the top- and bottom-ranked recruitment methods—"staff responsible for the IR working one-on-one with early adopters" (top) and "reducing user dependence on the library's print collection" (bottom). The rank order of three remaining top methods differs depending on respondent type. Respondents are not especially concerned that IRs could increase citation counts to their institution's intellectual output.

Table 6. Bottom-ranked Methods

Methods	IMP	PPT	PO	NP
Reducing user dependence on your library's print collection	16	16	16	16
Providing maximal access to the results of publicly funded research	15	(14)	(11)	(10)
An increase in citation counts to your institution's intellectual output	(13)	15	15	15

A follow-up question asks decision-makers who are (IMP) or who they think would be (PPT and PO) the major contributors to their institution's IR. Figure 2 summarizes results.
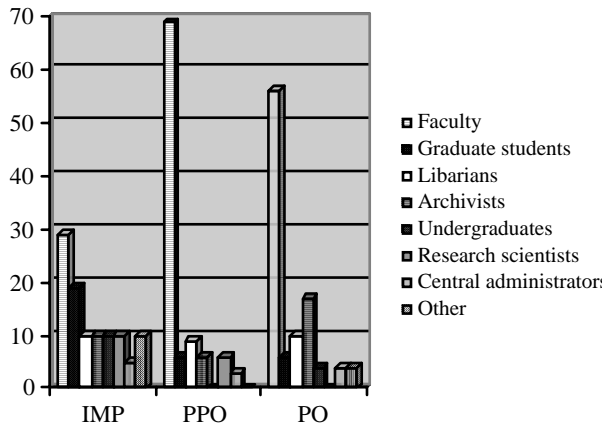

Figure 2. Contributors to the IR

Decision-makers at PPT and PO institutions expect that contributions by faculty will far outnumber contributions by other groups. In fact, faculty are contributing much less content to operational IRs, about 30% for the institutions in our survey and graduate students are not far behind at about 20%.

5.6 Respondents' Institutional Repositories

Of the 28 decision-makers whose institutions have operational IRs, 21 have one, four have two, and one has three such IRs available to members of their institution's learning community. Not all decision-makers identify their IR's software package by name. Those who do identify these packages: (1) 9 for Dspace, (2) 5 for bePress, (3) 4 for ProQuest's Digital Commons, (4) 2 for local solutions, and (5) 1 each for Ex Libris' DigiTools and Virginia Tech's ETD. Of the 42 decision-makers at institutions involved in IR planning and pilot testing, 31 estimate that they would make one, four estimate they would make two, and three estimate they would make three or more IRs available to members of their institution's learning community. Decision-makers are pilot testing these packages: (1) 17 for DSpace, (2) 9 for OCLC's ContentDM, (3) 5 for Fedora, (4) 3 each for bePress, DigiTool, ePrints, and Greenstone, (5) 2 each for Innovative Interfaces, Luna, and ETD, and (6) 1 each for Digital Commons, Encompass, a local solution, and Opus.

Decision-makers are asked to estimate the number of documents in their operational or pilot-tested IR. Table 7 gives the results.

Table 7. Numbers of Documents in IRs

Document range	IMP%	PPT%
Less than 501	41	67
501 to 1,500	12	8
1,501 to 2,000	12	4
2,000 to 3,000	12	4
3,001 to 5,000	6	8
Over 5,000	17	8
Total	100	100

Generally, operational IRs contain more documents than IRs in the pilot-testing phase. About two-thirds of pilot-tested IRs contain up to 500 documents. This proportion drops to about two-fifths for operational IRs. Over 5,000 documents are contained in 17% of the latter.

Decision-makers at IMP and PPT institutions are asked to rate the adequacy of 14 features of IR systems. Because they give high ratings to listed features, we weight their ratings so that we can rank order them from top to bottom. Tables 8 and 9 list the top three and bottom three features of IR systems, respectively. Ranks lower than three are enclosed in parentheses.

Table 8. Top-ranked IR-system Features

Features	IMP	PPT
Supported file formats	1	2
Scalability	2	(5)
Technical support	3	(8)
Formulating metadata for digital documents	(7)	3
Adherence to open access standards	(4)	1

Respondents do not agree on top-ranked features except for the "supported file formats" feature. They do, however, agree on bottom-ranked features. Interestingly, two of the three bottom-ranked features are likely to negatively impact end-user searching of IRs. This project's follow-up activities will probe whether end-user searching features of these systems generally may need major improvement and reasons for disagreement on top-ranked features.

