

# Institutional Repositories and the Institutional Repository: College and University Archives and Special Collections in an Era of Change

Elizabeth Yakel, Soo Young Rieh, Beth St. Jean, Karen Markey, and Jihyun Kim

## Abstract

This article reports on the archival implications of the 2006 *Census of Institutional Repositories in the United States*<sup>1</sup> and follow-up interviews with the developers of institutional repositories in selected colleges and universities. The findings indicate that archivists generally play a quiet but persistent role in institutional repositories, and archival and special collections materials are a major source of content in institutional repositories. Institutional repositories (IRs) are becoming an extension of the institutional repository (archives). Still, a great deal of uncertainty surrounds preservation in IRs, and a potential, albeit currently unfilled, role for the archivist exists in providing digital preservation expertise for the IR.

## Introduction

In this article, we report on the 2006 *Census of Institutional Repositories in the United States*<sup>2</sup> and subsequent interviews with the developers of selected institutional repositories. In particular, we will highlight the implications for archives

This research is part of a larger project—the Making Institutional Repositories a Collaborative Learning Environment (MIRACLE) project funded by the Institute of Museum and Library Services (IMLS) through its National Leadership Grant Program (grant number LG-06-05-0126-05). For more information about the project, see <http://miracle.si.umich.edu/>, accessed 1 May 2007.

<sup>1</sup> Karen Markey, Soo Young Rieh, Beth St. Jean, Jihyun Kim, and Elizabeth Yakel, *Census of Institutional Repositories in the United States: MIRACLE Project Research Findings*, Washington, D.C.: Council on Library and Information Resources, 2007, available at <http://www.clir.org/pubs/reports/pub140/pub140.pdf>, accessed 1 May 2007.

<sup>2</sup> Markey et al., *Census of Institutional Repositories in the United States*.

and special collections in colleges and universities as greater numbers of institutional repositories are implemented. Clifford Lynch describes institutional repositories (IRs) as “a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.”<sup>3</sup> Given this definition, IRs present both opportunities and challenges for college and university archivists. Will IRs be an opportunity to improve management of digital data for their institutions? Are IRs a sign that archivists are losing control of electronic institutional records? The *Census* provides some insight into these and other questions concerning a college’s or university’s archives as its institutional repository.

We will discuss three issues regarding institutional repositories that most pertain to academic archives: 1) the role of archivists in the development of IRs, 2) content recruitment of archival and manuscript materials for IRs, and 3) preservation of materials in the IR. These issues are important because they have an impact on the functions of the archivist, the ability of the archives to carry out its own collection development program, who takes the lead, and who possesses the expertise in preserving digital materials on campuses.

### Literature Review

Institutional repositories are the latest development in a series of systems aimed at managing digital content. At many colleges and universities, though, institutional repositories operate alongside digital libraries, content management systems, and digital asset management systems as part of the digital information management and provision universe. Each of these systems has multiple definitions and implementations. Christine Borgman analyzed different definitions of *digital library* and concludes that, broadly speaking, it is 1) a service; 2) an architecture; 3) a set of information resources: databases of text, numbers, graphics, sound, video, and so on; and 4) a set of tools and capabilities to locate, retrieve, and utilize the information resources available.<sup>4</sup> These information resources include both primary and secondary sources. Content management systems, enterprise content management systems, and digital asset management systems are also recordkeeping systems used to track and store institutional information.<sup>5</sup> IRs overlap with all of these types of information systems. They can focus on the library and

<sup>3</sup> Clifford A. Lynch, “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age,” *ARL Bimonthly Report* 226 (2003): 1–7, available at <http://www.arl.org/bm~doc/br226.pdf>, accessed 1 August 2008.

<sup>4</sup> Christine L. Borgman, “What Are Digital Libraries? Competing Visions,” *Information Processing and Management*, 35, no. 3 (1999): 233.

<sup>5</sup> Bryant Duhon, “Enterprise Content Management: What Is It? Why Should You Care?,” *AIIM E-Doc Magazine* 17 (November/December 2003): 10.

special collections materials and look very much like digital libraries, or they can resemble content and digital asset management systems by collecting materials created by members of the institutional community. Furthermore, a considerable amount has been written on institutional repositories.<sup>6</sup> We will briefly review what we know about three aspects of IR development and implementation: the role of the archivist, archival content in IRs, and IRs and preservation. These three topics bear on the findings we present later in this article.

Lynch's vision of an effective IR includes "collaboration among librarians, information technologists, archives and records managers, faculty, and university administrators and policymakers."<sup>7</sup> In reality, researchers agree that librarians are or should be taking the lead.<sup>8</sup> While Charles Bailey and his colleagues also find that librarians are the "driving force" behind IR efforts, they note that following IR implementation, libraries usually collaborate with other departments such as technical services, archives, and/or cataloging.<sup>9</sup> Archivists' roles vis-à-vis IRs are less clear and can border on competition. Raym Crow observes this phenomenon:

Depending on the university, an institutional repository may complement or compete with the role served by the university archives. University archives often serve two purposes: 1) to manage administrative records to satisfy legally mandated retention requirements, and 2) to preserve materials pertaining to the institution's history and to the activities and achievements of its officers, faculty, staff, students, and alumni. Compared to institutional repositories, which aim to preserve the entire intellectual output of the institution, university archivists exercise broad discretion in determining which papers and other digital objects to collect and store. Still, the potential overlap of roles of the two repository types merits consideration at institutions that support both.<sup>10</sup>

Crow's custodial view of archivists and archives is an older one. In recent years, archivists have become more involved in the research and particularly the teaching mission of colleges and universities. Crow also downplays the difficulties associated with documenting all of the intellectual output of an institution and the benefits of a selection or appraisal process in an IR collection development plan. Douglas Bicknese argues for a more selective approach to faculty

<sup>6</sup> As of June 2008, MIRACLE Project investigators identified nearly 700 citations. See <http://miracle.si.umich.edu/bibliography.html>, accessed 4 June 2008.

<sup>7</sup> Lynch, "Institutional Repositories."

<sup>8</sup> See Eugenio Pelizzari, "Opinions, Research, and Analysis: Harvesting for Disseminating: Open Archives and the Role of Academic Libraries," *The Acquisitions Librarian* 17, nos. 33–34 (2005): 35–51, and Barbara Quint, "Now or Never!," *Searcher* 10 (January 2002): 6.

<sup>9</sup> Charles W. Bailey, *Institutional Repositories*, SPEC Kit 292 (Washington, D.C.: Association of Research Libraries, 2006), 14–15.

<sup>10</sup> Raym Crow, *The Case for Institutional Repositories: A SPARC Position Paper* (2002): 17, available at [http://www.arl.org/sparc/bm-doc/ir\\_final\\_release\\_102.pdf](http://www.arl.org/sparc/bm-doc/ir_final_release_102.pdf), accessed 24 August 2007.

output, suggesting that archivists bring substantial expertise in collecting and appraising faculty papers and other university records that IRs could well utilize.<sup>11</sup>

While Crow and Lynch envision IRs capturing scholarly output and populated by original research papers, preprints, and postprints, research demonstrates that this is not necessarily the case. In fact, in terms of content recruitment, “the majority [of IRs] are clearly in a very early stage of development but even more of the longer-established sites have a relatively small number of documents compared to the research outputs of their institutions.”<sup>12</sup> Content recruitment is also hard to determine because different IRs enumerate digital documents in different ways. Clifford Lynch and Joan Lippincott surveyed members of the Coalition for Networked Information (CNI). Although they refrain from estimating numbers of digital documents in IRs, they find that the types of materials CNI members have in their IRs represent the research, teaching, service, and publicity activities of colleges and universities, and range from text-based items to multimedia artifacts—images, audio, video, software, blogs, and e-portfolios.<sup>13</sup>

Other surveys of IR content are equally vague. Kathleen Shearer’s study of the libraries associated with the Canadian Association of Research Libraries (CARL) finds that most IRs are sparsely populated with an equally wide range of materials. “Regarding content types, in most cases, the CARL institutional repository projects are accepting a wide variety of content types including: journal articles, learning objects, maps, theses and dissertations, electronic journals, photographs, conference proceedings, music scores, data sets, etc.”<sup>14</sup> For an international comparison of IRs jointly sponsored by CNI, UK Joint Information Systems Committee (JISC), and the SURF Foundation in the Netherlands, Gerard van Westrienen and Clifford Lynch surveyed IRs in thirteen nations. They note that IRs typically contain a few hundred digital objects and that the variety appears to be slightly narrower than in the United States and Canada, principally including articles, theses, books, primary data, video, music, and course material. In general, van Westrienen and Lynch find that Western

<sup>11</sup> Douglas Bicknese, “Institutional Repositories and the Institution’s Repository: What Is the Role of the University Archives with an Institution’s On-line Digital Repository?,” *Archival Issues* 28, no. 2 (2003–2004): 81–93.

