



## Tiers of Research Data Support Services

Rebecca C. Reznik-Zellen, Jessica Adamick, Stephen McGinty

University of Massachusetts Amherst, Amherst, MA, USA

### Abstract

**Objective:** To describe three tiers of research data support services that emerged from national environmental scanning of data management needs and activities.

**Setting:** The University of Massachusetts Amherst (UMass Amherst) is a top fifty recipient of NSF funding, with the agency providing over 40% of the University's sponsored research, and is classified as a Research University with Very High research activity by the Carnegie Foundation. After determining a need for data management services, a library Data Working Group performed national environmental scanning. Environmental scanning used public information available from 18 UMass Amherst peer and model institutions to determine the range of data management and curation services that are available to various research communities.

**Methods:** Environmental scanning activities include a web audit.

**Results:** National practices demonstrate a wide range of potential data management services. UMass Amherst's Data Working Group has generalized data management services into three tiers, creating a useful rubric for determining one's current service level and for setting goals to meet the needs of one's research community.

**Conclusions:** The Tiers of Research Data Support Services, generalized from local needs and national activities, describe different levels of support of increasing cost and involvement scales for supporting researchers' data management and curation needs: education, consultation, and infrastructure.

### Introduction

Academic libraries have shifted some focus to the management of research and teaching output, demonstrated in part by an increasing involvement with research data. Over the past decade, the federal government, funding agencies, and others invested in the research enterprise have formalized data management practices and recommendations. Major data management policy milestones include a National Institutes of Health data sharing plan requirement (NIH 2003), a National Science Foundation (NSF) Report on cyberinfrastructure (Atkins 2003), an NSF

data management plan requirement (NSF 2011), and an Office of Digital Humanities of the National Endowment for the Humanities data management plan requirement (NEH 2011). Several agency reports and special publications have highlighted the growing importance of data management. These include a National Science and Technology Council (NSTC 2009) strategic plan for the Federal government to facilitate the development of preservation and access methods for data, focusing on specific components of data management plans, and a 2009 National Academy of Science report advocating for open access research data and data man-

**Correspondence to** Jessica Adamick: jadamick@library.umass.edu

**Keywords:** research data, library services, tiers of service, NSF data management, University of Massachusetts Amherst, education, consultation, infrastructure

agement training for all researchers (Committee on Science, Engineering and Public Policy 2009). The concerns described by these agency reports made their way into the broader community; *Nature*, *The Economist*, and *Science* all published special issues on the proliferation and the management of scientific data in 2008 and 2009, 2010, and 2011, respectively.

Academic libraries recognize their role in data management. The Association of Research Libraries/Digital Library Federation 2011-2012 eScience Institute demonstrates library interest; 72 of 126 members have enrolled as sponsors and supporters of the institute, which is designed to help individuals develop their eScience support roles (ARL 2011). While libraries have been articulating their role in data management for at least half a decade (Friedlander 2006), the National Science Foundation data management plan requirement has catalyzed the development of library data services.

For an institution like the University of Massachusetts Amherst (UMass Amherst), a top fifty recipient of NSF funding which is classified as a Research University with Very High research activity by the Carnegie Foundation, the need for the development of library data services is critical. To address data management needs from a library perspective and to develop services for faculty and graduate students, the UMass Amherst Libraries formed a Data Working Group (DWG). The DWG was charged with determining if the University Libraries should accept broad responsibility for curating research data and, if so, how that should be done, what would be expected, and who would be involved. Through exploratory interviews with faculty and focus groups with graduate students, the DWG found that many researchers' data management strategies vary widely up until the point of publication and do not typically support preservation or sharing, which is the objective of the NSF mandate (University of Massachusetts, 2010). These efforts identified a clear need

for the development of local data management services and support.

To understand the current research data support environment more clearly, the DWG conducted national research on data services offered by peer and model institutions. In collating the findings of these activities, the DWG has identified and described three tiers of research data support: education, consultation, and infrastructure.

### Literature Review

ERIC, LITA, and Google Scholar were searched for research on data support services in higher education, with particular attention to the role of libraries. The journal literature is intermittent on this issue. Publications generally fall into two categories: case studies at individual universities and essays on the response of higher education as a whole to the growing importance of data management.

