

## Institutional Repositories

# SPEC KITS

Supporting Effective Library Management for Over Thirty Years

Committed to assisting research and academic libraries in the continuous improvement of management systems, ARL has worked since 1970 to gather and disseminate the best practices for library needs. As part of its commitment, ARL maintains an active publications program best known for its SPEC Kits. Through the Collaborative Research/Writing Program, librarians work with ARL staff to design SPEC surveys and write publications. Originally established as an information source for ARL member libraries, the SPEC series has grown to serve the needs of the library community worldwide.

## What are SPEC Kits?

Published six times per year, SPEC Kits contain the most valuable, up-to-date information on the latest issues of concern to libraries and librarians today. They are the result of a systematic survey of ARL member libraries on a particular topic related to current practice in the field. Each SPEC Kit contains an executive summary of the survey results; survey questions with tallies and selected comments; the best representative documents from survey participants, such as policies, procedures, handbooks, guidelines, Web sites, records, brochures, and statements; and a selected reading list—both print and online sources—containing the most current literature available on the topic for further study.

## Subscribe to SPEC Kits

Subscribers tell us that the information contained in SPEC Kits is valuable to a variety of users, both inside and outside the library. SPEC Kit purchasers use the documentation found in SPEC Kits as a point of departure for research and problem solving because they lend immediate authority to proposals and set standards for designing programs or writing procedure statements. SPEC Kits also function as an important reference tool for library administrators, staff, students, and professionals in allied disciplines who may not have access to this kind of information.

SPEC Kits can be ordered directly from the ARL Publications Distribution Center. To order, call **(301) 362-8196**, fax **(301) 206-9789**, e-mail [pubs@arl.org](mailto:pubs@arl.org), or go to <http://www.arl.org/pubscat/>.

Information on SPEC Kits and the SPEC survey program can be found at <http://www.arl.org/spec/>. The executive summary for each kit after December 1993 can be accessed free of charge at <http://www.arl.org/spec/complete.html>.



---

# SPEC Kit 292

---

Institutional Repositories

July 2006

---

**University of Houston Libraries**

**Institutional Repository Task Force**

Charles W. Bailey, Jr., Chair



ASSOCIATION OF RESEARCH LIBRARIES

---

Series Editor: Lee Anne George

SPEC Kits are published by the

Association of Research Libraries  
21 Dupont Circle, NW, Suite 800  
Washington, DC 20036-1118  
(202) 296-2296 Fax (202) 872-0884  
[http://www.arl.org/spec/  
pubs@arl.org](http://www.arl.org/spec/pubs@arl.org)

ISSN 0160 3582

ISBN 1-59407-708-8

Copyright © 2006

This compilation is copyrighted by the Association of Research Libraries. ARL grants blanket permission to reproduce and distribute copies of this work for nonprofit, educational, or library purposes, provided that copies are distributed at or below cost and that ARL, the source, and copyright notice are included on each copy. This permission is in addition to rights of reproduction granted under Sections 107, 108, and other provisions of the U.S. Copyright Act.

∞ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (R1997) Permanence of Paper for Publications and Documents in Libraries and Archives.

## **ABOUT THE AUTHORS**

---

This document was collaboratively authored by members of the University of Houston Libraries' Institutional Repository Task Force:

**Charles W. Bailey, Jr.**, Assistant Dean for Digital Library Planning and Development, Chair

**Karen Coombs**, Head, Web Services

**Jill Emery**, Director, Electronic Resources Program (now Head Librarian, Serials and Electronic Resources, University of Texas at Austin Libraries)

**Anne Mitchell**, Metadata Coordinator

**Chris Morris**, Web Developer 2

**Spencer Simons**, Director of the University of Houston's O'Quinn Law Library

**Robert Wright**, Coordinator of the Pharmacy Library



### SURVEY RESULTS

Executive Summary.....	13
Survey Questions and Responses.....	23
Responding Institutions .....	90