<sup>12</sup> Mark Ware Consulting Ltd., *Publisher and Library/Learning Solutions (PALS): Pathfinder Research on Web-Based Repositories: Final Report* (January 2004), 33, available at [http://www.palsgroup.org.uk/palsweb/palsweb.nsf/0/8c43ce800a9c67cd80256e370051e88a/\\$FILE/PALS%20report%20on%20Institutional%20Repositories.pdf](http://www.palsgroup.org.uk/palsweb/palsweb.nsf/0/8c43ce800a9c67cd80256e370051e88a/$FILE/PALS%20report%20on%20Institutional%20Repositories.pdf), accessed 24 August 2007. Now behind a firewall, it is still listed on Mark Ware’s homepage, <http://mrkwr.wordpress.com/articles>.

<sup>13</sup> Clifford A. Lynch and Joan K. Lippincott, “Institutional Repository Deployment in the United States as of Early 2005,” *D-Lib Magazine* 11 (September 2005), available at <http://www.dlib.org/dlib/september05/lynch/09lynch.html>, accessed 24 August 2007.

<sup>14</sup> Kathleen Shearer, “Survey Results—Summer 2004: CARL Institutional Repositories Project” (September 2004), 3, available at [http://www.carl-abrc.ca/projects/institutional\\_repositories/pdf/survey\\_results\\_2004-e.pdf](http://www.carl-abrc.ca/projects/institutional_repositories/pdf/survey_results_2004-e.pdf), accessed 24 August 2007.

European IRs are divided between journal articles and theses, while 88% of the content of Australian IRs is considered primary data.<sup>15</sup> We see from these descriptions of content in previous studies, that archival materials are present but on the periphery. Leaving our study aside for the moment, the most recent study in the United States, commissioned by Ithaka, finds that “repositories are most commonly used to hold digital images, followed by special collections.”<sup>16</sup> This points to some real differences between U.S. IRs and those in the rest of the world. Western Europe, and to a lesser extent Australia, are more directly challenging the publishing model both in terms of government and institutional policies and in content recruitment. In the United States, institutional repositories contain a much broader range of materials and often go far beyond faculty output.

Lynch claims that long-term preservation of scholarly content is an essential role of IRs.<sup>17</sup> However, while Lynch and Lippincott also note that the open access argument for IRs has been articulately made and largely accepted for faculty scholarly output, they see preservation as a more persuasive argument for IRs to make as they expand their collecting beyond faculty publications:

There is some reason to believe that, at least for faculty publications, institutional norms (as promulgated by the academic senate) may increasingly encourage faculty to place their writings into institutional repositories. For the other products of e-scholarship and e-research—the datasets, software, simulations, and related materials—we believe that for the foreseeable future the case will still need to be made in terms of continuity, quality and consistency of access, preservation, curation and similar issues.<sup>18</sup>

Digital preservation always comprises a complex series of activities. As noted by Richard Fyffe and his colleagues, long-term preservation consists of specific functions, such as ingesting digital objects and their metadata, storing these objects and retaining the link to their associated metadata, monitoring technological obsolescence, and migrating the objects over time.<sup>19</sup> A report by Mark Ware Consulting argues that “One of the likely largest costs over the long term will be . . . preservation . . . also by far the least known and indeed least

<sup>15</sup> Gerard van Westrienen and Clifford A. Lynch, “Academic Institutional Repositories: Deployment Status in 13 Nations as of Mid 2005,” *D-Lib Magazine* 11 (September 2005), available at <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>, accessed 24 August 2007.

<sup>16</sup> Ithaka, *Ithaka's 2006 Librarian and Faculty Studies: Overview of Key Findings* (New York: Ithaka, 2006), 4–5, available at <http://www.ithaka.org/research/Ithaka.Surveys.2006.Overview.pdf>, accessed 15 August 2007.

<sup>17</sup> Lynch, “Institutional Repositories.”

<sup>18</sup> Lynch and Lippincott, “Institutional Repository Deployment in the United States.”

<sup>19</sup> Richard Fyffe, Deborah Ludwig, Mary Roach, Becky Schulte, and Beth Forrest Warner, *Preservation Planning for Digital Information: Final Report of the HVC2 Digital Preservation Task Force* (Lawrence: University of Kansas), available at <http://hdl.handle.net/1808/166>, accessed 24 August 2007.

knowable . . . [s]o a commitment to an IR amounts to an implicit commitment to an unknown amount of work at some point in the future.”<sup>20</sup>

At the moment, IRs have not become the equivalent of trusted digital repositories with “a mission to provide reliable, long-term access to managed digital resources to its designated community, now and into the future.”<sup>21</sup> Current implementation of digital preservation methods and tools in IRs is in its infancy. The extent to which IRs employ preservation-related standards, such as the Open Archival Information System reference model and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is not known.

Bailey et al. find that preservation is viewed as one of the top three benefits; 36% of respondents to their survey ranked preservation high.<sup>22</sup> IR contributors share the preservation concerns of librarians. Swan and Brown note that 42% of non-open access journal authors are worried about preservation of open access journal contents.<sup>23</sup> From a preservation policy standpoint, identifying the file formats for which IRs will provide long-term access and thus preservation is a crucial decision. When making decisions about preserving file formats, Richard Jones, Theo Andrew, and John MacColl suggest IR staff answer these questions: 1) Is the file format an open standard/format?, 2) Is the file format widely used?, 3) Is the file format and associated technology likely to be preserved?, 4) Are the contents of the file human readable?, and 5) Is the file format itself human readable?”<sup>24</sup> Bailey et al. reported that 74% of the operational IRs in ARL libraries accept any digital file type, adding that “Relatively few (26%) are committed to functional preservation of every file type.”<sup>25</sup>

Discussion surrounding trusted digital repositories has escalated recently with the publication of the *Trusted Repositories Audit and Certification: Criteria and Checklist*.<sup>26</sup> While not always explicit, the preservation-worthiness of IRs contin-

<sup>20</sup> Mark Ware Consulting, *Publisher and Library/Learning Solutions*, 24.

<sup>21</sup> *Trusted Digital Repositories: Attributes and Responsibilities* (Mountain View, Calif.: RLG, 2002), available at <http://www.rlg.org/en/pdfs/repositories.pdf>, accessed 24 August 2007, as quoted in Center for Research Libraries (CRL), *Trustworthy Repositories Audit and Certification: Criteria and Checklist* (Chicago: CRL, 2007), 3, available at <http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91>, accessed 1 February 2008.

<sup>22</sup> Bailey et al., *Institutional Repositories*, 87.

<sup>23</sup> Alma Swan and Sheridan Brown, “Authors and Open Access Publishing,” *Learned Publishing* 17, (July 2004): 219–24, and Alma P. Swan and Sheridan N. Brown, *JISC/OSI Journal Authors Survey Report* (Truro, U.K.: Key Perspectives Ltd., February 2004), available at [http://www.jisc.ac.uk/uploaded\\_documents/JISCOAreport1.pdf](http://www.jisc.ac.uk/uploaded_documents/JISCOAreport1.pdf), accessed 24 August 2007.

<sup>24</sup> Richard Jones, Theo Andrew, and John MacColl, *The Institutional Repository* (Oxford, U.K.: Chandos Publishing, 2006), 80.

<sup>25</sup> Bailey et al., *Institutional Repositories*, 17.

<sup>26</sup> Center for Research Libraries, *Trustworthy Repositories Audit and Certification: Criteria and Checklist*.

ues to be an issue. At least one article by MacKenzie Smith and Reagan W. Moore discusses how an IR (in this case MIT's DSpace) can be configured both technologically and policywise to be a trusted digital repository (TDR) and how much of the certification work can be automatic. It is also interesting to note that Smith and Moore refer to archivists as being involved in this process, although they finally state that

It is our assumption that archivists who manage digital archives work primarily at the policy level. . . . This mapping from concrete policies established by archivists at the data curation level (e.g. DSpace) into specific system capabilities that are enforced at the data management level (e.g. SRB) will normally be done by technology experts developing the rules engines, rather than the archivists themselves.<sup>27</sup>

Their dismissal of a large number of archivists having the technical knowledge necessary for this work is disappointing and certainly narrows the role archivists could play in preservation in IRs and other digital repositories.