The case studies describe librarians' efforts to gain an understanding of researcher data management practices. Librarians are aware that different disciplines have different needs. By delineating the differences among the various disciplines, they bring that perspective to the construction of data management services. Witt and colleagues (Witt, et al. 2009, 95) interviewed 19 faculty members at two large research institutions in order to inform librarian practice, developing "curation profiles" within various domains. The profiles include information on all aspects of data management; typical data formats and file sizes, intellectual property, data description, and interoperability are some examples. A similar study was conducted at the University of Colorado, where librarians conducted structured interviews with twenty-six researchers from nine departments within the sciences (Lage et al 2011, 918). They developed profiles, or "personas" of a typical array of researchers, and offer suggestions for relationship building.

The University of Otago, New Zealand, analyzed questionnaire data from 71 researchers in an effort to learn more about their data management practices and to determine their level of interest in services (Eliot 2008, 6). The author found considerable interest in data curation. The majority of researchers manage their own data and apply their own metadata, as cooperation and standardization are minimal. Similarly, at Oxford University and the University of Edinburgh, McDonald and Martinez-Urbe found that strategies for effective data management, particularly in the context of utilizing data repositories, require a range of skills including information management, computing, and social dynamics, and that participation from various institutional entities beyond the researcher is necessary (McDonald and Martinez-Urbe 2010, 5).

The articles that review the responses of higher education to data management practices define data management and curation and focus on establishing infrastructure for handling research data. Much of the discussion examines the scope of the issue and the appropriate roles for different campus entities. A prominent example in the latter category is Friedlander's report on a 2006 Association of Research Libraries workshop that explored new partnerships, infrastructure development, and sustainable economic models for data management (Friedlander 2006). Five years before the NSF mandate, the ARL emphasized "that digital data stewardship is fundamental to the future of scientific and engineering research and the education enterprise, and hence to innovation and competitiveness." Similarly, Hey and Hey describe a "data deluge" about to impact the scholarly community: "One of the key drivers underpinning the e-Science movement is the imminent availability of large amounts of data arising from the new generations of scientific experiments and surveys. New high-throughput experimental devices are now being deployed in many fields of science - from astronomy to biology - and this will lead to a veritable deluge of

scientific data over the next five years or so" (Hey and Hey 2006, 519). EDUCAUSE has conducted one of the few multi-university examinations of data management, based on a quantitative survey, qualitative interviews and two case studies, with the participation of 309 EDUCAUSE members (Yanosky 2009). Although they concluded that campuses are not "drowning" in data, they voice a concern that declining budgets will have serious repercussions on the variety of research data in the near future.

Discussion of library roles in data curation was initiated in 2007 by Anna Gold in the second part of her two-part article on Cyberinfrastructure, Data, and Libraries (Gold 2007) and taken up again in 2010 when she presents the developments within the library community that have been shaping library roles in data management since 2006 (Gold 2010). In the latter work, Gold articulates tiers of digital data curation support at a national level: national infrastructure, campus infrastructure, and professional development and education. Also in 2010, an ARL survey of member institutions' data support services was released. The authors report that "approximately seventy-three per cent of the respondents (29 of 40) indicated the library was involved in e-science at their institutions. Most of the services are limited to consultation and reference" (Soehner 2010, 8). Some examples of more recent discussion of library involvement in data management include Heidorn's description of the current landscape of participants and roles in data curation. He embeds this function squarely within the mission of the library: "Libraries have the organizational culture and the mandate that would allow these data to be properly curated over long periods of time" (Heidorn 2011, 664). In an article that chronicles the change of perception among scientists—where their focus has changed from seeing a "data bottleneck" to seeing a "data deluge"—Baraniuk identifies many opportunities for data managers and libraries in this new scenario (Baraniuk 2011, 717).

**Table 1:** Web Audit Evaluation Areas

Infrastructure	Services	Organization	Marketing
Institutional repository	Education	Dedicated jobs	Promotion of services
Data repository	Consulting, including grant support and mediation	Departmental structure	Publications
Data storage	Metadata	On-campus collaborations	
	Informational web pages	Off-campus collaborations	

This article contributes to this body of work with a description of the variety of library data management services currently being provided by a sample of academic libraries across the nation, categorized by degrees of researcher involvement and increasing levels of institutional commitment.