### REPRESENTATIVE DOCUMENTS

#### IR Home Pages

University of California CDL eScholarship Repository .....	94
University of Connecticut DigitalCommons@UConn.....	95
University of Delaware University of Delaware Library Institutional Repository .....	96
Université Laval Archimède .....	97
University of Toronto U of T Research Repository.....	98

#### IR Usage Statistics

University of Oregon Statistics for University of Oregon Scholars' Bank.....	100
University of Rochester UR Research.....	101

## **IR Deposit Policies**

University of California CDL Repository Policies.....	104
University of Nebraska–Lincoln DigitalCommons@UNL.....	106

## **IR Deposit Agreements**

Brigham Young University BYU Library DSpace Policies and Guidelines. Deposit License .....	110
University of Waterloo Author’s Declaration for Electronic Submission of a Thesis .....	111
Wayne State University License Agreement.....	112

## **IR Metadata Policies**

University of Kansas Describing a Resource (Metadata) .....	114
University of Texas at Austin Metadata Registry Project.....	117
University of Utah University of Utah Institutional Repository Metadata Best Practices .....	118

## **IR Digital Preservation Policies**

University of California CDL Overview of the UC Libraries Digital Preservation Repository (DPR) .....	122
University of Illinois at Urbana-Champaign IDEALS Digital Preservation Support Policy .....	128
University of Kansas KU ScholarWorks. Preservation Planning for Digital Information .....	131
Vanderbilt University VU e-Archive Supported Formats .....	133

## **IR Proposals**

University of Illinois at Urbana-Champaign	
Illinois Digital Scholarship: Preserving and Accessing the Digital Past, Present and Future .....	136
Kent State University	
OhioLINK Digital Resource Commons.....	153
Queen's University	
QSpace - Queen's Institutional Repository Project Plan.....	155

## **IR Promotion**

University of California CDL	
eScholarship Liaison's Library.....	162
University of Kansas	
7 reasons to use KU ScholarWorks .....	163
University of Toronto	
T-Space at The University of Toronto Libraries. Marketing Materials.....	165

## **SELECTED RESOURCES**

<b>General Works</b> .....	169
<b>DSpace</b> .....	170
<b>eScholarship</b> .....	171
<b>Fedora</b> .....	171
<b>Other</b> .....	172





## **SURVEY RESULTS**



---

## EXECUTIVE SUMMARY

---

### Introduction

Since 2002, when DSpace and other institutional repository (IR) software began to be available, an increasing number of research libraries and their parent institutions have established institutional repositories to collect and provide access to diverse, locally produced digital materials. This emerging technology holds great promise to transform scholarly communication, but it is still in its infancy.

For the purposes of this survey, an IR was simply defined as a permanent, institution-wide repository of diverse, locally produced digital works (e.g., article preprints and postprints, data sets, electronic theses and dissertations, learning objects, and technical reports) that is available for public use and supports metadata harvesting. If an institution shares an IR with other institutions, it was within the scope of this survey. Not included in this definition were scholars' personal Web sites; academic department, school, or other unit digital archives that are primarily intended to store digital materials created by members of that unit; or disciplinary archives that include digital materials about one or multiple subjects that have been created by authors from many different institutions (e.g., arXiv.org).

In this analysis, the authors have chosen to report the data as percentages based on the relevant number of responses (which can vary by question and within question) unless number totals are clearer. Percent figures have been rounded according to standard rules and they may total to slightly over or under 100%. For questions where respondents

supplied textual answers, the authors have only used unambiguous replies in their analysis and calculated percentages accordingly. Respondents who indicated that their institutions currently have IRs are referred to as "implementers." Respondents who indicated they have plans to develop an IR are referred to as "planners."

### Background

The survey was distributed to the 123 ARL member libraries in January 2006. Eighty-seven libraries (71%) responded to the survey. Of those, 37 (43%) have an operational IR, 31 (35%) are planning for one by 2007 at the latest, and 19 (22%) have no immediate plans to develop an IR.