As we see from this review of selected previous research, many questions abound about archivists' roles in the institutional repository, archival content, and digital preservation in IRs. In later sections of this paper, we present findings from the *Census* that complement, contradict, and attempt to provide more in-depth descriptions of these issues.

## Methods

In our research design for the *Census*, we targeted library directors as the most likely individuals to be aware of and involved in IR activities on campuses. We obtained electronic mailing lists with the names and email addresses of academic library directors and senior library administrators at U.S. educational institutions from two sources: 1) American Library Directory (ALD), and 2) an online database produced by Thompson-Peterson, which provided a comprehensive list (2,207 records) of all college and university main libraries in the U.S. Still, ALD's online database did not provide email addresses for specific individuals so we supplemented this with the less comprehensive data from Thompson-Peterson that contained email addresses. After comparing the lists, deleting community colleges and duplicate entries, we identified a total of 2,147 email addresses for the nationwide census.

Developing the questionnaire was more complex. To draft survey instruments, we reviewed published and open access literature on IRs, talked to colleagues, and asked our advisory board to comment on pilot draft instruments. We anticipated

<sup>27</sup> MacKenzie Smith and Reagan W. Moore, *The International Journal of Digital Curation* 1 (June 2007): 95, available at <http://www.ijdc.net/ijdc/article/view/27/30>, accessed 1 February 2008.

that survey respondents would come from colleges and universities at various stages in IR development. Accordingly, we created four instruments corresponding to the stage of IR involvement:

1. no planning to date (NP),
2. planning only to date (PO),
3. both planning and pilot testing one or more IR systems (PPT), and
4. public implementation of an IR system at the respondent's institution (IMP).

The *Census* covered such topics as needs assessment activities, funding, perceived benefits, personnel involved, content, systems, and policies. Most of the response options were scales, offering respondents four primary choices, such as very important, somewhat important, somewhat unimportant, and very unimportant. Respondents could also indicate that they had no opinion or did not know the answer to questions. The full text of the survey was published in the *Census* report and on the project website.<sup>28</sup>

We established a detailed survey administration plan. First, we selected SurveyMonkey to administer the *Census*. After pretesting several approaches, we settled on first sending an email message to each institution's academic library director or a senior administrator to tell them about the *Census* and queried them about the stage of their institutional repository: "Please tell me how you would characterize the current status of your institutional repository (IR)?" Based on their selection of one of the four options (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 at your institution, we replied with a second email message bearing a link to the appropriate web-administered questionnaire. We used SurveyMonkey's List Management Tool to send out initial survey links and to perform two subsequent follow-ups with people who had agreed to participate but who had failed to respond to our inquiries. The *Census* was conducted from 19 April through 25 June 2006.

After closing the *Census* in SurveyMonkey, we exported census data from SurveyMonkey into four Microsoft Excel files (one for each version of the survey—NP, PO, PPT, and IMP). We cleaned up census data, deleting the responses of people who did not agree to sign the informed consent form as well as the surveys of people who responded *only* to the informed consent form and/or to the one question about the number of IRs at their institution. We also deleted empty questionnaires, one entry submitted from a two-year college, and multiple answer sets, keeping only the most comprehensive response sets from respondents. When data clean-up was done, the *Census* response rate was 20.8% (n=446). We exported these data to SPSS, a statistical software package, for analysis.

<sup>28</sup> Markey et al., *Census of Institutional Repositories in the United States*; and MIRACLE Project website.



In a follow-up to the survey, we conducted a series of semistructured interviews to understand some of the phenomena in greater depth. Using findings from the survey, we again created four interview protocols based on how respondents categorized their stage of IR development. One of the final questions on the *Census* asked survey respondents to indicate whether they would be willing to volunteer for a follow-up interview and, if so, to provide their names and email addresses. This generated 176 volunteers. We wanted to conduct approximately forty interviews so we created a purposive sample from these volunteers taking into account IR stage of development (from no planning or only planning to implementation and planning and pilot testing), geographic region, the size and Carnegie classification of parent institution (from small colleges to research universities), the range and types of content represented in the IR, and the position of respondents. In the end, we recruited 36 participants for the phone interviews holding the following positions: 11 library staff, 9 library directors, 4 assistant-associate library directors, 4 archivists or directors of archives, 4 heads or directors in libraries, 3 CIOs, and 1 associate dean for research. These interviewees also hailed from different types of academic institutions: 16 from research universities, 8 from master's colleges and universities, 1 from a doctoral/research university, and 11 from baccalaureate colleges.<sup>29</sup> 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. The hour-and-a-half telephone interviews were conducted from October to December 2006. We developed an initial coding scheme for the interviews and ingested them into N-Vivo 7, a qualitative data analysis application, for coding and analysis. We conducted intercoder reliability testing prior to full coding of these data.

## Findings

As noted, we will concentrate on the three areas most closely related to archives and archivists in the *Census* and the follow-up interviews: 1) the role of archivists in the development of IRs, 2) the types of content recruitment currently being done in IRs, and 3) preservation of materials in IRs. In these data, we heard from library directors and other library and information technology staff, as well as archivists themselves. Taken together, these data provide a number of different perspectives on archivists and archives vis-à-vis institutional repositories.

<sup>29</sup> The Carnegie Foundation for the Advancement of Teaching, *Carnegie Classifications: Lookup and Listings* (2006), available at <http://www.carnegiefoundation.org/classifications/index.asp?key=807>, accessed 15 August 2007.

### The Role of the Archivist

We sent the *Census* to library directors and senior library administrators but we asked them to pass our questionnaire on to staff members who were most involved in IR development. One of the final questions in the *Census* asked respondents to identify their position. Figure 1 shows who completed questionnaires. Almost three-quarters of respondents (n=288) were library directors followed by library staff (10%, n=40) and assistant-associate library directors (8%, n=31). Archivists responded 3% (n=11) of the time.

The number of archival respondents to the *Census* is inverse to the stage of IR development. Four archivists responded in institutions that had done no planning (NP), 4 in planning only (PO), 2 in the planning and pilot testing (PPT), and 1 in the implementation (IMP) phase. The inverse relationship between maturity of the IR and archival respondents is interesting but ultimately uninformative. We completed interviews with 4 archivists. Two of those individuals confused the intent of our survey, thinking that we were surveying archives as institutional repositories, not the institutional repository.

We asked *Census* respondents about the activity level of people involved in their institution's IR: "How active were people in the following positions in terms of leading the charge to get involved with IRs at your institution?" We asked respondents from institutions with no IR planning (NP), "How active do you

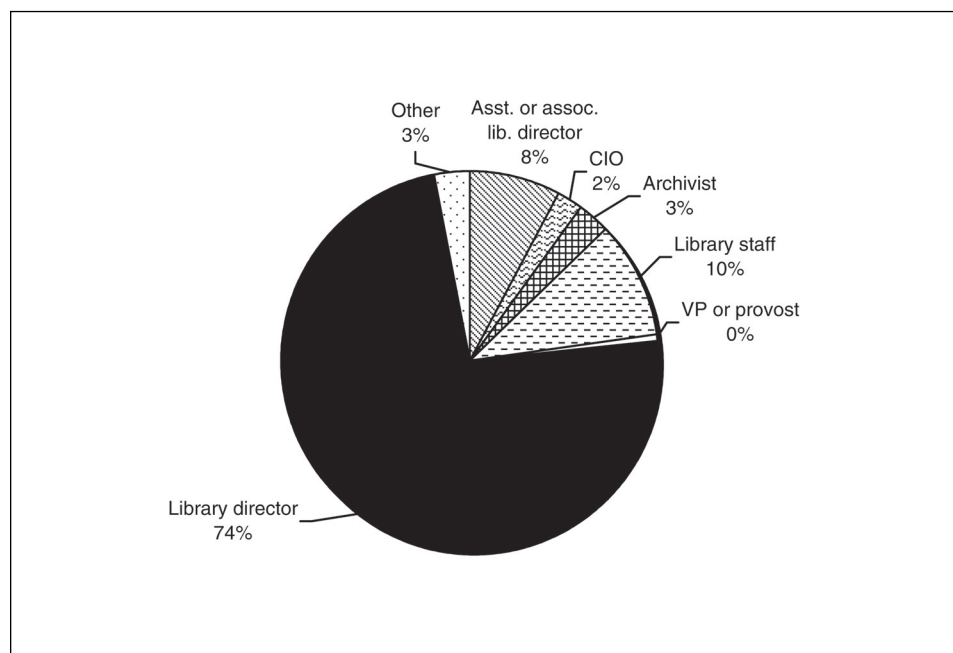


FIGURE 1. Respondents who completed questionnaires.

think that the people in these positions would have to be to light the spark for IR planning at your institution?” To answer this question, respondents rated each position from a list of 13 (see Figure 2); some respondents also wrote in a response for “other.” Respondents were given a 4-point scale ranging from “very active” to “very inactive.” Figure 2 shows that 75% of respondents consider archivists to be engaged in the IR process and overall rank archivists as one of the top five (fourth, in fact) in activity level.

