

SPEC KITS

Supporting Effective Library Management for Over 40 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 Kit 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 are available in print and online. The executive summary for each kit after December 1993 can be accessed online free of charge. For more information visit: <http://www.arl.org/publications-resources>.



SPEC Kit 334

Research Data Management Services

July 2013

David Fearon, Jr., Betsy Gunia, and Barbara E. Pralle

Johns Hopkins University

Sherry Lake and Andrew L. Sallans

University of Virginia



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

P (202) 296-2296 F (202) 872-0884

<http://www.arl.org/publications-resources>
pubs@arl.org

ISSN 0160 3582

ISBN 1-59407-901-3 / 978-1-59407-901-6 print

ISBN 1-59407-902-1 / 978-1-59407-902-3 online

Copyright © 2013

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 US 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.

SPEC Kit 334

Research Data Management Services

July 2013

SURVEY RESULTS

Executive Summary	11
Survey Questions and Responses	23
Responding Institutions	105

REPRESENTATIVE DOCUMENTS

RDM Web Pages

University of California, San Diego	
Research Cyberinfrastructure	110
Emory University	
Research Data Management	111
University of Wisconsin-Madison	
Research Data Services	112

IRs and Data Archives

Columbia University	
Academic Commons	114
University of Florida	
IR@UF: The Institutional Repository at the University of Florida	115
Harvard University	
Harvard Dataverse Network	116
University of Nebraska-Lincoln	
UNL Data Repository	117
University of New Mexico	
Lobo Vault – UNM Research Data	118
University of Oregon	
Scholars' Bank	119
Pennsylvania State University	
ScholarSphere	120

Purdue University	
Purdue University Research Repository (PURR).....	121
University of Virginia	
LIBRA: Online Archive of University of Virginia Scholarship.....	122

Data Management Plan Tools

Indiana University	
Indiana University Guidance on NSF Data Management Plans	124
Johns Hopkins University	
Questionnaire to Help with the Creation of a Data Management Plan.....	128
University of Minnesota	
Data Management Plan	131
University of North Carolina	
Research Data Toolkit	134

Job Descriptions

Johns Hopkins University	
Data Management Consultant	136
University of Maryland	
Research Data Librarian	138
University of Ottawa	
Social Sciences Research Liaison Librarian.....	141
Purdue University	
Digital Library Software Developer	144
Digital Data Repository Specialist	148
University of Tennessee	
Data Curation Librarian	152
University of Virginia	
Data Consultant.....	154

Organization Charts

Brown University	
Center for Digital Scholarship	156
Cornell University	
Research Data Management Service Group Organizational Structure	157
Duke University	
Information Technology Services.....	158
Massachusetts Institute of Technology	
Specialized Content and Services	159
University of Rochester	
River Campus Libraries Organizational Chart.....	160

Research Data Policies

University of Iowa	
Data Management: Research Records	162
Johns Hopkins University	
Policy on Access and Retention of Research Data and Materials	164
McGill University	
Regulation on the Conduct of Research excerpt	167
University of Tennessee	
Research Data Policy	169

Data Retention Policies

Duke University	
Policy on Research Records: Sharing, Retention and Ownership	174
Emory University	
Policy 7.9: Guidelines for Responsible Conduct of Scholarship and Research	175
Section 7.9.02C: Access to and Retention of Scientific Research Protocols and Data	176
Johns Hopkins University	
Appendix V from Policy on Access and Retention of Research Data and Materials	177
Northwestern University	
Research Data: Ownership, Retention and Access	178
University of Virginia	
Policy: Laboratory Notebook and Recordkeeping	182
University of Wisconsin – Madison	
Policy on Data Stewardship, Access, and Retention	183

Data Needs Assessment

Cornell University	
Research Data Management Group Survey of NSF Principal Investigators at Cornell University	188
Emory University	
Research Data Management. Faculty Survey Results	197
Massachusetts Institute of Technology	
A Data Planning Checklist	198
Purdue University	
Data Curation Profiles Toolkit	199
University of Virginia	
Data Interview Protocol	200

RDM Staff Resources

University of Florida	
Data Management/Curation Task Force	206

Purdue University	
Supporting Information for Data Services.....	207
Yale University	
Data & eScience Group.....	208

Outreach Materials

University of Colorado at Boulder	
Research Data Services Outreach Card.....	210
University of Florida	
Speaker Series	211
Indiana University Bloomington	
Workshops	212

SELECTED RESOURCES

Research Data Management Services (RDMS) and the Library	215
Supporting Data Management Planning	216
Data Archiving.....	216
Staffing and Staff Training	217
RDMS Policies	218
Scanning and Assessment of the Landscape for RDMS	218
Key Papers in the Development of RDMS	219
ARL/DLF/Duraspace eScience Institute	219



SURVEY RESULTS

EXECUTIVE SUMMARY

Introduction

With an increased emphasis on open data, data management planning requirements, and potential “big data” research opportunities, research institutions are recognizing an emerging demand to provide a wider and more refined array of data services to meet needs at different points in the research process. Many research libraries are answering that demand by expanding or adopting new research data services, most within the last three years. The timing was appropriate, therefore, to survey ARL member libraries, assessing early endeavors, and benchmarking future growth as we anticipate demands for these services to expand and for libraries to find new opportunities for relevant services.