One respondent had an operational IR as early as 1999 and a few more came online in 2002 and 2003. Implementation surged in 2004 as 12 repositories became publicly accessible; 14 followed in 2005. Two more were operational in early 2006 and an additional 11 are planned for later this year. Seven others expect their IRs to become accessible in 2007. (One planner indicated that planning and implementation has been ongoing since 2004.)

This data indicate that 30% of all ARL institutions had an operational IR at the beginning of 2006; by the end of 2007 the total may reach at least 55%. While the growth rate appears to be leveling off at this point, IRs will continue to be developed and implemented in the near future.

Implementers and planners are in general agreement about their motivations for starting an IR. The

top three reasons are to increase global visibility of, preserve, and provide free access to the institution's scholarship. These goals are followed closely by a desire to collect and organize the institution's scholarship in a single system (implementers, 89%; planners, 83%). Thirty-eight percent of implementers and 47% of planners were responding to requests for an IR from faculty, staff, and students. Among the other motivating factors was a desire to "Change scholarly communication by demonstrating alternative mechanisms," "Provide a solution to researcher's data management and data publication needs," and "Position the university as a leader in managing digital assets."

All respondents, implementers and planners alike, indicated that the library has been a driving force in the creation of or planning for an IR. Information technology and academic departments advocated for an IR about equally but trailed the library significantly. The administration was an advocate at only about a quarter of the responding institutions.

A wide range of academic units were specifically identified as advocating IRs, such as Aerospace Engineering, Anthropology, Art, Biology, Computer Science, Environmental Studies, Geography, Journalism and Mass Communication, Law, Mathematics, Medicine, Political Science, and Romance Languages. It should be noted that several respondents indicated Graduate Studies or Graduate School Services, with the latter unit explicitly mentioning an interest in electronic theses and dissertations (ETDs).

A variety of other areas on campus were also identified as advocates, such as the Center for Teaching & Learning Excellence, Honors Program, Institute for Policy Studies, Knowledge Media Design Institute, Senate Library Committee, University Archives, and University Press.

### **Planning, Implementation, and Assessment**

Thirty percent of the implementers engaged in planning for six months to a year. Twenty-four

percent took from one to six months and an equal number took more than a year to complete the planning stage. The planning process is ongoing for 19%. Only one implementer spent less than a month on the planning stage. More than half of the planners report that this stage is ongoing. For most of the others planning started within the last year.

Roughly a third of the implementers needed less than six months for the implementation phase. Another third took six months to one year to complete the process. Only two needed more than a year. For most of the remaining implementers, that task is ongoing. While almost half of the planners report that they have not reached the implementation phase, the others have either recently entered it or are simultaneously planning and implementing their IRs.

While more than a third of the implementers have not reached the initial assessment phase, yet, for many (43%) assessment is ongoing. A small number (8 or 23%) have completed some assessment. The majority of planners are not ready to assess their IR, but a few report some assessment activity.

Most institutions have conducted or will conduct a pilot project before making their IR public (implementers, 73%; planners, 86%). The pilot project serves multiple purposes. The top two are to determine potential difficulties or problems and plan contingencies and to test processes or procedures. Slightly less important are determining staffing needs (59% and 80%), evaluating and testing IR system options (41% and 80%), estimating costs (41% and 72%), and determining needed material resources (37% and 76%). Other purposes include testing campus interest, building support, and seeding the repository.

### **Staffing**

The overwhelming majority of respondents have appointed or will appoint a project group for planning and implementation (implementers, 92%; planners, 93%). Planning groups range from 2 to

26 members but most are fairly large. The mean number of group members for the implementers is 7.8 and the median is 6; for the planners the mean is 9.1 and the median is 7. Both implementers and planners reported a few groups larger than 15 members.