The *Census* confirms the presence of archivists in IR development and implementation. Figure 3 further demonstrates IR committee membership in PO, PPT, and IMP institutions. The archival presence increases as the IR enters the planning and pilot testing stage and nears implementation. Archivists are present 32% of the time in PO, 56% in PPT, and 52% in IMP institutions.

We also asked respondents who was leading the IR effort at their institutions. As expected, few archivists are directing the IR process and none are overseeing implemented IRs. When archivists do head the IR effort, it is more likely to be during the PPT stage in institutions classified as master’s and baccalaureate institutions according to the Carnegie Classification of Higher Education. In these institutions at that stage, archivists lead 6.5% of the IRs.

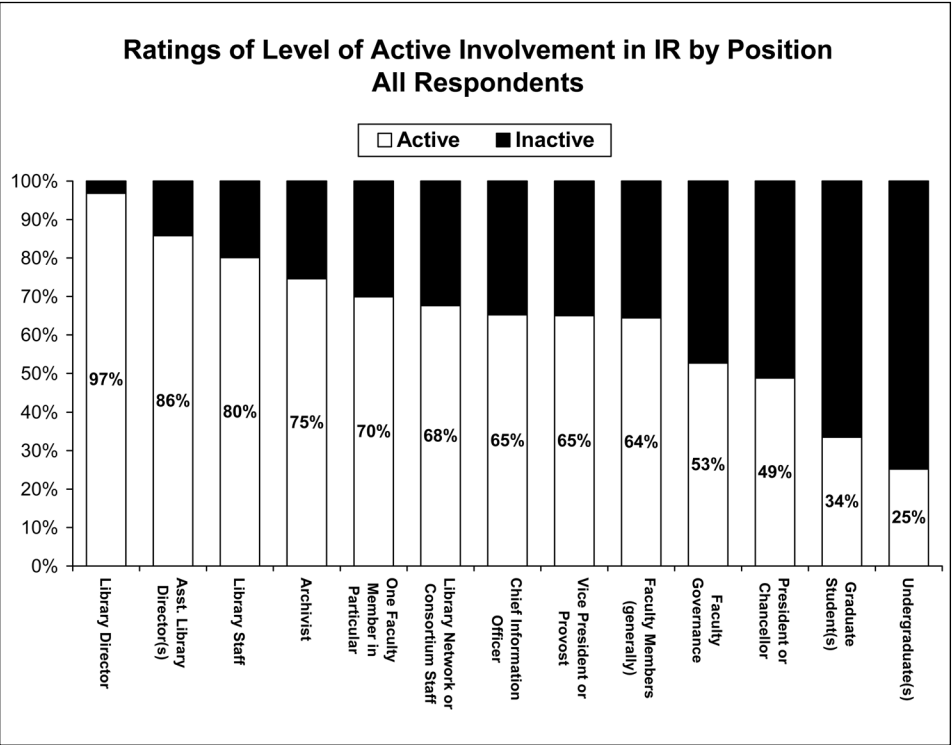


FIGURE 2. All IR census respondents ratings of level of active involvement in IR by position.

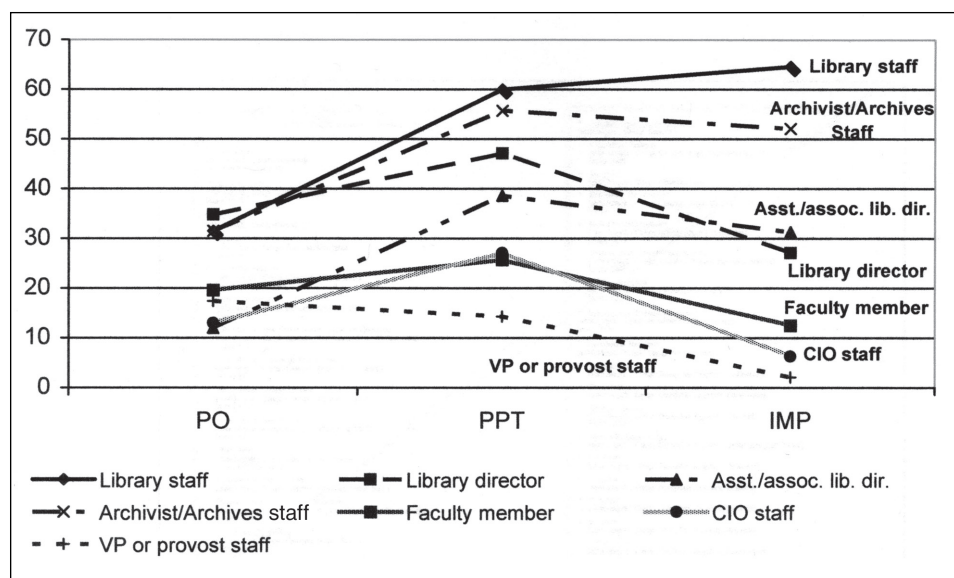


FIGURE 3. IR committee membership.

We asked, “If your efforts to implement an IR involved early adopters of IR technology, from what academic colleges, departments, schools, and service units have they come?” Archivists are also perceived to be early adopters of IR technology. Among all respondents in IMP and PPT institutions, the archives ranks third in terms of early adoption and utilization of the IR. (See Table 1.) About two-fifths of PPT and one-quarter of IMP respondents have worked with their institution’s archives.

As previously noted, only 11 archivists responded to the *Census*, and the number of archivist-respondents to any single question was often much smaller. Given this small sample size, it is difficult to make any generalizations about how archivists’ perceptions differ from those of other respondents. However, in several questions, a majority of the archivists responded differently. Not surprisingly, archivists perceive themselves to be more active; 90.9% of the archivists judge themselves to be “very active” or “somewhat active” ( $n=10$ ) as opposed to 74.6% of all respondents ( $n=226$ ). When we asked archivists, “When planning for an IR, what did you think would be the most important reasons why members of your institution’s learning community would contribute to the IR?,” all 7 who responded agree that it is “very important” or “somewhat important” to place the burden of preservation on the IR rather than on individual faculty members. This contrasts with respondents as a whole; only 76.9% ( $n=133$ ) think that relieving faculty of the preservation burden is a compelling reason for contribution. In the end, the few “archivist” responses provide less evidence about

**Table I.** Early Adopters of IR Technology

Rank	Academic Unit	Number	Percent
1	Your institution's library	70	59.3%
2	A particular academic college, department, or school	67	56.8%
3	Your institution's archives	38	32.2%
4	A particular service unit	12	10.2%
5	Your institution's central computing unit	9	7.6%
6	Other (please specify)	7	5.9%
	Not applicable/Don't know	12	10.2%

the archival role in IRs than the entire pool of responses provides about archivists, archival content, and preservation.

**Content Recruitment in the IR**

Content recruitment is key to a successful IR.<sup>30</sup> We identified 37 digital document types that might be included in institutional repositories and asked respondents what types of materials they recruit. If we take a very broad view of what constitutes university records or archives, 30 of these content types could be considered archival. Institutions that had implemented (IMP) an IR and those in the planning and pilot testing (PPT) stage both rely heavily on archival materials for content. The total number of digital documents in IMP institutions is 76,477, of which 53,780, or 70.3%, could be considered archival. Likewise, respondents count 32,667 digital documents in PPT institutions, of which 23,189, or 71.0%, could be considered archival. (See Appendix A.)

Table 2 shows the *average* number of documents by type in PPT and IMP IRs according to frequency: high (over 200 documents), medium high (51–200 documents), medium low (6–50 documents), and low (less than 6 documents). Dissertations are the most frequent document type in both IMP and PPT institutions. After that, the average digital document frequencies ratings do not match. Furthermore, it should be noted that even though the averages for some document types are high, the actual number of respondents can be low and is given in parentheses. Of note, there are low numbers of all document types in IRs.