In this context, we identified two emerging areas of services that are relatively new for member libraries. These are research data management, which enters the research process at the grant proposal stage in meeting data management planning requirements, and in various ways during the research process. Second is support for data archiving, at a project’s end for preservation and online dissemination to facilitate data sharing, and in providing new data resources for discovery. This survey will use the term “research data management services” (RDMS) to refer collectively to library activities surrounding data management and archiving.

Our survey also addresses contextual requirements for planning, developing, staffing, and managing new research data management services. Finding necessary expertise and funding for new positions is challenging, but creative new models of service provision are emerging. Collaboration across institutional units is one route toward unearthing expertise

and knowledge to help researchers at all stages of the research process and to provide data management support. Our survey findings, however, point to the “growing pains” of new service development, with challenges such as initiating and encouraging campus-wide coordination that addresses gaps and overlapping services. Responses throughout the survey show that libraries are still in the early stages of development and implementation of RDMS. In most cases, services are evolving ahead of evidence of which models and strategies will prove most effective or successful. Variables for that development include the structured and unstructured institutional environment for new research services: Are more universities recognizing a community need formally through data policies (See Q1)? Are administration and researchers looking to libraries for solutions or are libraries taking their own initiative (Q6)? At what stage are libraries in conducting needs analysis to guide service development, engaging in active outreach to communicate their provision of new services (Q50), or assessing their early efforts (Q51)?

This SPEC survey of research data management services at ARL libraries joins a growing literature of surveys and case studies covering various dimensions of this emerging domain.¹ To a degree, we are still learning what questions we should ask to assess current practices and provide benchmarks for assessing future developments. We address a breadth of aspects that has not been consolidated before to encourage further research, but perhaps more importantly, to give libraries a timely orientation to the challenges and benefits of offering research data management services (henceforth referred to as RDM services).

Background

This survey builds upon and extends the findings of the 2010 report entitled “E-Science and Data Support Services: A Study of ARL Member Institutions,” authored by Catherine Soehner, Catherine Steeves, and Jennifer Ward, and sponsored by the ARL E-Science Working Group.² In preparing for our survey, we spoke with the authors of the report to discuss which aspects of the study went well and which areas would have benefited from continued investigation. Among the strengths, the authors specifically noted valuable responses to inquiries about how organizations began their efforts, collaborative approaches, educational efforts, pressure points or barriers to entry, and refocusing of professional interests. As key areas to explore further in our survey, the authors identified semantic and contextual clarity, more detail of academic and work backgrounds, ties between data management and cyberinfrastructure, perceptions of responsibility, areas of intended investment, and actions which defined success. While we did not have enough space to address all new questions, we have attempted to incorporate many of those points into our survey questions. We recognize that the absence of new, comprehensive case studies is a limitation of our study. This may be a fruitful direction for a next round of inquiry.

One of the areas that we recognized as important in building upon the 2010 report was further refinement of terminology, which may have varying interpretations or ambiguity among respondents and more broadly in this field of service. Examples of such terms include “eScience,” “cyberinfrastructure,” “data archiving,” and distinguishing research data management from data services more broadly, the latter sometimes including institutional records management. Perhaps the most ambiguous term still emanating through these conversations is “digital services.” As the survey results show, in application this can range from digitization to research support to intellectual property to repository management. While we attempted to specify particular meanings and control response variables for the purpose of a more systematic analysis, this process and some of the responses we received highlight the fact that there is still great variation in maturity and definition of

services throughout the ARL community. Based on anecdotes, we believe that there is also even broader variation and understanding of the meaning of these terms and services beyond the ARL community, such as with the consumers of the services. Among other goals, we hope that this type of study may shed some light on the variations of meaning, and provide some opportunity for further maturation and convergence of terminology.

Broad Data Support Services

Seventy-three of the 125 ARL member libraries responded to the survey. All are academic libraries. Respondents were asked whether and how long they have offered 11 broad types of data support services (Q3). All 73 libraries offer at least one of the listed services. Helping researchers locate and use data sources is the most common and long-established service (68 responses, or 93%). Also common and long-offered are support for geospatial analysis (61, or 84%), dataset acquisition (58, or 79%), and copyright and patent advising (53, or 74%). Sixty-four respondents (88%) provide an institutional repository; 48 for more than three years. About half of the respondents plan to add one or more services, particularly data visualization, data mining, and data analysis.