**Methods**

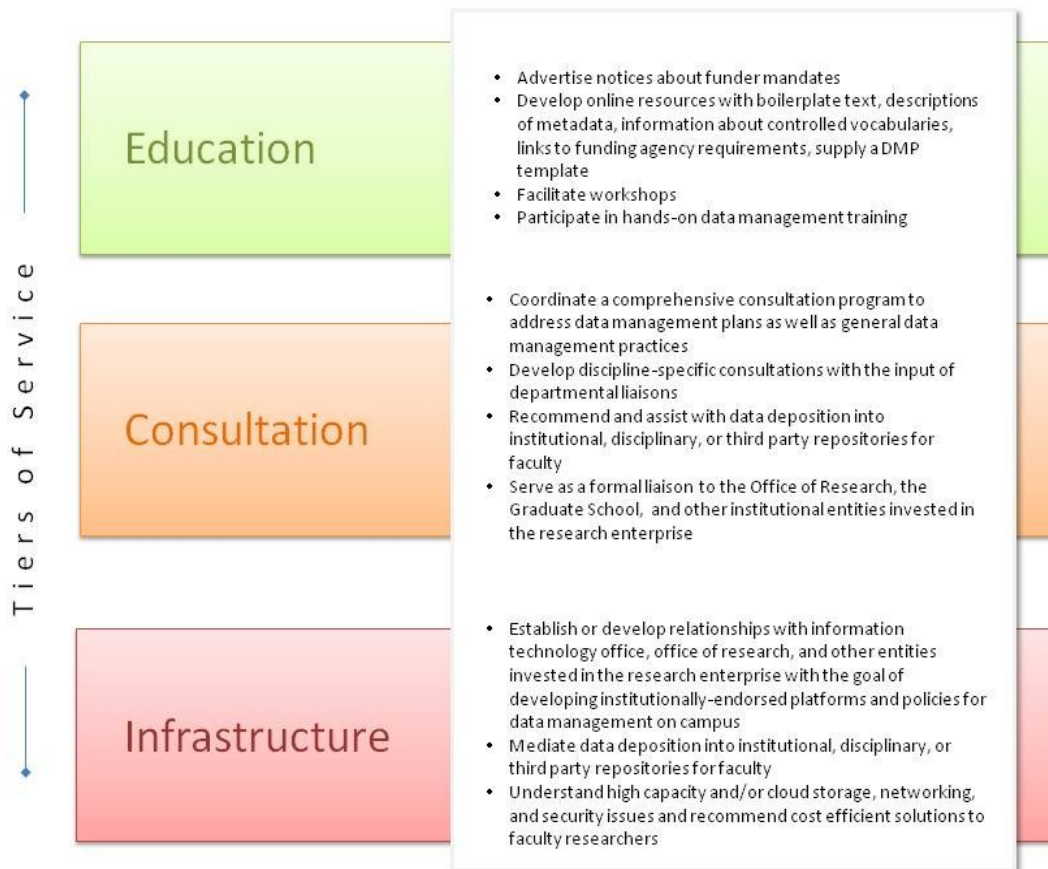
Wanting to make informed recommendations for the development of data management services on campus, the DWG looked nation-wide for examples of library data services. A web audit/environmental scan of peer and model institutions was performed during spring 2011 to survey library approaches to data management at other universities. A list of peer institutions that is used for evaluation at a university-wide level was surveyed. These institutions include Indiana University Bloomington, Iowa State University, Rutgers University, SUNY Stony Brook, University of California Santa Barbara, University of Colorado Boulder, University of Connecticut, University of Delaware, University of Maryland, and University of Nebraska Lincoln. Institutions that have set a national example for supporting the management and curation of research data have been chosen as model institutions for this audit. They include: Purdue University, Massachusetts

Institute of Technology, University of Minnesota, Johns Hopkins University, Cornell University, Oregon State University, University of Virginia, and the University of Wisconsin-Madison. The DWG inventoried numerous library data services using public information available through the peer and model institution web sites. The institutions were audited in the following categories: infrastructure, services, organization, and marketing (Table 1).