How library staff are selected to work on an IR implementation varied across the respondents. The top criteria for group members is their functional expertise (94% and 92%), particularly with technology, software, systems, and metadata. Managerial/supervisory expertise is the second criteria for implementers (62%) but only the fifth for planners (35%). Planners report that representation of a particular constituency, such as archives, IT, or faculty, is more important (81%). Being a liaison to a particular stakeholding group, especially faculty, and prior successful project experience are important criteria for a majority of respondents. The least important criteria is workload (15% and 31%), but among the other criteria reported was a willingness to take on additional work.

Survey respondents identified up to four units that were major players in the institution's ongoing IR operations. They indicated the unit name, its responsibilities, the title of the unit manager, the title of the person that the unit reported to, the number of individuals in each staff category (i.e., librarian, other professional, support staff, student assistant, and other), and total FTE in each category.

The data reveal that libraries and their staff are leading the campus IR effort and providing the majority of staffing support for it. There are no campus IT reporting lines for the first units, though there is one joint library-IT reporting line. A small number are found in the second to fourth units and they are often performing server support and similar functions. Out of 58 identified first units, there is only one non-library unit—a graduate studies office.

By far, the most frequent types of units reported for the first unit were digital library/initiatives or systems units within libraries. Also in the list were administrative units, archives, and research or technical services.

It is clear that IRs are a library-wide effort involving many different departments. Aside from technical support units, the lists of second to fourth units includes archives, cataloging, branch libraries, collection development, instruction, metadata, preservation, reference, and special collections. Most of these units report to upper levels of management, such as a library dean or associate/assistant director.

If the mean FTE values for each of the four units are added together, the average number of staff working on an implementers' IR is 28.1. The breakdown by staff category is 7.4 librarians, 7.3 other professional staff, 9.5 support staff, and 3.9 students. The average number of staff working on a planners' IR is 61.2. The breakdown by staff category is 8.8 librarians, 20.8 other professional, 22.2 support staff, and 9.4 students.

### **Budget**

Only 44% of implementers report having a dedicated budget for start-up costs; 48% have a dedicated budget for ongoing operations. Half of the planners anticipate having a dedicated budget for start-up costs and 40% expect to have a dedicated budget for ongoing operations. Many of the respondents who do not have a dedicated budget explained that costs for staff, equipment, etc. were either supported by general library operations already represented within the library's budget, or that existing budget lines were reallocated, or that a consortium or other third party absorbed the costs.

Implementers report a range of start-up costs from \$8,000 to \$1,800,000, with a mean of \$182,550 and a median of \$45,000. Planners report a range of \$12,000 to \$160,000, with a mean of \$81,667 and a median of \$75,000. The range for ongoing operations budgets for implementers is \$8,600 to \$500,000, with a mean of \$113,543 and median of \$41,750. Only two planners knew their budgets for ongoing operations—\$100,000 and \$133,000, with a mean and median of \$116,500.

The distribution of both start-up and ongoing budgets shows concentrations of responses at the

lower and upper ranges, with few in the middle ranges. For start-up, 67% of budgets fall below \$75,000, 14% are \$75,000 to \$125,000, and 19% are \$150,000 or greater. The maximum start-up budget (\$1,800,000) is far greater than the next highest (\$400,000) and is from an institution that included extensive software development and testing costs in its start-up budget.

For ongoing budgets, there is a similar concentration at the ends of the ranges: 50% are below \$50,000 and 50% are \$100,000 or greater. The maximum ongoing budget (\$500,000) is also much greater than the next highest (\$300,000) and is reported by an institution that has a major role in a state-wide IR initiative.

The primary method of funding both start-up and ongoing costs is reallocation from existing budgets, but respondents also reported a significant use of new funds. New funds for start-up costs most often came from grants (implementers, 83%; planners, 22%), the parent institution (50% and 33%), one-time supplemental funds (33% and 33%), and other sources (50% and 33%) such as provosts. In almost all cases, reallocated funds are or will be provided by the library (91% and 80%). A few respondents got reallocations from one-time supplemental funds, the information technology department or parent institution, and other sources such as student fees.