Research Data Management Services

Respondents were also asked whether they offer research data management services, defined as “providing information, consulting, training or active involvement in: data management planning, data management guidance during research (e.g., advice on data storage or file security), research documentation and metadata, research data sharing and curation (selection, preservation, archiving, citation) of completed projects and published data” (Q4). Almost three-quarters do (54, or 74%). Seventeen others plan to (23%). Only two have no plans to offer RDM services. One of those two commented, “Rather, no, we wish we could, but we have no formal plan to do so in the short term because of lack of resources.” The other explained that there is little to no demand for such services.

Origin of Research Data Management Services

Several survey questions addressed the current and transitioning climate of support for data services at each library and their institution. The responses suggest that libraries are developing data support services ahead of formal policy requirements of their institutions. Fewer than a quarter of the 73 respondents (16, or 22%) reported that their institutions have some form of policy for research data management or retention (Q1). However, those policies vary from IRB guidelines for sensitive data to institutional records policies, with few specifying that research data be kept and managed. A third of the respondents (24, or 33%) reported that policies are planned in the next one to three years, suggesting a trend of institutions to keep up with expanding funder requirements.

Only four libraries initiated RDM services before 2005 (Q5). The earliest reported was the 1966 library-supported Latin American Data Bank project. Ten others started their RDM services between 2005 and 2009. This correlates to the early initiatives for eScience, which was the hot topic of many papers and task forces.³ For these early providers the most important reasons for beginning service were researchers' requests for help and a library initiative to expand support of faculty research (Q6). One reports their AUL "at the time was a visionary in terms of DRM and initiated a broad range of services."

Five of the 11 libraries that started their services in 2010 also reported the influence of library initiatives. Four others stated the main reason was the National Science Foundation announcement that they would begin requiring data management plans on January 18, 2011. The NSF requirement was the main reason for 11 of the 16 libraries that started RDM services in 2011, as well.

In 2011 and 2012, ARL and CLIR/Digital Library Federation co-sponsored the E-Science Institute, a workshop series to help libraries develop e-research strategic agendas. The institute is now operated by DuraSpace and is open to non-ARL institutions.⁴ Forty-nine survey respondents (67%) have attended one or both of the previous sessions; seven (10%) say they plan to attend a future session, four for the first time (Q2). Of the 49 who have attended the institute, 40 provide some level of RDM service (82%). The four

who are planning to attend a future institute for the first time also already provide RDM services.

The core of the survey focused on the RDM services that support the management and curation of research data throughout its life cycle. The following table presents the range of RDM service categories discussed below, with their corresponding survey questions.

RDM Services (N=54)	N	%	Q
Online Data Management Plan (DMP) resources	47	87%	7
DMP training	33	61%	10
DMP consulting	48	89%	11
RDMS besides DMP support	53	98%	18
Data archiving by library	40	74%	19
Data-specific archive (other than institutional repositories)	5	9%	21

Data Management Plans

Many libraries began their RDM service to help researchers create data management plans (DMP), most often for NSF proposals. Two and a half years since the NSF's DMP requirement began, and with several other funders adding similar requirements, the survey asked several questions to gauge the extent to which the libraries with RDM services currently provide online DMP resources, or training and consulting on plan preparation. As seen in the table above, 47 libraries provide online resources related to data management plans. All of those libraries include an explanation of funding agencies' DMP requirements and guidelines for creating DMPs. All but a few have a tool or resource for DMP creation and DMP template examples. In addition to DMP planning information, other online resources include information about digital repository services, long-term data management and preservation, links to related campus services, information on copyright, and workshops offered.

Most respondents are providing links to external resources, but 70% have also created their own, and almost half have customized others' content. Forty-one libraries (75%) have linked to the DMPTool, a prominent online self-service resource for researchers

operated by the California Digital Library and a consortium of contributing institutions, and 29 of those offer training and support for the tool (Q8 & 9).

A rough indicator of the breadth of RDM support at the responding libraries is whether they have gone beyond online resources to add training on DMP preparation and/or consultation on DMPs for grant proposals. Forty-eight libraries offer consultation services. Ten began in 2010, 23 in 2011 (again due to the NSF DMP requirement), 12 in 2012, and 2 more in 2013. Thirty-three of these 48 also provide DMP training. At 25 institutions DMP support is also provided by various other departments, most frequently the office of research and office of sponsored programs (Q10–Q12).

Most direct consultations (other than workshops) are done via e-mail/chat/phone (94%) or meeting with researchers at their office, lab, or other location (92%) (Q13). Of all consultation methods used, many respondents commented that workshops, training, and tutorials have provided the most contacts. One commented on a unique way they get consultations: “In addition to providing our contact information to faculty through departmental grant administrators and our partners in the institution, faculty can also indicate that they would like a consultation through our internal grant tracking system.”