Table 9. Bottom-ranked IR-system Features

Features	IMP	PPT
Authority control	14	13
Controlled vocabulary searching	13	12
Extensibility = Access to other campus systems and data	12	14

5.7 IR Costs

Several questions address costs of IRs. For example, one question asks decision-makers what percentage of their IR's annual budget is allocated to each of 7 line items. Respondents type a percentage from 0% to 100% into a dialog box. Project staff average the percentages for each line item to make sense of the responses. Figure 3 summarizes the results.
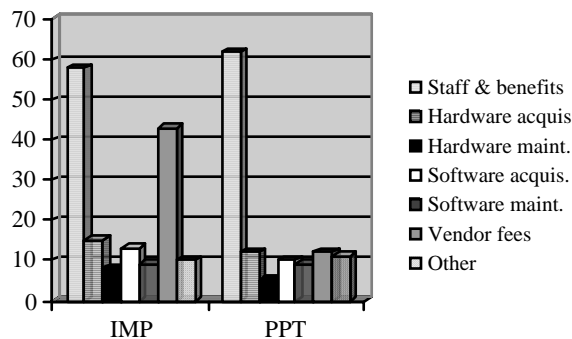
Figure 3. Costs of IRs

Except for vendors fees, decision-makers at IMP and PPT institutions agree on the relative distribution of costs. Vendor fees are not an insignificant line item, accounting for about 40% of the total budget for an IR at IMP institutions.

5.8 Respondents' Next Steps

Decision-makers at IMP institutions are asked how long they would stick with their IR system before migrating to a new one. Up to three years is the response of 54% and four to six years is the response of 31%. Only 15% say they would stick with their current IR for more than six years. On average, respondents will stay with their current system for four years before migrating.

Decision-makers at PPT and PO institutions are asked to assess their next steps connected with IR implementation. The top-ranked next step chosen by both respondent types is "Your institution supports implementation of an IR software package." Tied for the top rank amongst decision-makers at PO institutions is the step "Your institution widens the scope of its investigation into IRs." Ranked last for both respondent types is the step "Your institution terminates its investigation of IRs." Clearly, in terms of this project's respondents who are in the PPT and PO phases, the momentum is on the side of effort that eventually culminates in IR implementation.

We ask decision-makers at institutions where there is no planning for an IR to rate the importance of 15 reasons why no such planning has taken place to date. Their top two reasons are:

- Other priorities, issues, activities, etc., are more pressing than an IR, and
- We have no available resources to support planning.

Their bottom three reasons are:

- We do not understand or believe in the value or effectiveness of an IR
- We have no support from our library's administration, and
- We do not need an IR

These top and bottom reasons indicate that decision-makers at NP institutions are knowledgeable about IRs, have considered them

for their institution, and are probably too busy with other tasks including finding resources to support IR planning in the future. We can safely say that for our study's respondents, few had entirely dismissed the idea of IRs for their institution.

6. CONCLUSIONS

The results presented herein are based on data collected through mid-May 2006. In early June 2006, MIRACLE Project investigators will send one more set of reminders to prospective census participants and then terminate the census in mid June 2006. We are hesitant to draw conclusions here because are still actively collecting data in the first phase of a multi-phased project.

Subsequent project activities will include qualitative approaches to data collection such as phone interviews and case studies. Combining questionnaire results with qualitative data will enable us to gain greater insight into the preliminary findings this paper presents. Also we can pursue hypotheses regarding non-respondents to our email invitations, for example, contacting by phone decision-makers who have ruled out IR implementation entirely at their institution and probe their reasons for such a decision.

Please consult the MIRACLE Project's web site for more up-to-date information (http://miracle.si.umich.edu/) in the months and years to come.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Lim, E. Preprint servers: A new model for scholarly publishing? *Australian Acad. & Res. Libs., 27* (1996), 21–30.

[2] Prosser, D. Institutional repositories and open access: The future of scholarly communication. *Info. Servs. & Use, 23* (2003), 167–170.

[3] George, L. A. SPEC online surveys: SPEC survey on institutional repositories. Retrieved May 30, 2006 from http://createchange.org/spec/surveys/index.html

[4] Lynch, C. A. & Lippincott, J. K. Institutional repository deployment in the United States as of early 2005. *D-Lib Mag., 11,* 9 (Sept. 2005).

[5] Van Westrienen, G. & Lynch, C. A. Academic institutional repositories: Deployment status in 13 nations as of mid 2005. *D-Lib Mag., 11,* 9 (Sept. 2005).

[6] Shearer, K. Survey results: Summer 2004, CARL Institutional Repositories Project. Retrieved May 30, 2006 from http://www.carl-abrc.ca/projects/institutional_repositories/pdf/survey_results_2004-e.pdf

[7] Mark Ware Consulting Ltd. Publisher and Library/Learning Solutions (PALS): Pathfinder research on Web-based repositories: Final report. Retrieved May 30, 2006 from http://www.palsgroup.org.uk/palsweb/palsweb.nsf/