New funds for ongoing operations most often are or will be provided by the parent institution, grants, or the library. Reallocating funds is almost always the responsibility of the library (78% and 67%). As with start-up funds, a few respondents got reallocations from one-time supplemental funds, the information technology department or parent institution, and student fees.

Not surprisingly, for the majority of implementers, salaries and benefits account for the largest component of the budget—63% of start-up budgets and 68% of ongoing budgets, on average. This is exceeded only by vendor fees for the small number of institutions whose IR is hosted by an exter-

nal vendor—70% to 74%, on average. Hardware and software acquisition each account for about a quarter of the start-up budgets and hardware and software maintenance account for under 10%, on average. The allocation for acquisition decreases slightly for ongoing operations and the maintenance allocations increase correspondingly.

Planners allocate the largest percentage of their start-up budgets for hardware acquisition (about 58%, on average) and software acquisition (38%) and a small amount for hardware maintenance (10%). Only one respondent reported a figure for staffing and benefits (57%). For the few planners who have an ongoing operations budget about three-fourths of the budget is allocated for salary and benefits. Much of the rest goes to hardware maintenance.

### **Hardware and Software**

Fifty-three respondents identified the software that is being used to support their IRs. By far, the open source DSpace software is the most common choice of both implementers and planners. Twenty-three of the 33 responding implementers and 14 of the 20 planners (70% each) use DSpace; 20 implementers and 11 planners use it exclusively. Two of the implementers use it in conjunction with CONTENTdm (commercial software); one of these also uses the vendor-hosted DigitalCommons system. One implementer uses DSpace in conjunction with ETD-db and Open Conference Systems (both open source software). Two planners have chosen it in conjunction with open source Fedora software; another with commercial software Digitool.

Of the respondents that don't use DSpace, one implementer uses open source Archimède software and two use commercial CONTENTdm software. Two planners will use open source Fedora software and one will use open source Greenstone software. CONTENTdm, Digitool, and Documentum are the intended commercial systems for one planner each. The ProQuest DigitalCommons system (or the software from the Berkeley Electronic Press

it is based on) is used on all the vendor-hosted platforms.

There is a greater variety of hardware in use. Implementers are about evenly divided between Intel-based servers (Dell models in particular) with either Linux or Windows operating systems and Sun servers with Solaris. Only one uses an Apple Xserve running OS X. All but two of the planners use Intel-based servers, again primarily Dell models, with Linux or Windows operating systems. One uses an IBM RISC server and the other a Sun system.

Memory on the implementers' Sun systems ranges from 2 GB to 64 GB and disk storage ranges from 10 GB to 6 TB. (One institution reported a 12 TB storage unit, but it was not dedicated to the IR.) The Intel-based servers have memory ranging from 512 MB to 4 GB and disk storage ranging from 28 GB to 3 TB. The Macintosh server has 1 TB of storage space. The planners' Intel-based servers have memory ranging from 512 MB to 12 GB and disk storage ranging from 108 GB to 3 TB. The IBM RISC-based server has 4 GB of memory and 275 GB of storage space.

Roughly a third of respondents have made no modifications to the IR software and another third have made minor modifications. About 20% of both implementers and planners have made major modifications to IR software. Implementers are much more likely than planners to have made frequent changes, 22% vs 6%.

### **Policies and Procedures**

Seventy-five percent of implementers and 71% of planners indicated they have or will have written policies and procedures for their IRs. For both groups, 54% have submitted their policies and procedures to an institutional authority for review, or are planning to do so. Most of those who identified the reviewing authority indicated that their policy documents went to the University Counsel.

Respondents place a wide variety of materials in their repositories. Electronic theses and disserta-

tions are the most common type of deposit (implementers, 67%; planners, 79%). Articles, including preprints and postprints follow closely. The majority of respondents include official published versions of articles, conference presentations, technical reports, and working papers. Only a few include university catalogs, yearbooks, or alumni publications.

Only a handful of respondents are actively negotiating with publishers to secure permanent deposit of e-prints from published serials, but 46% of implementers and 63% of planners are considering doing so in the future.