While all 48 libraries that offer DMP consultation services interact directly with researchers, it was challenging to gauge the extent of engagement, and how many researchers are actually looking to their library for DMP assistance. When asked about the number and depth of consultations, only 28 libraries reported that they kept track of the number of consultations (Q14). About half of that group reports having more than 10 consultations since their DMP service began. Seven of the ten libraries that started in 2012 and 2013 have consulted on fewer than five plans total. Only two libraries average more than three consultations per month. It is possible that there was confusion of what we meant by “DMP sessions” in our survey question, which had aimed to measure individual one-on-one consultations via any means, e-mail, face-to-face, etc. One response to this question was 300 DMP sessions. This high number may have included

number of participants in group sessions; the next highest number was 96 sessions.

When asked which departments/fields of research at their institutions use the DMP services, the majority reported that natural sciences, social sciences, engineering, and humanities use them at least occasionally (Q17). Overall, most of the libraries that are offering more than online DMP resources are receiving modest participation from their researcher communities. If more funders add DMP requirements for proposals, and/or increase compliance and accountability for the quality and follow-through on plans in the next few years, use of these services should be resurveyed, ideally with more detailed usage metrics and qualitative details on models of service provision.

RDM Services Other than DMP Preparation

Helping researchers prepare data management plans for grant proposals is a relatively focused category of service, and for many of the libraries surveyed, an entry point into this relatively new area of support for the research process. The survey next explored whether libraries offer services beyond DMP assistance, and asked about nine categories of additional RDM services that we expected some portion of ARL libraries to offer (Q18). A majority of the 54 responding libraries (36 to 48) offer eight of the nine services, including data management best practices (both online resources and workshops), helping researchers identify (and apply) appropriate metadata standards, research file organization and naming, data citation, data sharing and access, and data storage and backup. The last two services are commonly offered by both the library and elsewhere on campus. Only 14 libraries provide help with securing & anonymizing data. Half said this is only offered elsewhere, such as by the institutional review board, privacy office, or research compliance office. In addition to the listed services, a few libraries report they support data publication, data rights management, and analog to digital image data conversion.

Data Archiving Services

Even with moderately enforced requirements to share datasets from grant projects, publications, and other research by public funders such as NSF, most of the

responding libraries have made initial forays into data archiving services. As funders expand data sharing requirements and efforts at compliance,⁵ libraries may see an increasing role and call from researchers to assist in finding data archiving and dissemination solutions.

The survey defined data archiving as providing “longer-term retention of and access to research data by others.” A data archive is distinguished from other digital file storage systems by features that include online file access, file integrity checks, and permanent identifiers for locating files (Q19). Forty of the 54 respondents who provide RDM services (74%) report their library offers data archiving. At 13 of these institutions archiving is also offered elsewhere. Five institutions (9%) only offer data archiving elsewhere than the library. The other archiving units are most often central IT and research centers; however, in some cases such services might have been considered centralized storage rather than archiving by our definition.

Whether or not libraries offer their own archiving solution, nearly all the libraries with RDM services offer assistance locating data archiving solutions for research data, for example, at data repositories for a specific domain of research such as neuroscience (Q20). Twenty-six libraries (48%) help researchers deposit data at such repositories, and 21 (39%) have assisted with depositing data at journals, which are increasingly asking for supporting datasets. All but one of the 40 libraries that have their own archives offer direct assistance with depositing data in that archive.

Data Archive Characteristics

From prior literature and informal observations, we expected current archiving solutions among libraries to fall into two categories: repositories designed specifically for research data, and those repositories built for other purposes that can contain datasets. For the latter, we expected institutional repositories (IRs) used primarily for publications to be most common, as well as datasets archived in digital repositories that are used for institutional special collections such as historical photos. We expected data-specific archives built specifically for retaining and accessing research data to be less prevalent because of the higher resource

requirements for setting up a system, the lack of ready-made data repository platforms (in contrast to the more mature enterprise of institutional repository software), and the relative lack of demand for libraries to initiate such platforms at this time.

Question 21 attempted to distinguish these two categories, but after comparing Q21 responses to those for other questions, and to links respondents supplied to their archive websites, we found varying interpretations of what respondents reported as a “research data archive dedicated to data deposits and access.” All but 5 of the 13 “data archives” appear to be institutional or digital repositories by our definition. The survey questions and responses section lists the original responses, but the table below redistributes the 40 data archiving libraries’ primary method into three categories: institutional repositories (IRs) that include datasets in addition to publications; digital repositories used for a variety of institutional collections besides data, such as digital photos, rather than faculty publications; and archiving solutions more clearly dedicated to research datasets.

Archive Type (N=40)	N	%
IR with datasets	30	75%
Digital repository with datasets	5	13%
Data-specific repository	5	13%

Since 88% of the libraries that archive research data use either publication- or digital collection-centered repositories, the analysis of the survey’s data archiving questions distinguishes between responses of the five libraries we labeled as having “data archives” from those with institutional repositories or digital collections, which we combined into a single category “IRs with data.” This categorization may not be exact in all cases, so we encourage those doing further studies on library data archiving to verify directly with the libraries their current methods.