**Findings**

As expected, the audited model institutions offer more varied and sophisticated services than do many, but not all, of the audited peer intuitions. Peer institutions demonstrate a wide range of services offered and tremendous variety in the degree to which data management had a high or low profile on their respective library web sites. Generally speaking, each institution has different resources and campus needs to address, and this was reflected in the type and amount of services offered by the libraries. For example, three of ten peer institutions audited appear to be actively involved in data management and curation on their campuses. Two of these have organizational structures in place to grapple with data curation and management and utilize their institutional reposi-

**Figure 1: Tiers of Service**



ties to house data; two also have (or are developing) dedicated infrastructure to house and preserve research data and provide services based on that infrastructure. The remaining seven institutions demonstrated comparatively lower levels of service activity at the time of the audit, but did collect and disseminate information, minimally about the NSF mandate, and more robustly about data management best practices. The model institutions demonstrate a sophisticated array of services and infrastructure and organizational models for data management. The model institutions uniformly have staff dedicated to data in some capacity and seven of eight have relationships or partnerships with other entities to support the research enterprise on their campuses.

Categories of service levels naturally emerged by examining the variety of activities at peer and model institutions. These service levels can encompass many different activities, but are distinguished from one another based on the degree of involvement with researchers and their data: providing information to faculty; interacting with faculty on a one-to-one basis; and taking stewardship of faculty research data (Figure 1). The DWG created three tiers of service based on these distinctions and inferred degrees of financial and staffing support: education, consultation, and infrastructure. The tiers described in detail below refine levels of activity and resources required for an infrastructure for the campus-based data curation services presented by Gold (2010). Librar-

ies that are building research data support services may use the tiers as a rubric for determining their current service level and for setting goals to meet the needs of their research community that are consistent with their institutional mission and environment.

## **Tiers of Service**

### ***Tier One Level of Service—Education***

*Libraries educate their communities about data management.*

Even libraries that appear to be doing relatively little to support the research data infrastructure at their institutions still engage in some manner of education. The most basic level of information provided to campus communities are notices about the NSF mandate. In addition, some libraries host a Lib-Guide or a set of web pages that contain a variety of current “how-to” information. A resource site may include boilerplate text, descriptions of metadata, information about controlled vocabularies, file naming conventions, up-to-date links to funding agencies, and a data management plan template. The most comprehensive educational information includes pointers to tools and services on campus for faculty to help in the management of their research data and tutorials or workshops on data management basics. Examples of libraries with well-constructed and thorough data management education packages are the University of Nebraska, Massachusetts Institute of Technology, and the University of Minnesota.

In the education tier, there is an opportunity to develop a closer relationship with campus offices of research. For example, the library could work with research administrators to keep the campus up to date on funder policies through workshop series. Research administrators could refer principal investigators to the library liaisons, who could provide discipline-specific guidance.

While the primary target of educational resources and workshops is the campus com-

munity, hands-on data management training should also be directed toward library staff in order to build the necessary in-house expertise to continue developing useful educational resources. Investing in the acquisition of in-house expertise is a critical component for providing solid educational data services to a university. The maintenance of comprehensive educational materials and staff data management training both require resources in the shape of a formal library group/committee or dedicated staff time and a strong commitment to professional development to support them. Providing education is a low-investment strategy, but it offers limited opportunities for formal engagement with the campus community. While educating is a traditional role for libraries, education in data management expands the boundary of traditional library services by targeting researchers at an earlier stage of the research process.

### ***Tier Two Level of Service—Consultation***

*Libraries consult with faculty and researchers on a variety of issues relevant to the management of research data.*

The most accessible and common point of consultation involves funder requirements and the execution of data management plans. At this tier, libraries provide individual or group consultations for faculty and graduate students to review and enhance proposed data management plans. Libraries also provide metadata services to researchers in the form of online or in-person tutorials and free or fee-based metadata consultation. Libraries also offer to identify and assist with the deposition of material into data repositories at this service level. Discipline-specific metadata consultations and repository identification require the involvement of subject specialists. More developed consultation services come in the form of dedicated offices or centers of activity around digital scholarship. These offices will frequently be staffed and/or supported by cross-institutional entities, such as an office of research, and will be focused on the entire re-

search enterprise, from data collection to publication to preservation. Examples of libraries with well-developed consultation services include Johns Hopkins University and the University of Wisconsin-Madison.