The widespread inclusion of traditionally unpublished material in IRs may reflect the relative ease of recruiting this type of content as well as the fact that these materials in print format do not have robust publishing avenues. Data sets, learning objects, and multimedia materials are the most prevalent non-traditional materials deposited, with over a third of all respondents indicating they include or will include these materials in the IR. Several respondents mentioned using the IR to house retrospectively digitized images and other archival material.

Seventy-four percent of implementers and 83% of planners indicated that they accept any digital file type into the IR, but relatively few (26% and 39%) are committed to functional preservation of every file type. Eighteen percent of implementers and 17% of planners will only accept and preserve specified file types. A few accept certain file types but do not preserve them. Several respondents mentioned following the support levels outlined in MIT's DSpace guidelines (<http://www.dspace.org/implement/policy-issues.html#digformats>), which include full support and preservation for common file types such as PDF, XML, AIFF for audio, and GIF, JPEG, and TIFF for images, among others.

Most deposits to the IR are or will be made by authorized depositors (implementers, 89%; planners, 79%). A significant number of IR staff also

deposit documents for authorized users (78% and 63%). Most respondents are using both methods for deposit rather than one or the other, and many respondents indicated that their deposit procedures are still under construction. Almost all respondents indicated that faculty may deposit their materials in the IR and both implementers and planners allow the work of professional staff, students, and support staff to be deposited, as well. Several respondents also mentioned faculty sponsorship as a means of bringing in work for students or outside contributors.

Fifty-nine percent of implementers and 73% of planners have some method for reviewing documents for copyright compliance or other reasons. In most cases, the review is by individuals outside the IR unit. Only nine respondents report that IR staff review and approve documents. These procedures are not always systematic, but vary by collection and type of document. Additionally, all but six respondents require depositors to sign a deposit agreement.

Sixty-nine percent of implementers and 62% of planners accept multiple versions of the same document. In most cases, versioning is achieved by appending additional files to the original item; most institutions do not permit the depositor to overwrite an earlier version of the document. Eighty-two percent of respondents indicate that IR staff or authors may withdraw documents from the IR. Comments suggest that most institutions permit withdrawal only in cases of copyright infringement or other legal issues. Institutions that permit the withdrawal of documents generally leave a “tombstone”—a reference to the withdrawn document—in the system as a record. Many respondents felt that policy in this area would solidify as they learned more about the legal landscape.

### **Metadata**

Roughly half of the respondents import metadata into their IRs from outside sources, typically by a process of automated mapping from a variety of

schemas into Dublin Core. Many are converting data from local schemas and a surprising number mentioned converting records from the MARC format. Ninety-four percent of implementers and 78% of planners allow depositors to enter simple metadata; many of these same respondents also enter metadata on behalf of depositors (implementers, 60%; planners, 56%) or enhance depositor supplied metadata after the fact depending on the material and source (57% and 72%).

Survey results indicate that many institutions are taking significant steps to ensure that their IRs are interoperable with other systems. Ninety-four percent of implementers and 88% of planners indicate that their IR supports the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and a similarly large percentage (89% and 82%) report using persistent identifiers for materials in the IR. This result is consistent with the widespread use of DSpace and other platforms to which these services are integral. Over half of implementers (54%) have ensured that their systems are OpenURL compliant and 41% of planners anticipate having this functionality. Sixty-three percent of implementers have incorporated their IRs into federated search mechanisms. Since only one planner reports doing so, it may be that institutions still in the planning stages have not yet considered how to incorporate the IR with other services, or these institutions may lack federated search tools altogether.

Although most IR platforms in widespread use are OAI-PMH compliant, only one respondent specifically mentioned being crawled by search engines. The issue of optimizing exposure to search engines may become more significant as IRs become more prevalent and stable.

### **Content Recruitment**

The difficulties faced by institutions when recruiting content for their IRs is clearly borne-out by respondents. Only one implementer found recruitment “very easy” and only seven “somewhat easy.” Fourteen (40%) found it “somewhat difficult,” and

eight (23%) “very difficult;” five (14%) were neutral. In other words, nearly two-thirds of implementers surveyed were sufficiently challenged by the task of recruitment to label their efforts “difficult.”