<sup>30</sup> See, for example, Nancy Fried Foster and Susan Gibbons, "Understanding Faculty to Improve Content Recruitment for Institutional Repositories," *D-Lib Magazine* 11 (January 2005), available at <http://www.dlib.org/dlib/january05/foster/01foster.html>, accessed 6 September 2007, and Barbara Jenkins, Elizabeth Breakstone, and Carol Hixson, "Content In, Content Out: The Dual Roles of the Reference Librarian in Institutional Repositories," *Reference Services Review* 33, no. 3 (2005): 312–24.

**Table 2.** Document Types in Pilot-Test and Operational IRs

<b>PPT high: Over 200 documents</b>	Average number	<b>IMP high: Over 200 documents</b>	Average number
<i>Doctoral dissertations (n=9)</i>	1,288	<i>Doctoral dissertations (n=18)</i>	1,518
<i>Preprints (n=7)</i>	900	<i>Working papers (n=18)</i>	716
<i>Other learning objects . . . prepared by faculty, lecturers, teaching assistants, etc. (n=4)</i>	550	<i>Journal articles (n=19)</i>	462
<i>Master's theses (n=10)</i>	230	<i>Raw data files that result from doctoral dissertation research (n=11)</i>	457
		<i>Master's theses (n=16)</i>	419
<b>PPT medium high: 51 to 200 documents</b>	Average number	<b>IMP medium high: 51 to 200 documents</b>	Average number
<i>Journal articles (n=14)</i>	172	<i>Committee meeting agendas and minutes (n=8)</i>	90
<i>Working papers (n=18)</i>	124	<i>Preprints (n=10)</i>	84
<i>Your institution's course catalogs (n=7)</i>	109	<i>Your institution's newspapers (n=7)</i>	81
<i>Books (n=4)</i>	96	<i>Senior theses (n=12)</i>	68
<i>Video recordings of performances (n=6)</i>	76	<i>Committee meeting documents, e.g., budgets, reports, memos (n=8)</i>	68
<i>Senior theses (n=7)</i>	68	<i>Maps (n=9)</i>	61
<b>PPT medium low: 6 to 50 documents</b>	Average number	<b>IMP medium low: 6 to 50 documents</b>	Average number
<i>Faculty senate agendas and minutes (n=5)</i>	51	<i>Other learning objects . . . prepared by faculty, lecturers, teaching assistants, etc. (n=9)</i>	31
<i>Interview transcripts (n=6)</i>	49	<i>Written papers or transcripts of conference presentations (n=12)</i>	27
<i>Your institution's alumni publications (n=4)</i>	43	<i>Undergraduates' class notes, outlines, assignments, papers, and projects (n=10)</i>	18
<i>Sound recordings of interview transcripts (n=7)</i>	36	<i>Conference presentations (e.g., summaries, abstracts, notes, outlines, etc.) (n=10)</i>	16
<i>Your institution's newspapers (n=2)</i>	34	<i>Interim and final reports to funders (n=9)</i>	13
<i>Course syllabi, class notes, handouts, etc., prepared by faculty, lecturers, etc. (n=3)</i>	33	<i>College, departmental, and school alumni publications (n=9)</i>	13
<i>Regent, trustee, board meeting agendas and minutes (n=4)</i>	30	<i>Faculty senate agendas and minutes (n=8)</i>	13
<i>Committee meeting documents, e.g., budgets, reports, memos (n=5)</i>	28	<i>Regent, trustee, board meeting agendas and minutes (n=9)</i>	11
<i>Committee meeting agendas and minutes (n=5)</i>	28	<i>Video recordings of performances (n=12)</i>	9
<i>College, departmental, and school alumni publications (n=4)</i>	22	<i>Books (n=15)</i>	6
<i>Journals (n=16)</i>	22	<i>Sound recordings of interview transcripts (n=11)</i>	6
<i>Written papers or transcripts of conference presentations (n=3)</i>	17		
<i>Conference presentations (e.g., summaries, abstracts, notes, outlines, etc.) (n=4)</i>	8		
<i>Undergraduate student e-portfolios (n=2)</i>	8		

**Table 2.** Continued

<b>PPT low: 5 or fewer documents</b>	<b>Average number</b>	<b>IMP low: 5 or fewer documents</b>	<b>Average number</b>
<i>Maps (n=3)</i>	5	<i>Journals (n=6)</i>	4
<i>Raw data files from senior thesis research (n=3)</i>	5	<i>Interview transcripts (n=10)</i>	3
<i>Raw data files from faculty research projects (n=5)</i>	3	<i>Raw data files that result from master's thesis research (n=9)</i>	2
<i>Software (n=2)</i>	3	<i>Software (n=9)</i>	2
<i>Undergraduates' class notes, outlines, assignments, papers, and projects (n=2)</i>	0	<i>Course syllabi, class notes, handouts, etc., prepared by faculty, lecturers, etc. (n=9)</i>	2
<i>Raw data files that result from doctoral dissertation research (n=1)</i>	0	<i>Software documentation (n=9)</i>	2
<i>Interim and final reports to funders (n=1)</i>	0	<i>Raw data files from faculty research projects (n=7)</i>	1
<i>Raw data files that result from master's thesis research (n=1)</i>	0	<i>Your institution's course catalogs (n=7)</i>	1
<i>Software documentation (n=1)</i>	0	<i>Raw data files from senior thesis research (n=8)</i>	1
<i>Graduate students' class notes, outlines, assignments, papers, and projects (n=1)</i>	0	<i>Graduate students' class notes, outlines, assignments, papers, and projects (n=8)</i>	1
<i>Graduate student e-portfolios (n=2)</i>	0	<i>Your institution's alumni publications (n=9)</i>	0
<i>Blogs (n=1)</i>	0	<i>Undergraduate student e-portfolios (n=2)</i>	0
		<i>Graduate student e-portfolios (n=7)</i>	0
		<i>Blogs (n=8)</i>	0

Particularly interesting about this list is that in the medium-high and medium-low categories, document types associated with institutional records and faculty papers begin to be listed. These manuscripts in paper form traditionally collected by university archives are now being targeted by IR staff with or without the cooperation/consent of the university archives and special collections. In addition to the document types we listed, *Census* respondents volunteered additional document types in the "Other" category, such as government documents, archives, institutional historical documents including photographs, art history slide collections, faculty spatial data sets, staff project reports, research reports from centers and institutes, self-study reports and other documentation from academic accreditation events, posters, newsletters, musical scores, and scrapbooks. Many of these materials could also be considered archival.

We grouped the archival content types into four categories: faculty papers, student papers, special collections, and university records. The faculty papers are research data, curricular materials, and other types of personal papers. Special collections material consist largely of groups of images. Student papers include student work, such as e-portfolios and class assignments, but 90% of the documents in this category are theses and dissertations. University records are items generated by university offices. Table 3 shows the distribution of IR archival content across these categories. We recognize that the inclusion of theses and dissertations is problematic because they are "low-hanging fruit" for IRs; however,

**Table 3.** Types of Archival Content in IRs

Archival Digital Document Category	PPT		IMP		PPT + IMP	
	#	%	#	%	#	%
Faculty papers	2,931	12.6%	1,009	1.9%	3,940	5.1%
Special collections	15	0.1%	550	1.0%	565	0.7%
Student papers	14,402	62.1%	40,220	74.8%	54,622	71.0%
University records	2,270	9.8%	2,262	4.2%	4,532	5.9%
Other	3,571	15.4%	9,739	18.1%	13,310	17.3%
Sum	23,189	100.0%	53,780	100.0%	76,969	100.0%

these are important “archival” records documenting teaching and research. Theses and dissertations (along with other student work) account for 45% of all documents in all IRs. If we include dissertations and theses as archival material, archival material makes up 70.6% of all digital documents in IRs. If we omit dissertations and similar materials, archival materials make up 20.5% of IR content.

Figure 4 shows that even without the inclusion of theses and dissertations, 15 IRs still contain between 90% and 100% archival content. Out of 66 operational IRs, 26 have over 50% archival content.

Archival content is not the only archival contribution to IR content recruitment. Interviewees described repeatedly how they use existing liaisons and subject bibliographers to meet with professors to solicit content (Interviewees IMP 7, 8, 12, 16, 17, 18). This new role for the subject specialists is not always an easy fit. Interviewee PPT 4 noted that

In addition to the usual kinds of liaison work that subject specialists do, we’re sort of driving this initiative where we re-educate our librarians to think about asking broader questions. Not just what teaching are you doing and how can we support that by having things on reserve, but what research are you doing? And it’s really when you go and talk to people about do they have data or information management, organization access needs, they start talking to you from their perspective and say you know what? I’m not able to get a hold of this thing.

Interviewee IMP 18 discussed the difficulty of getting the bibliographers to do this type of content recruitment.