Much of the activities in the consultative tier are oriented toward a campus community, though librarians would benefit from trainings and consultation services to the extent that they interact with researchers with data management needs or are conducting research themselves. A comprehensive consultation program would be able to meet researchers' needs regardless of where they are in the research cycle and provide direction for hardware and software choices, data description and metadata standards, data storage and backup scenarios, data management training for graduate students, provisions for data access and sharing, data publication and citation, and data archiving and preservation. There is also the opportunity to expand a liaison program to create a formal liaison to an office of research, graduate school, and any other institutional entities invested in the research enterprise. To provide full-spectrum consultation services, dedicated staff hours are required to develop expertise in the range of competencies that support data management. Consultation is a mid-range investment that requires the commitment of librarians and library administration. It offers opportunities for campus engagement and expands the battery of library services in a relevant and timely manner.

### ***Tier Three Level of Service—Infrastructure***

*Libraries provide infrastructure for data management and data curation to their campus communities.*

Though most audited institutions link to infrastructure services provided by other campus entities or third party providers, there are a few different local infrastructure scenarios demonstrated. These include data staging platforms for active data sets (Cornell) to robust repositories for publication and long-

term storage of completed data sets (Rutgers).

Campus-wide infrastructure solutions require a campus-wide conversation. It is a large-scale problem for the entire community that involves many stakeholders. This level of commitment to infrastructure requires minimally the ability to store, secure, backup, describe, provide access to, and preserve research data of different kinds. Collaborative approaches to infrastructure can provide large-scale solutions and meet multiple needs, but not without the investment of significant time and resources from established entities such as an information technology office, an office of research, the library, and others. The libraries can play an effective role here as the coordinators of infrastructure providers and/or managers of hardware platforms for data management. At this level of service, a library's ability to contribute relies on the development of in-house expertise in data management through professional development support. Providing technical infrastructure is a high-investment, long-term strategy that requires the support of the library and other entities on campus that are invested in the research enterprise. While providing infrastructure might expand a library's role on campus, it requires extensive scoping and financial support before it can be implemented.

### **Conclusion**

After completing national research on current library data management services, the UMass Amherst Libraries Data Working Group identified and articulated tiers of support service for library data management. Libraries can participate in data management and offer meaningful services to their researchers at various levels: education, consulting, and infrastructure. These tiers create a useful rubric for determining one's current service level and for setting goals to meet the needs of one's research community.



## References

Association of Research Libraries. Association of Research Libraries, accessed December 19, 2011. <http://www.arl.org/rtl/eresearch/escien/escieninstitute/esciensponsors.shtml>.

Atkins, Daniel E. and National Science Foundation (U.S.). Blue-Ribbon Advisory Panel on Cyberinfrastructure. 2003. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Arlington, VA: National Science Foundation.

Baraniuk, Richard G. 2011. "More is Less: Signal Processing and the Data Deluge." *Science* 331 (6018): 717-719.

Committee on Science, Engineering, and Public Policy (U.S.). Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. 2009. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. Washington, D.C.: National Academies Press.

Elliot, Gillian. 2008. *Otago Biodiversity Data Management Project Report Part 1: Questionnaire Report*. New Zealand: University of Otago Library. Accessed on December 22, 2011. [http://www.library.otago.ac.nz/pdf/2008\\_OBDMP\\_questionnaire.pdf](http://www.library.otago.ac.nz/pdf/2008_OBDMP_questionnaire.pdf).

Friedlander, Amy, and Prudence Adler. 2006. *To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering*. ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe (2006: Arlington, VA). Washington, DC: Association of Research Libraries.

Gold, Anna. 2007. Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries. *D-Lib Magazine* 13 (9/10). Accessed on December 21, 2011. <http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>.