By contrast, about half of the responding planners were neutral. The remainder were evenly divided between “easy” and “difficult.”

This difference in perceptions between implementers and planners may reflect both a simple difference in experience and/or the change in perceptions of implementers after an initial recruitment phase of easily identified departmental content. It may be that it becomes increasingly difficult to recruit content after this initial set of objects is added to the IR.

A variety of recruitment strategies are employed or planned by respondents. The majority have tried subject specialist advocacy, identifying likely depositors, presentations to faculty, and offering to deposit electronic materials for authors. Implementers appear to be more aggressive with additional strategies, such as sending electronic announcements, faculty co-recruiting, offering to digitize and deposit printed material for authors, and holding awareness-raising symposia. This practice may indicate that implementers have reacted to recruitment difficulties by trying more and more recruitment strategies.

One recruitment strategy not mentioned above is institutional pressure on authors to submit content to IRs. Only one implementer requires authors to submit content to the IR. One implementer and one planner are considering such a requirement. Half of the implementers and two-thirds of the planners report there is no pressure on authors to submit content. The rest encourage, but do not require, authors to submit content.

### **Assessment**

A small number of implementers (8 or 22%) have conducted research on why users do or do not contribute to the IR; only five planners (28%) have decided to conduct any research. This seems odd

since the success of an IR is highly dependant on users contributing to the IR. One explanation for this might be that about a third of the implementers and 71% of the planners answered that they had not yet reached the assessment phase. Because few institutions have conducted assessment of contributor motivation, there is likely to be limited data regarding what factors influence users who contribute to repositories.

While close to 70% of the implementers who have done some form of assessment of the success of the IR have gathered direct feedback from IR users through interviews, surveys, or focus groups, the majority (23 or 79%) have tracked hits on IR content. This is likely due to the fact that it is fairly simple to collect “hit” data from server log files, while the collection and analysis process for more ethnographic user data is significantly more time consuming.

It is clear from the comments that there are many different viewpoints on what constitutes “success” for a repository. One respondent commented about assessing the usability of the interface, while another responded about counting full-text downloads. Clearly, there are many aspects of an IR which need to be examined to determine success.

### **Current Status of IR**

Because the survey respondents have repositories at various stages of development, the numbers of digital objects in the IRs differ significantly. Implementers report a range of 20 objects to over 19,000. Planners report between 4 and 4,500 objects in their repositories. Interestingly, not all the materials stored in the repositories are available to everyone. Forty-four percent of the implementers (16) have material within their repository that is available to only a specific user group, while 36% of the planners (5) intend to restrict access to parts of their IR to specific groups.

Comments from the respondents indicate that there are different reasons for these restrictions as well as different groups to whom use is being re-

stricted. For example, some repositories contain materials that are limited only to campus users, while others have materials that are limited to a specific department or groups of people (such as a specific group of research faculty). Copyright is only one reason that access to materials is limited. Cultural concerns with primary source materials and pending patents were also cited as reasons for restricting access.

Although some institutions restrict access to materials within their repository, few implementers (3 or 9%) supply IR documents to external users for a fee. Primarily, fees appear to be for re-use of images or electronic thesis and dissertations. This seems somewhat incongruous when one considers that 44% of the implementers limit access to materials within their repository. However, there are several possible explanations for this. First, institutions could be storing institutionally licensed materials, such as images, in their IR. Second, the process of collecting per-use fees is missing from several popular open source software packages for IRs. This makes it difficult for institutions to collect fees on a per-use basis without extending the software.

### **Benefits**

Respondents' comments indicate that the top two benefits of IRs are enhanced visibility and increased dissemination of the institution's scholarship (34 responses or 68%) and free, open, timely access to scholarship (23 or 46%). Preservation and stewardship of digital content and preservation of and long-term access to the institution's scholarship are close seconds (18 responses each or 36%), followed by collecting and organizing assets in a central location (12 or 24%). Four respondents (8%) report that another benefit of an IR is the opportunity to educate faculty about copyright, open access, and scholarly communication.