Funding Data Archives

As is the case for funding RDM services overall (Q49), most of the libraries with data archiving services (31 of 37, or 84%) are absorbing those costs through their internal budgets (Q24). Nine fund archiving through

grants (24%), five charge researchers (14%), and seven have found funds through other means (19%). No library charges users for data access.

There are clear differences between IRs with data and data-specific archives. Thirty of the IRs (94%) absorb any extra costs for research data into the repository budget. Only one of the data archives reported funding from their general budget. Two of the data archives are grant funded, and three charge researchers for archiving. Charging researchers or fees upon grants are much less common for IRs with data. Understanding the costs of archiving in many cases is still under review, and institutions recognize that archiving costs will need to factor in the volume of data and length of hold.

Data Archive Infrastructure

The survey asked respondents to describe the platform and software used for their data archiving solution (Q22). Most of the 38 respondents use open source software for all or part of their solutions; one developed their own software. DSpace is the most commonly used institutional repository and digital collection platform and interface (17, or 43%). Fedora is the platform for eight of the IRs (20%), often along with additional software interfaces such as Hydra or iRODS. The five data archives use Fedora and Data Conservancy software, Chronopolis, a customization of HubZero, a multi-component system that includes Fedora, Archivematica, Dataverse,⁶ and iRODS, and a custom-built repository. All are in active development and/or in “beta” phase of implementation.

To assess the use of repositories for data, the survey asked for estimates of the number of researchers currently depositing datasets in the archives (Q23), the typical sources of archived data (Q25), and total deposit size (Q26). Twenty-eight of those with IRs reported that zero to 1000+ researchers have deposited data (a median of 10 and an average of about 91 researchers). Four of those with data archives reported that between two and 100 researchers have deposited data. Twenty-two of the respondents (66%) reported that data deposits are in the gigabyte range; all but three are under 100GB. Eleven others reported total deposits between 1 and 75 terabytes. Follow-up with respondents might yield more precise numbers

and distinguish among archives with single large and many small deposits. Clearly, however, these are early days for both data-specific archives and IRs with data, and possibly also for researcher’s awareness and adoption of these archiving options.

Data files in both IRs and data archives are coming from a range of sources. Most of the respondents report that datasets are associated with particular publications (88%), full research projects (85%), and graduate theses/dissertations (80%). Twenty-five (63%) report that data was moved from another archive to the library. As data-specific archives expand in use, there may be shifts in data sources that institutional repositories cannot accommodate as well.

The survey also asked about data deposit options. Institutional repositories are generally set up for self-deposit by researchers, and 23 of the IRs with data (65%) do allow data deposits without direct assistance (Q28). However, all but one of these also provides assistance, and 19 say they will deposit data collections for their researchers. Three data archives allow researchers to self-deposit, and they also provide assistance and will deposit data for the researcher. A trend to follow is whether data archive software and support models become more “self-service” for researchers or remain a staff-mediated service.

The final set of data archiving questions addressed details on their architecture for access and preservation. The survey responses show that open access is the policy and intention for all but three of the libraries with archiving solutions, as one would assume based on the literature and public funder requirements. Six of the IRs and data archives also allow controlled access, such as administrative or researcher approval to access data. For data archives in particular, the type of access may be a technical issue, not just policy. Datasets for two institutions are essentially “dark archives” for preservation without a public interface as a direct component of the system, and at least one archive does not currently have the capacity to control access.

Another feature generally considered essential to data archives is support for persistent identifiers so that datasets can be located long-term and reliably cited in publications (including, in some cases, citing particular versions of collections updated with new

data). All but one of the libraries with archives use persistent identifiers. Most common for IRs are the Handle System (21, or 64%). Most of the data archives use DOIs for datasets. Only eight archives use ARKs. Several generate their own identifiers.

Finally, the survey asked about preservation capacities of archives, choosing standards most typical for digital repositories. All five data archives provide file integrity/fixity checking and multiple copy replication, as do 27 IRs (84%). Four data archives practice geographic separation of backups, but this is less common for IRs (22, or 69%). Three data archives and half of the IRs also provide format migration and conversion for data files over time.

Overall responses about archive architecture reflect the different purposes of publication-oriented IR platforms and archives that focus on the specific needs of large and diverse research data collections. If data archiving platforms become less resource intensive for libraries to install and operate, future surveys could gauge corresponding attitudes toward making the library a center for data archiving. Here, the traditional cultural emphasis of libraries and archives on long-term preservation and curation may support libraries' justification for taking on their operation. By contrast, science domains and academic publishers may emphasize shorter-term requirements of data dissemination, yet may hesitate to invest in archiving infrastructure. As interest grows in meeting public funder requirements, and as research practices shift toward data sharing for accelerated discovery and collaboration, academic institutions may recognize libraries as facilitators for research data. Survey responses suggest that infrastructure requirements are significant, and implementation and adoption may be slow. Data archiving by academic libraries, however, is clearly an emerging field that future surveys and case studies should follow.