### **Preservation**

The *Census* addressed preservation issues. Initially, we asked respondents whether they dealt with preservation issues during the IR needs assessment process and whether they perceive preservation as a benefit of the IR. Later, we queried them about their actual preservation activities, such as policies about preservation of file formats. Finally, we asked respondents about migration of



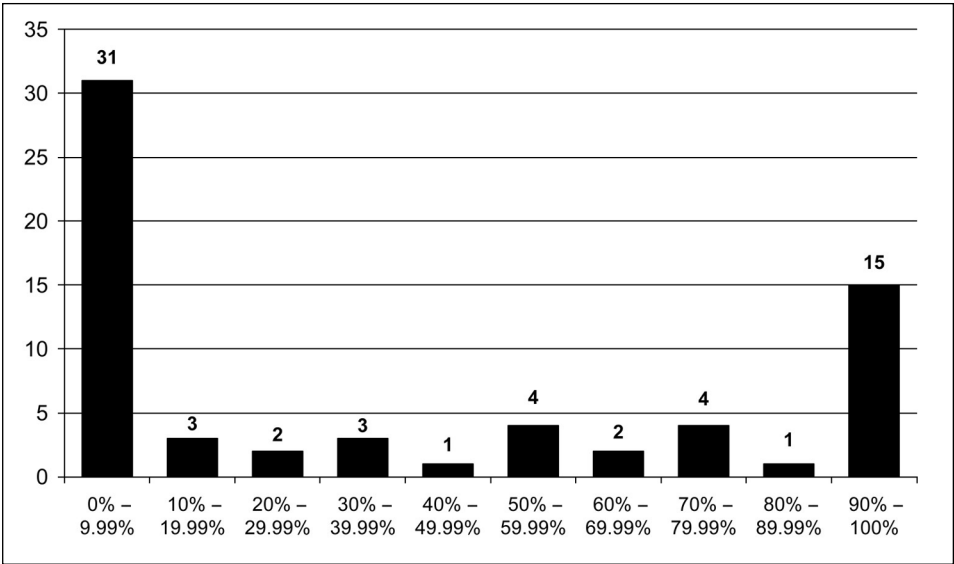


FIGURE 4. Percentage of archival content in IRs excluding theses and dissertations.

the IR system itself. The responses to these varied questions demonstrate a great deal of uncertainty about preservation in IRs.

Needs assessment prior to launching an IR is not universal; few IMP institutions and even fewer PPT institutions are doing this. Among those few IMP institutions that do perform some form of needs assessment, we asked how important this activity is for accomplishing 11 IR-related tasks, among them the identification of preservation techniques. We also asked whether the result of that investigation influenced the decision to proceed with the IR. Respondents rated preservation as the least important IR-related task with regard to the needs assessment. (See Table 4.)

We anticipated that preservation would be a perceived benefit of an IR and asked several questions in this regard. In asking them, we attempted to assess changes in attitudes from the beginning of the IR planning process through pilot testing and implementation. We first asked, “At the beginning of IR planning at your institution, how important did you think these anticipated benefits of IRs would be to your institution?” Respondents were asked to rate 16 potential benefits in terms of importance, two of which concerned preservation. PPT and IMP institutions ranked the preservation options high: “longtime preservation of your institution’s digital output” as the fifth (PPT) and fourth (IMP) most important benefit of an IR, and “a solution to the problem of preserving your institution’s intellectual output” as sixth (PPT) and fourth (IMP).

As a corollary to this question, we later asked IMP respondents about benefits a second time: “Now that you are implementing or have implemented an

**Table 4.** Percent Rating Needs Assessment as Very Important or Somewhat Important for Each Task

Rank	IR-related tasks	Percentage
1	Formulating IR policies	90%
2	Identifying first adopters of an IR	84%
3	Recruiting digital content for the IR	83%
4	Choosing an IR software package	82%
4	Streamlining IR planning and implementation	82%
6	Increasing faculty awareness of the IR	79%
7	Identifying especially active contributors to the IR	78%
8	Identifying new services to build onto the IR	72%
9	Scheduling the rollout of various IR services	69%
10	Making the decision to implement an IR	68%
11	Identifying preservation techniques	63%

IR, reassess these same anticipated benefits of IRs and tell whether you think they are less important or more important than you originally thought.” Approximately three-quarters of respondents made no changes in the importance of listed benefits. When they did revise their initial ratings, the change was nearly always an *increase* in importance. Both of the preservation options increased in importance over 30% of the time. (See Table 5.)

In a final contribution- and benefits-related question, we asked respondents to rate their perceptions of what other members of their learning community valued in an IR. For IMP institutions, the question was “When planning for an IR, what did you think would be the most important reasons why members of your institution’s learning community would contribute to the IR?” For PPT and OP, the question was “Why do you think members of your institution’s learning community will contribute to an IR?” Respondents were again asked to rate 15 reasons according to the following categories: “very important,” “somewhat important,” “somewhat unimportant,” “very unimportant,” “no opinion,” “don’t know,” and “not applicable.” Two of the options concerned preservation: to solve the problem of preserving the institution’s intellectual output and to place the burden of preservation on the IR instead of on the individual faculty members. Table 6 shows that respondents perceive that preservation issues are a key benefit for potential contributors and users of the IR. Over 75% of the respondents said these are “very important” or “somewhat important” to members of their institution’s learning community.

Preservation systems and policies lag behind perceived importance of preservation issues. Preservation considerations are in play in the IR system selection process, and preservation issues are viewed as a component of success. We asked, “To what extent do you think the following are likely to inhibit your ability to deploy a successful IR?” and provided 13 possibilities for respondents to rate. Table 7 shows that approximately 35% of the respondents perceive that it is “very

Table 5. Increases in Benefits' Importance

Rank	Benefit	% Increase
1	An increase in your library's role as a viable partner in the research enterprise	50.0%
2	An increase in the accessibility to knowledge assets such as numeric, video, audio, and multimedia datasets	36.8%
2	Longtime preservation of your institution's digital output	36.8%
4	Better service to contributors	35.9%
5	Better services to your institution's learning community	35.0%
6	New services to learning communities beyond your institution	34.2%
7	A solution to the problem of preserving your institution's intellectual output	33.3%
7	Exposing your institution's intellectual output to researchers around the world who would not otherwise have access to it through traditional channels	33.3%
9	A boost to your institution's prestige	32.5%
10	Capturing the intellectual capital of your institution	30.8%
10	Maintaining control over your institution's intellectual property	30.8%

Table 6. Perceived Preservation Benefits of IRs

	Very Important	Somewhat Important	Total
To solve the problem of preserving your institution's intellectual output	31.6%	45.2%	76.8%
To place the burden of preservation on the IR instead of on the individual faculty members	34.7%	42.2%	76.9%

likely” or “somewhat likely” that “inadequacy of the IR system’s digital preservation capabilities” would negatively impact the successful adoption of an IR.

Given that most respondents view preservation as a benefit, we asked those who are in the PPT and IMP stages to rate their actual systems against a list of 14 attributes: technical support, technical documentation, adherence to open access standards, scalability, customization, extensibility, supported file formats, authentication, formulating metadata, browsing and searching, end-user interface, controlled vocabulary, authority control, and digital preservation. 78.2% of respondents perceive that their system is “somewhat adequate” or “very adequate” for digital preservation (see Table 8); however, less than 30% rated their system as “very adequate” in this regard. Again, a high number (approximately 20) of respondents skipped this question, and an approximately 20 individuals answered “no opinion,” “don’t know,” or “not applicable.”

Some of the IMP respondents are already planning for system migration. When asked, “How long do you think your institution will stick to this IR system before migrating to a new system?,” 14 individuals answered in “3 or fewer years” and 10 in “4–6 years.” Only 1 person envisions sticking with his or her current IR system for over 7 years. This means that, on average, the implementers see changing systems in 3.4 years. Again, a large number of respondents (23) skipped this question.

**Table 7.** Barriers to Successful Deployment of an IR System

Inadequacy of the IR System's Digital Preservation Capabilities		
	Number	Percentage
Very Likely	14	9.6%
Somewhat Likely	37	25.3%
Somewhat Unlikely	69	47.3%
Very Unlikely	26	17.8%
No Opinion	6	
Don't Know	17	
N/A	2	
Sum	171	
Skipped	39	

**Table 8.** Rating of IR Systems and Preservation

Digital preservation	#	%
Very Adequate	23	29.5%
Somewhat Adequate	38	48.7%
Somewhat Inadequate	13	16.7%
Very Inadequate	4	5.1%
No Opinion/Don't Know/Not Applicable	20	
Total	98	

We also queried respondents about their impetuses for migrating: "What do you think will be the most important reasons for migrating to a new IR system?" More than 95% of the respondents identify preservation as "very important" or "somewhat important" in selecting the next system. Table 9 shows that this ranks first among all of the possible reasons for migrating.