### **Challenges**

Among the top three challenges that respondents face in implementing, promoting, and running an

IR are content recruitment/building a critical mass of content (16 responses or 32%), staffing (15 or 30%), and faculty awareness/buy-in/interest/engagement (14 or 28%). Copyright issues and communicating the benefits of the IR to faculty are close behind. Adequate funding and other resources and integrating the staff and workflow of IRs into existing structures were also recognized as challenges.

### **Conclusion**

Based on the survey, what were the major characteristics of operational ARL institutional repositories at the start of 2006? Most IRs had been established in the last two years (or had just been established). By far, the library was likely to have been the most active institutional advocate of the IR. It was also likely to have been the primary unit leading and supporting the IR effort, sometimes in partnership with the institutional information technology unit. The main reasons for establishing an IR were to increase the global visibility of, preserve, provide free access to, and collect and organize the institution's scholarship. In most cases, a project team had been used to plan and implement the IR and a pilot project had been used to determine IR-related issues. If it was not still ongoing, the IR implementation process had most frequently taken six months to a year, with one to six months being the next most common duration.

By a large majority, the most frequently used local IR software was DSpace, with DigitalCommons (or the bepress software it is based on) being the system of choice for vendor-hosted systems. Local systems usually either ran under variants of Linux or Windows on an Intel-based server or under Solaris on a Sun server. A typical IR held about 3,800 digital objects, with ETDs, article preprints and postprints, conference presentations, technical reports, working papers, conference proceedings, and multimedia materials being the most common types of documents. IRs normally support OAI-PMH and, a little over half the time, OpenURL.

Most IRs had written policies and procedures and the majority of them had been submitted to an

institutional authority for review. Faculty members were almost always authorized to directly deposit digital objects in the IR, and professional staff and students were typically able to do so as well. These depositors almost always signed a deposit agreement. Most institutions also authorized IR staff to deposit objects on behalf of users. A deposit review process was common, with documents most frequently being reviewed by department or other institutional officials. Authorized depositors were almost always allowed to enter metadata; IR staff could typically do so as well, plus enhance existing metadata. Most IRs accepted multiple versions of the same document. Document withdrawal was usually possible, but typically had to be done by IR staff under specific circumstances. The vast majority of institutions intended to preserve IR documents, but most of those doing so limited the types of files that would be preserved. Most institutions found IR content recruitment to be somewhat or very difficult and they usually engaged in a variety of recruitment strategies to increase deposits.

The average IR start-up cost had been around \$182,500 and its average ongoing operation budget was about \$113,500. Reallocated funds from the library's budget were a key source of IR support, as were new funds from grants and the parent institution. Staff had been the largest single IR budget item during start-up and it remained so in ongoing budgets. Many IRs were funded without

dedicated budgets, using existing personnel and technical resources.

The typical IR was supported by about 28 FTE from a variety of units within the library and elsewhere, a digital library/initiatives unit managed it, and that unit reported to a high-level library administrator, such as an assistant or associate dean/director. Most institutions modified their IR software to some degree to enhance its functionality.

As one would expect, the perceptions of institutions still planning IRs did not always match the experience of implementers as outlined above, with differences most often occurring over resource and time requirements as well as levels of difficulty. Since these matters can be difficult to accurately project and little data existed at the time the survey was administered that offered guidance, this is not surprising.

Although institutional repositories are at an early stage of development, ARL libraries have demonstrated a strong preliminary commitment to them: 78% of the 87 survey respondents had either implemented an IR or were planning to do so by the end of 2007. Since IRs represent a significant long-term organizational commitment, this is a major expansion of ARL libraries' service role and, along with digital library functions, aptly illustrates how these libraries are rapidly evolving into global digital information providers.