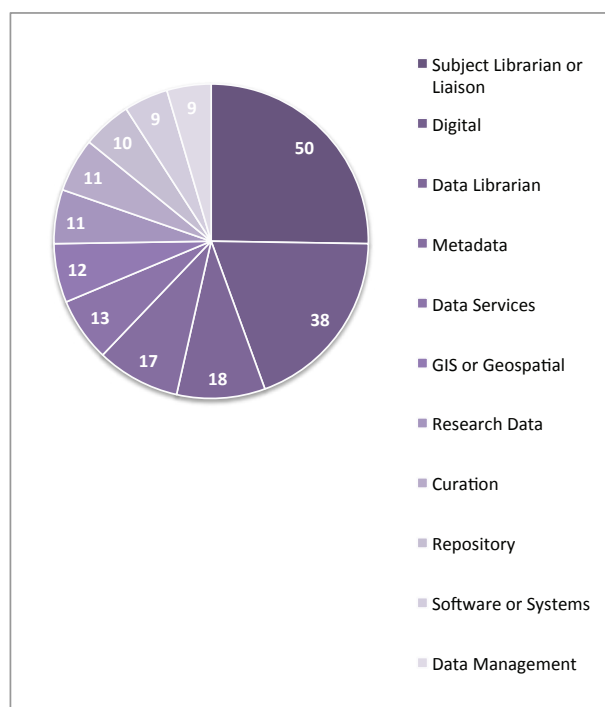
It will be increasingly relevant to follow up on libraries' forays into data archiving, since 30 institutions indicated they plan to offer data archiving within two years (Q56), but nearly a quarter of the respondents included providing data archiving in their top three challenges for RDM services (Q54), most commenting on the difficulty in setting up infrastructure, from software to storage requirements.

RDM Service Staffing

The depth and range of RDM services that libraries offer are, of course, directly proportional to staffing, both in the number of positions and the amount of time given to RDM activities when a position has other responsibilities. Currently, the most prevalent organizational structure for providing RDM services at the 53 responding libraries is a committee of staff from departments within the library (27, or 51%)(Q32). Less common organizational structures include a committee/group comprised of staff from across the university, including the library (9, or 17%), a single position within the library (8, or 15%), and a single department within the library (6, or 11%).

At the libraries where RDM services are provided by staff from different departments, no single department dominates (Q33). About a quarter of the departments provide reference/liaison services, followed by work in collections (19%), digital services (12%), research/instruction (12%), and systems/IT (11%).

The range in RDMS position titles reported shows that staff expertise is diverse and that no one type of position dominates either (Q38).



Out of 231 positions, the most frequently reported title is subject liaison/librarian (50 positions). One might expect RDMS roles to be more prevalent for liaisons to

the hard sciences, but subject specialists for all academic disciplines were represented nearly equally. The next most common position titles include the words “digital” (38 positions), “data librarian” (18 positions), or “metadata” (17 positions). Given that these positions come from across the library, it is not surprising that the majority (146, or 63%) have RDM activities added to their existing job duties, though 49 are new RDMS positions (22%). Only 34 positions were substantially redesigned (15%) to focus on RDM services (Q39).

The survey asked about these positions’ responsibility for three broad RDM service roles: data management plan support, RDM guidance other than DMPs, and data archiving assistance. All but a few of the 53 libraries provide some level of service for all three categories (Q42). RDM guidance is the most common activity among both libraries (51, or 96%) and positions (183, or 82%). Data archiving and DMP support for grants follow closely (49 libraries, 154 and 150 positions, respectively). The majority of positions perform all three roles. In addition to providing RDM service, these positions devote a significant portion of their time to subject reference services (42%), cataloging/ collection development services (27%), administrative tasks (25%), and other data services (24%), among many other duties (Q44).

How much staffing is required to provide RDM service so comprehensively? The number of positions and time spent on RDM activities varies widely across the responding libraries. Groups made up of staff from both the library and other departments in the institution have 10 members on average (Q34). Library committees/groups average about 8 members (Q33). Single library departments that provide RDM services average about 6 staff (Q36). Eight libraries have a single position that provides RDM services.

Although 90% of the positions are full-time, most of these individuals spend only a portion of their time on RDM activities. Respondents were asked to estimate the percentage of time spent on RDM for up to six positions within their institutions (Q43). Only 27 of the 213 reported positions (13%) spend 100% of their time on RDM services. Twelve spend between 60% and 90%, and 17 spend half their time on RDM activities. The remaining 147 positions (69%) spend

less than half their time on RDM activities; nearly a quarter (47) spends less than 10%.