[september07/gold/09gold-pt2.html](http://www.dlib.org/dlib/september07/gold/09gold-pt2.html).

Gold, Anna. 2010. Data Curation and Libraries: Short Term Developments, Long-Term Prospects. Prepared for a 2009 talk at AGU on "Data and Libraries". Accessed on December 21, 2011. Available from: [http://digitalcommons.calpoly.edu/lib\\_dean/27](http://digitalcommons.calpoly.edu/lib_dean/27).

Heidorn, P. B. 2011. "The Emerging Role of Libraries in Data Curation and E-Science." *Journal of Library Administration* 51 (7): 662-672.

Hey, Tony and Jessie Hey. 2006. "E-Science and its Implications for the Library Community." *Library Hi Tech* 24 (4): 515-528.

Lage, Kathryn, Barbara Losoff, and Jack Maness. 2011. Receptivity to Library Involvement in Scientific Data Curation: A Case Study at the University of Colorado Boulder. *Libraries and the Academy* 11 (4): 915-937.

Macdonald, Stuart and Luis Martinez-Urbe. 2010. "Collaboration to Data Curation: Harnessing Institutional Expertise." *New Review of Academic Librarianship* 16: 4-16.

National Endowment for the Humanities (U.S.). "Digital Humanities Implementation Grants." National Endowment for the Humanities, accessed December 14, 2011. <http://www.neh.gov/grants/guidelines/digitalhumanitiesimplementation.html>.

National Institute of Health (U.S.). 2003. *Final H.I.H. Statement on Sharing Research Data*. Press Release. <http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html>.

National Science and Technology Council (U.S.). Committee on Science and National Science and Technology Council (U.S.). Interagency Working Group on Digital Data. 2009. *Harnessing the Power of Digital Data for Science and Society: Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council*. Washington,



D.C.: Interagency Working Group on Digital Data.

National Science Foundation (U.S.). "NSF Data Management Plan Requirements." National Science Foundation. Accessed December 16, 2011. <http://www.nsf.gov/eng/general/dmp.jsp>.

Soehner, Catherine, Catherine Steeves, and Jennifer L. Ward. "E-Science and Data Support Services a Study of ARL Member Institutions." Washington, D.C.: Association of Research Libraries.

University of Massachusetts Amherst Libraries' Data Working Group. 2010. *Data Working Group Report*. Accessed on December 22, 2011. <http://www.library.umass.edu/assets/aboutus/attachments/Report2010-Final.pdf>.

Wilson, James A. J., Michael A. Fraser, Luis Martinez-Uribe, Meriel Patrick, Asif Akram, and Tahir Mansoori. 2010. "Developing Infrastructure for Research Data Management at the University of Oxford." *Ariadne* 65 Accessed on December 22, 2011. <http://www.ariadne.ac.uk/issue65/wilson-et-al/>

Witt, Michael, Jacob Carlson, D. Scott Brandt, and Melissa H. Cragin. 2009. "Constructing Data Curation Profiles." *International Journal of Digital Curation* 4 (3): 93 - 103. Accessed on December 22, 2011. <http://ijdc.net/index.php/ijdc/article/view/137>.

Woods Hole Oceanographic Institute. 2001. *Data Management for Marine Biology and Geophysics, Tools for Archiving, Analysis and Visualization, Workshop Report*. Woods Hole, MA: Woods Hole Oceanographic Institute.

Yanosky, Ronald. 2009. *Institutional Data Management in Higher Education. ECAR Key Findings*. Boulder, CO: EDUCAUSE.

## Acknowledgement

*Tiers of Research Data Support Services* is based on the work of the University of Massachusetts Amherst Libraries Data Working Group: Jessica Adamick, MJ Canavan, Stephen McGinty, Rebecca Reznik-Zellen, Matt Sheridan, and Maxine Schmidt.

*Disclosure:* The authors report no conflicts of interest.

All content in *Journal of eScience Librarianship*, unless otherwise noted, is licensed under a Creative Commons Attribution-Noncommercial-Share Alike License <http://creativecommons.org/licenses/by-nc-sa/3.0/>

ISSN 2161-3974