Preservation policies are also more in flux. We asked PO, PPT, and IMP respondents to characterize the status of 16 different types of policies as follows: "no policy," "drafted policy," "implemented policy," "don't know," and "not applicable." Many of these policies are directly related to preservation issues. Figure 5 shows the status of these policies at IMP institutions. Interestingly, 16 IMP respondents skipped the policy question altogether, and another 3 chose either the "do not know" or "not applicable" categories. As a result, only 60% of the total 48 IMP respondents answered this question making generalizations difficult. IMP respondents report the highest number of implemented policies for 1) acceptable file formats (73.3%); 2) determining who is authorized to make contributions to the IR (68.6%); 3) defining collections (63.6%); 4) restricting access to IR content (61.3%); 5) identifying metadata formats and authorized metadata creators (61.3%); and 6) determining what is acceptable content (60.6%). Most IMP institutions have implemented or drafted all policies

**Table 9.** Potential Reasons for Migrating to a New IR System

Rank	Reason	% Important
1	Greater capacity for handling preservation	96.6%
2	Greater opportunities for customization	89.7%
3	Advanced searching features	86.2%
3	Greater versatility with the wide range of digital formats	86.2%
5	Friendlier digital content submission procedure	85.2%
6	Friendlier user interface	82.8%
7	Better tools for assisting contributors with metadata creation	82.1%
8	Greater versatility for linking to other campus systems and data	75.9%
9	Around-the-clock technical support	44.4%

except: 1) charging for IR services (16.7%); 2) licensing IR content (29.2%); and 3) providing access management services (37.1%).

Since only 60% of the 48 IMP respondents answered this question, we assume that some operating IRs are still developing policies. The policies of greatest interest to archivists are the preservation policy concerning acceptable file formats and preserving content in general. Guaranteeing file formats in perpetuity poses problems for both PPT and IMP institutions. The PPT respondents were less likely than IMPs to indicate file formats guaranteed in perpetuity. For nearly any given file format, more than 65% of the PPT respondents either skipped the question or selected “do not know,” “no opinion,” or “not applicable.” In response to a question asking them to explain these answers, one respondent noted, “We don’t guarantee anything. At least not until we have a preservation plan in production.”

Colleges and universities with operational IRs are slightly more definitive. Table 10 lists the digital file formats that at least 12.5% of IMP respondents guarantee in perpetuity and those IMP institutions that skipped, did not know, or who had no opinion. These latter numbers are striking. Other than PDF format, approximately 45% of the IMP respondents skipped this question. Most of the IMP institutions claim to preserve text, numerical, or image data in some way; however, the numbers for audio and moving image data are abysmal. Fewer than 12.5% of IMP respondents guarantee multimedia formats such as QuickTime (10.4%), MPEG-4 (10.4%), Windows Media Video (6.3%), and AVI (6.3%); sound formats such as AIFF (8.3%), Real Audio (6.3%), and Wave (6.3%); and image formats such as BMP (10.4%) and PHOTO CD (6.3%).

Initially, we hypothesized that the greater the archival involvement in IRs, the greater the role or importance of preservation. This does not prove to be the case. IRs where archivists are active show no difference in their perception of IR benefits, the importance of preservation features in systems, and the number or types of file formats identified for long-term preservation from those where archivists are inactive. This indicates that archivists may not be viewed as digital preservation experts in the process of IR development or that this

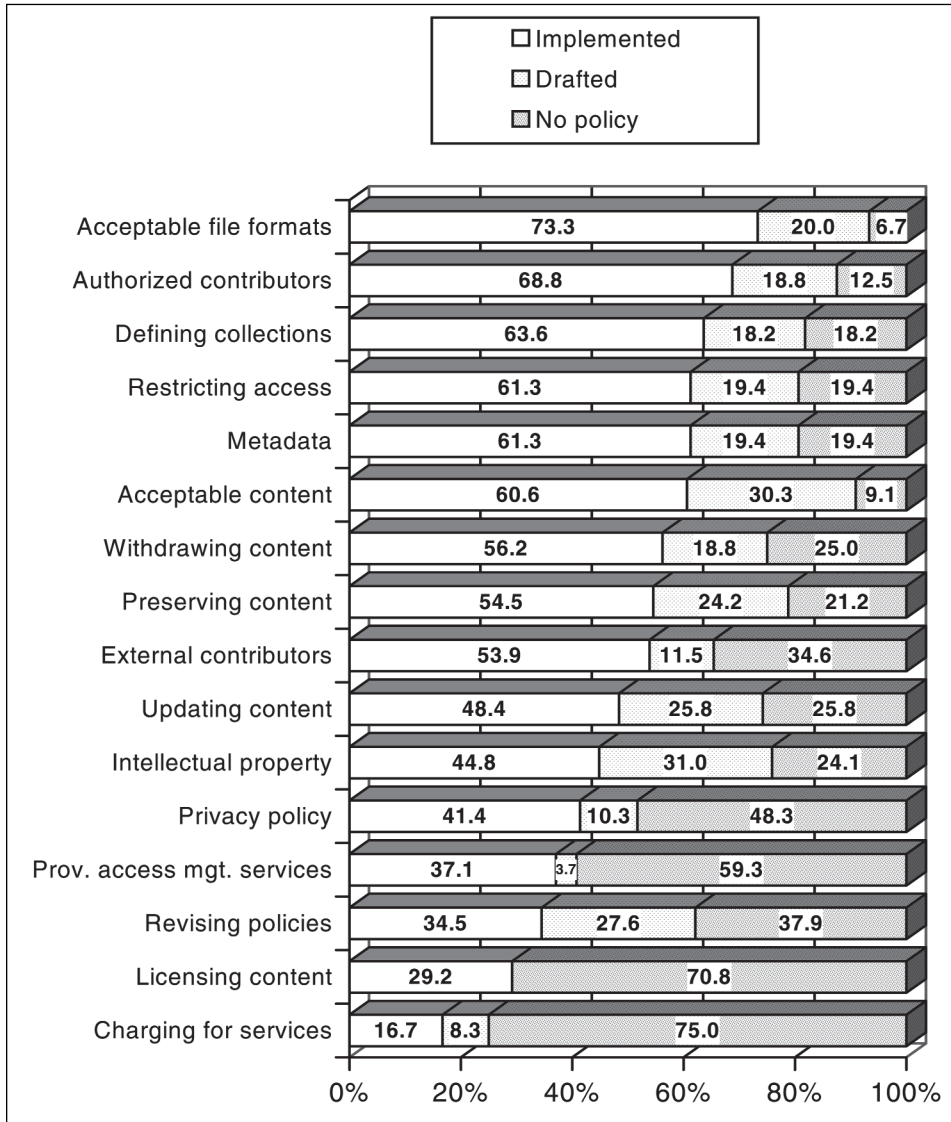


FIGURE 5. Status of IR policies at IMP institutions.

expertise, if it exists, is elsewhere in the library or the university. Archivists may not have the digital preservation skills to step up and take charge of digital preservation activities in the IR.<sup>31</sup>

The interviewees revealed a range of approaches, expertise, and activities concerning preservation. Only one interviewee (IMP 17) claimed to have “no

<sup>31</sup> Susan E. Davis, “Electronic Records Planning in ‘Collecting’ Repositories,” *American Archivist* 71 (Spring/Summer 2008): 167–89.

**Table 10.** Guaranteed Digital File Formats in IMP Institutions

File Format	Skipped	"Do not know", "no opinion," and "not applicable"	Guaranteed by IMP respondents	
	No.	No.	No.	%
PDF	15	6	27	56.3
JPEG	24	6	18	37.5
TIFF	22	8	18	37.5
GIF	24	10	14	29.2
XML	24	11	13	27.1
Microsoft Word	25	10	13	27.1
Microsoft Excel	26	11	11	22.9
PDF/A	22	15	11	22.9
Rich text	26	12	10	20.8
Microsoft PowerPoint	26	12	10	20.8
Postscript	24	15	9	18.8
MPEG audio	26	14	8	16.7
Plain text ANSI X3.4/ECMA-6/US-ASCII (7-bit)	26	15	7	14.8
Plain text UTF-8 (Unicode)	25	16	7	14.8
Plain text ISO 8859-x (8-bit)	26	16	6	12.5
Plain text (all other encodings, including, but not limited to ISO 646, national variants)	26	16	6	12.5
PNG	25	17	6	12.5
TeX	25	17	6	12.5

preservation program at all." Several interviewees are highly sophisticated about digital preservation and are weighing different approaches, such as the LOCKSS model or the DAITSS system. Some interviewees consider back-ups as digital preservation (e.g., Interviewees PPT 8 and IMP 6), while others admit that regular back-ups are all they can do, recognizing that this is not digital preservation (Interviewees IMP 12, 17, and 21). Interviewee IMP 12 summed up the reality and uncertainty regarding digital preservation: "As long as you have a coherent back up plan I think we'll have it. We don't see this as a preservation strategy because it's ... digital preservation is still very much an unknown."