Education and Skills

In addition to assessing staffing models, the survey explored the training and educational backgrounds of current RDM service providers. We expected this emerging service area to require new and diverse skills, not always found within library service environments, and which often draw upon experiences from other professions or disciplines. The majority of staff with RDMS roles have MLS/MLIS degrees (172 individuals, or 75%) (Q45). Thirty-eight of these have masters degrees in other disciplines (22%), and six have PhDs in other disciplines (3%). Forty-three individuals (19%) hold only graduate degrees in a discipline other than library science; 20 hold a masters, 17 hold a PhD, and six hold both. Fourteen individuals have a degree with a data curation emphasis. Eight others have degrees with an archives emphasis. The range of other disciplines is tremendous and includes very minimal concentration in any single area. Only ten individuals have only an undergraduate degree.

Recognizing that formal academic backgrounds may not be the only influential factor in building a qualified RDM service team, the survey also asked respondents to select the three most important skill areas for the staff now in these positions (Q46). The top three areas are subject domain expertise (38 responses, or 75%), digital/data curation training (31, or 60%), and IT experience (30, or 59%). Respondents also explicitly noted the importance of training and experience in research methods, data analysis, research data management best practices, and scholarly communication.

Training Needs

Although many respondents stated that the function of libraries (access and preservation of knowledge) and current skill set of librarians lends themselves to RDM, many acknowledged that new skills and training are needed to optimally perform RDM. For future planning purposes, the survey asked which skills RDM staff most need further advanced training on. The most frequent response was identification and

application of appropriate metadata standards (25, or 49%). A second cluster includes digital preservation (17, or 33%), data ownership policies (16, or 31%), ethical and legal issues (15, or 29%), and subject domain expertise (15, or 29%). Some respondents also noted the need for deeper technical skills in related areas such as data acquisition, wrangling, analysis, interpretation, visualization, and deeper knowledge of research administration practices and forces. A few also acknowledged that their services were not yet developed enough to know what they would need next.

The survey next looked at the training methods that libraries have used to develop their RDM staff. Perhaps predictably, workshop attendance (48 responses, or 92%), conference attendance (44, or 85%), independent study (35, or 67%), and training provided by professional organizations (32, or 62%) rose to the top.⁷ With more iSchools and MLIS programs offering data curation and digital collection emphases, and other fields of science and information technology emphasizing training in big data and digital data support, it will be interesting to follow trends in educational background of those entering the emerging library specialty of RDMS.

Funding RDM Services

Many survey respondents identified RDM service funding as a key challenge. This is not surprising since all but one library covers the costs through the regular library budget (Q49). Only a few have received external grants or a portion of research project funds. Only three have tapped endowment funds. Expectations for additional funding to support RDM services in the future don't indicate much change. Slightly more than half of the respondents to Q61 indicated that additional sources of funding have not yet been determined. The most frequently anticipated future funding sources are the regular library budget (21, or 36%) and external grant funding (15, or 26%). Some libraries expect to tap a temporary or special project budget, or receive funding from the parent institution. At the same time, 66% of survey respondents expect the allocation of funds for RDM services will increase in the next three years (Q62).

Partnerships

Building RDM services involves collaboration within the library, across a campus, and sometimes across institutions. Respondents' institutional models and levels of service development have varied widely for all the components of RDM services discussed thus far. Similarly, responses about which departments library RDMS staff refer researchers to reveal wide diversity in the degree to which these units interact. Survey respondents most frequently refer researchers to central IT and research administration, units with whom they also frequently collaborate on projects (Q52). Referrals are also directed to IRB and general counsel, but only a few libraries collaborate with these departments. A smaller number of respondents both make referrals to and collaborate with institutional administration, institutional archives, and other units ranging from statistical consulting groups to department or school IT units.

Although inter-institutional partnerships for service provision are happening throughout ARL libraries, the number of formal collaborations is still relatively small. Only 13 respondents (26%) have participated in an external partnership (Q53). Several of these were joint developers of the DMPTool. Other partnerships center on software and tool development, creation of training materials, and research for service provision.

Conclusion

Given the current technical and political environment, we (and most other followers of the research data curation field) predict that the need for universities to manage their research data for both access and preservation will grow, due primarily to two factors: the reliance in many fields of science upon technical ability to create large and complex digital data, and the increasing requirements and enforcement of data sharing policies by research funders. As the results from this survey show, at least 54 ARL libraries are responding to this need by providing RDM service in some capacity, and another 17 have plans to do so in the next few years. The majority of these libraries provide service across all three RDM activities: data

management plans for grants, guidance on data management, and research data archiving. Half draw on staff from different departments within the library to deliver these services.