One final preservation note concerns trusted digital repositories. When we did the interviews, the certification guidelines were still in draft form.<sup>32</sup> Still, 26 of the interviewees responded to our question on trusted digital repositories. Seventeen had never heard of them, but 9 were not only aware of trusted digital repositories but were interested in at least benchmarking to the guidelines if not pursuing certification. Interviewee IMP 16 characterized this interest: "We're developing our way forward to move into that status. The repository's working group is studying that and what it takes to get there."

<sup>32</sup> Center for Research Libraries, *Trustworthy Repositories Audit and Certification: Criteria and Checklist*.

## Conclusions

Institutional repositories are a growing presence on college and university campuses, and IRs are increasingly part of the institutional landscape. We began this article with a question concerning the archival implications of the *Census*. We found a real and persistent archival presence in IR development, both in terms of archivists working as part of an IR development team and in terms of content recruitment. However, in most IRs, archivists do not act as nor are they recognized as the digital preservation experts. The presence of archivists in IR planning does not appear to have any impact on the treatment or perceived importance of preservation issues; however, archival expertise would have a substantial impact in this area. This raises key questions about the role of college and university archivists. Does other digital preservation expertise exist at colleges and universities, or are college and university archivists perceived to be more paper or analog based? In a somewhat deprecating remark, one interviewee noted: "I think the real issue for a [institutional] repository is going to be how facile they are at projects that are a mix of video, audio, text, and graphics. And those are the projects that can't sit on a shelf in special collections and I think those are going to be the first candidates for a repository so that we can go back and view them or experience them later" (Interviewee PPT 6). Given the centrality of archival and special collections materials to IRs and the perceived importance of preservation to both IR developers and contributors, this may be a missed opportunity for archivists. Even if IRs are a passing trend, managing digital content is a major issue in colleges and universities, and expertise is needed in this area. If archivists do not fill this gap, others will.

Archivists sit on over 50% of the IR committees at both the IMP and PPT institutions. They even lead the effort at several small liberal arts colleges. When asked if archivists' presence is felt, 75% of respondents rated them as either "very active" or "somewhat active." Through our examination of these data, it appears that archivists are having an impact on IRs, though more in the area of content recruitment than preservation. Although the early IR literature focused on recruiting faculty preprints and postprints to challenge the current publishing paradigm,<sup>33</sup> IR developers have had difficulties recruiting these faculty materials and have sought out other types of content. This has led them to archives in search of potential content for the IR. Overall, IRs are still poorly populated, but archival and manuscript materials appear to be a mainstay in many. In fact, some IRs contain only archival content. Apparently, Institutional Repositories (IRs) are becoming an extension of the institutional repository (archives).

<sup>33</sup> Crow, *The Case for Institutional Repositories*.



Yet, much of this “archival” content is in the form of digitized materials from existing collections, rather than digitally born materials from administrative offices or research projects. This pattern aligns more with archivists’ traditional roles of curator and collector, and reinforces current modes of collection development. If archivists want to document colleges and universities in the future, and some would argue currently, collecting digitally born materials is mandatory.

In addition to actual content, librarians in IRs have borrowed content recruitment methods from archivists. Library subject bibliographers, the major content recruiters for IRs, are taking a page from archivists and doing active fieldwork with faculty. In the past, librarians interacted with faculty to solicit information about what resources faculty wanted the library to purchase or subscribe to in order to support teaching or research. The social dynamic is very different for librarians who now have to ask faculty for materials. This use of archival collection development techniques is noteworthy, and IRs can definitely benefit from archival expertise in this area.

One additional note concerning IRs is that in spite of the archival influence in planning and development, IRs, in general, are not very compatible with archival records; most cannot provide hierarchical display of data and require item-level description. So the IR becomes a place to publish digital images and files decontextualized from the rest of the collection. However, the current generation of IR systems is only a beginning, and if archivists do envision these applications as a means of displaying digital archival records, they are also going to have to become more articulate in presenting these issues during system selection.

College and university libraries have been under siege for the past decade. Decreasing budgets, rising costs for materials in analog and digital form, overdue infrastructure upgrades, and new expectations about the role of the library in higher education have left them reeling. While archives face budgetary concerns and increased pressure to demonstrate their relevance to the larger teaching, learning, and research mission of the university, their core purposes and role are not challenged to the same degree. The development of IRs presents a new set of challenges—competition for digital content and conundrums concerning digital preservation. How will the role of archivists vis-a-vis IRs evolve? Of particular importance is whether archivists can position themselves as digital preservation experts who will fill this role, not only in the IR but in academic institutions as a whole.

### Appendix A: Number of Digital Documents by Type in Institutional Repositories

Document Type	Classification	Archival (A)	PPT	IMP	Sum
Conference presentations (e.g., summaries, abstracts, notes, outlines, remarks, etc.)	Faculty papers	A	31	161	192
Course syllabi, class notes, handouts, outlines, assignments prepared by faculty, lecturers, teaching assistants, and other professional teaching personnel	Faculty papers	A	100	16	116
Interim and final reports to funding agencies	Faculty papers	A	0	115	115
Interview transcripts	Faculty papers	A	290	33	323
Other learning objects such as simulations, models, software demonstration files, images, video prepared by faculty, lecturers, teaching assistants, and other professional teaching personnel	Faculty papers	A	2,200	279	2,479
Raw data files that result from faculty research projects	Faculty papers	A	13	10	23
Sound recordings of interview transcripts	Faculty papers	A	251	66	317
Written papers or transcripts of conference presentations	Faculty papers	A	46	329	375
Maps	Special collections	A	15	550	565
Class notes, outlines, assignments, papers, and projects prepared by graduate students	Student papers	A	0	6	6
Class notes, outlines, assignments, papers, and projects prepared by undergraduate students	Student papers	A	0	176	176
Doctoral dissertations	Student papers	A	11,593	27,329	38,922
Graduate student e-portfolios	Student papers	A	0	0	0
Master's theses	Student papers	A	2,298	6,701	8,999
Raw data files that result from doctoral dissertation research	Student papers	A	0	5,023	5,023
Raw data files that result from master's thesis research	Student papers	A	0	21	21
Raw data files that result from senior thesis research	Student papers	A	10	10	20
Senior theses	Student papers	A	478	954	1,432
Undergraduate student e-portfolios	Student papers	A	23	0	23
Blogs	University records	A	0	0	0
College, departmental, and school alumni publications	University records	A	89	113	202
Committee meeting agendas and minutes	University records	A	139	720	859
Committee meeting documents (e.g., budgets, reports, memoranda)	University records	A	141	540	681
Faculty senate agendas and minutes	University records	A	253	100	353
Regent, trustee, board meeting agendas and minutes	University records	A	120	100	220
Video recordings of performances	University records	A	457	113	570

**Appendix A** *(continued)*

Document Type	Classification	Archival (A)	PPT	IMP	Sum
Your institution's alumni publications	University records	A	170	0	170
Your institution's course catalogs	University records	A	766	10	776
Your institution's newspapers	University records	A	135	566	701
Other	Other	A	3,571	9,739	13,310
<b>Archival Subtotal:</b>			<b>23,189</b>	<b>53,780</b>	<b>76,969</b>
Software			5	20	25
Software documentation			0	15	15
Books	Pub		380	94	474
Journal Articles	Pub		1,311	8,769	10,080
Journals	Pub		132	64	196
Preprints	Pub		6,290	842	7,132
Working papers	Pub		1,360	12,893	14,253
<b>Non-archival Subtotal:</b>			<b>9,478</b>	<b>22,697</b>	<b>32,175</b>
<b>Grand Total:</b>			<b>32,667</b>	<b>76,477</b>	<b>109,144</b>
<b>Archival % of Grand Total:</b>			<b>71.0%</b>	<b>70.3%</b>	<b>70.5%</b>