A common theme throughout the survey is the recognition that, in order to provide comprehensive RDM services and to support scientists throughout the data lifecycle, libraries need to collaborate, either formally or informally, with other units at the institution. This is true even where the library has a dedicated department for RDM. At the very least, these units will then be aware of the services the library offers and can refer researchers to them (e.g., research administrators can send proposal writers to the library for data management plans). Forming these partnerships is listed as the biggest challenge by respondents, and in some cases has led to uncertain roles at the institution-level over which units have primacy over RDM. With so many aspects of RDM services overlapping domains and defining new territories of collaboration among multiple units within an institution, it will be interesting to follow how libraries continue to position their roles within the school as such services expand, and how unified or diverse an approach an institution may choose to take overall in supporting research data management.

It costs money to provide quality services. RDMS requires a diverse range of skills, many outside the typical expertise of library staff and not all libraries can afford to hire new and/or retrain staff. Creating archiving infrastructure and curating research data are also expensive endeavors. Right now, the amount of archived data is relatively small for the majority of institutions; however, as funders become more stringent in data retention and sharing requirements, libraries will need to employ an alternative funding model, such as through fees to researchers or their grant projects.

Finally, the third biggest challenge reported is faculty (non)engagement due to a lack of awareness of services that the library provides, low perceived value of services, and resistance to data sharing. Respondents stated that the most effective marketing techniques were through workshops and presentation to researchers, referrals from research project (grants) administration, and direct emails to researchers (Q50).

The low perceived value of services is a more difficult issue to overcome. Some researchers do not view the library as a resource for data management, and as one respondent commented, “preservation and data sharing are a hard sell when the researcher only need[s] to write a plan.” Also, some respondents noted that funders are not taking mandates seriously, giving the researcher a low incentive to care about data management and sharing. As noted, the push for open data access for publicly funded research and compliance for data sharing policies, both in North America and around the world, may change the environment significantly. Our survey results suggest that many ARL libraries have at least a start toward growing services to meet an increasing demand.

This survey provides a snapshot of what RDM activities ARL libraries are currently involved in, what human resources are being used to provide these services, and projected service provision. Although providing RDM services is not easy and requires a heavy investment in hiring/retraining staff, building technical infrastructure, and continually reaching out to and collaborating with other data management players on campus, many respondents felt that library could and should support RDM activities to some degree (Q66). Although RMD services are relatively new, institutions are taking diverse approaches to providing them, and will likely evolve over the next few years. The exact nature of how service will be provided will likely depend on institutional and funder policy, technical skills of library staff, and the financial position of a library.

Limitations of Survey

The survey responses clearly show that RDM services do not happen just in the library. They require pan-institutional collaboration. One major limitation of this survey is the absence of responses from those outside the library. If RDM services go beyond the library but are unknown by the library, the answers in this survey may not be a complete picture of the institutions’ RDM services. One particular library commented: “We don’t know the answers to any of these and don’t want to speak for other units.”

The survey underwent review and testing before it was sent out. We adjusted many definitions and

questions, but that did not eliminate problems with respondents misinterpreting concepts and definitions. Several of the metrics of service, such as the extent of data management plan assistance and archive use, were particularly difficult to define and ask in ways that yielded precise responses. We recommend that further studies involve case studies, and focus on particular topics such as archiving or staffing, since we presented our respondents with a particularly long and complicated survey. We greatly appreciate their efforts to complete our survey, and hope these results will be a useful benchmark and basis for inspiration in this new and expanding field of research library service.

Acknowledgements

We would like to sincerely thank the following individuals and groups for reviewing our survey instrument: Andrea Denton, Mike Furlough, Brian Gunia, Patricia Hwse, David Lowe, Karl Nilsen, Susan Payne, Lizzy Rolando, Jennifer Ward, Jonathan Wheeler, Lynda White, Stephanie Wright, the Canada Institute for Scientific and Technical Information (CISTI), and the E-Research Working Group. We would also like to thank the authors of the e-science report for providing helpful background and context for the report and for advising us what we ought to consider including in the RDM services survey. Tim Dilauro, Johns Hopkins University, provided his perspective and expertise in developing the survey instrument. Finally, we'd like to thank Wendy Mann of George Mason University for arranging a space at the Fenwick Library for the University of Virginia and Johns Hopkins University authors to hold in-person meetings.

Endnotes

- 1 For a sample see the references in the Selected Resources section of this SPEC Kit.
- 2 <http://www.arl.org/storage/documents/publications/escience-report-2010.pdf>
- 3 See heading "Key Papers in the Development of RDMS" in the Selected Resources section of this SPEC Kit.
- 4 <http://www.arl.org/focus-areas/e-research/e-science-institute>
- 5 See [White House Office of Science and Technology Policy \(OSTP\)](#) Feb. 22, 2013 memorandum on open access to funded research data and publications.
- 6 Dataverse is in relatively widespread use as a repository for specific disciplines and research centers, but only one library reported being directly involved with a Dataverse implementation.
- 7 Attendance at an E-Science Institute workshop noted earlier in responses to Question 3 was another influential training resources for many respondents.

