Chapter 7

Data Management Services in Libraries

Patricia Hswe*

Publishing and Curation Services, University Libraries,
The Pennsylvania State University, W-311 Pattee Library,
University Park, Pennsylvania 16802
*E-mail: phswe@psu.edu

Activated in January 2011, the National Science Foundation requirement to include data management plans with grant proposal submissions has compelled many academic libraries to revisit their service offerings, particularly in the context of assisting researchers with managing their data in efficient, productive ways. This chapter covers the status of data management services in academic libraries in the years since the NSF requirement went into effect. It explores the reasons why data need to be managed, and why librarians and libraries are up to the task; describes the steps that various libraries have taken to determine service, staffing, and infrastructure requirements; provides a brief overview of current data management service offerings in libraries; and, in conclusion, touches on the emergence of new librarian roles, including postdoctoral positions, that have arisen to meet demands in data management and data curation in academic libraries.

Introduction

It’s the 21st century - do you know where your data are?

Research data exemplify the output of experimentation. As such, data can be considered “the recorded factual material commonly accepted in the scientific community as necessary to validate research findings” (1). When the National Science Foundation (NSF) announced in summer 2010 that, effective
January 18, 2011, data management plans (DMPs) would be required as part of grant application submissions to the agency, it may as well have been asking the question above. Data also tell a story, one that the NSF - which applies taxpayer dollars toward an estimated 20% of all research funded by the U.S. government - wants to ensure that scientists are recounting as fully as possible. The general guidance on DMPs provided by the NSF is intended to orient researchers with the essentials of managing data, even prior to their generation, when methodologies and other approaches are being determined, through their collection and continuing into their life cycle of use and relevance, with the overall aim of sharing data and making them publicly accessible. To this end, in particular for the directorates that have not issued more specific DMP guidelines, such as the Computer & Information Science & Engineering Directorate and the Division of Material Research (in the Directorate of Mathematical and Physical Sciences), the NSF outlines the components that make up a plan, which amount to five sections:

- a description of the types of data;
- how the data will be documented (i.e., metadata standards to be used);
- policies for sharing the data and providing access to them (also accounting for any confidentiality, privacy, or security protections);
- guidance and recommendations for reuse and redistribution of the data, as well as creation of derivatives from them;
- how the data will be archived, to enable preservation for continuing access.

These are the plot points necessary for telling the story of research data.

Although library literature has long acknowledged the “data deluge,” and calls for librarians to acquire skills in data curation and data management were appearing before the NSF announcement, this recent mandate has many academic libraries, whether large or small, marshaling resources toward new service models, new infrastructure, and new librarian roles (2–10). In addition, in the short time following this new implementation of the NSF’s data sharing policy, more library and information science degree programs are integrating courses that address research data management issues, including short-term, “boot camp” courses, or they are including coverage on this topic in already existing classes (11). Fellowships in data curation, with the goal of expanding a workforce of expertise in this area, have also emerged since the NSF went public with its requirement (12).

This chapter covers the status of data management services in academic libraries in the years since the NSF requirement went into effect. It explores the reasons why data need to be managed, and why librarians and libraries are up to the task; describes the steps that various libraries have taken to determine service, staffing, and infrastructure requirements; provides a brief overview of current data management service offerings in libraries; and, in conclusion, touches on the emergence of new librarian roles, including postdoctoral positions, that have arisen to meet demands in data management and data curation in academic libraries.
Why Data Need To Be Managed and the Key Role Librarians and Libraries Play

The story of data management services in libraries cannot be told without considering the subtext of the story first: the unstoppable torrent itself of data. The escalating availability to researchers of rapid and scalable computational methods, afforded by highly efficient and powerful supercomputing processors, has produced an ineluctable embarrassment of data riches. As Hay et al explain, scientists are struggling with a massive amount of data born of many types of sources, including instruments, simulations, and sensor networks (13). As a consequence, not managing data properly has come at great price to many scientists and scholars. Reports of data loss, security breaches, and lack of data authentication are not uncommon and ultimately erode both the sharing of data and access to them. The steepest cost is, likely, impeded progress in scientific research: the more time scientists must spend on management of data, the less time they have to do the research that new data and future findings are dependent upon.

Perhaps most significant, the mismanagement of data diminishes the reproducibility and replicability of data - those verification processes that constitute the gold standard in scientific research. A special issue of Science addressed the topics of data replication and reproducibility, noting an array of challenges and benefits, such as the following: processes observed in the field that call for certain parameters or conditions are difficult to capture and thus replicate; the discipline of computer science needs a set of base norms, or standards, for reproducibility of methods and code, so that researchers can arrive at identical, dependable outcomes with the unprocessed, initial data; and allowing for different groups, such as scientists and public policy makers, to view the same data - and engender new data - can yield rich results (14). This last point resonates with what José Muñoz, Chief Technology Officer (CTO) in the NSF’s Office of Cyberinfrastructure, stated in the May 10, 2010, press release announcing the funding agency’s imminent requirement: “Twenty-first century scientific inquiry will depend in large part on data exploration” (15). Because of the mandate, such data exploration can be enabled across communities more frequently, spurring new findings and - by extension - new stories and new questions.

One of these communities is, arguably, that of academic librarians and libraries. Libraries have long been in the practice of maintaining, as well as disseminating and sharing, the scholarly record, of which research data constitute an intrinsic part. Starting up new library services for data curation presents an opportunity for librarians who are subject experts to collaborate with archivists, who are experienced in the appraisal, selection, description, organization, preservation, and retention of a range of content types, from electronic mail, to scholarly papers, to university business records (16). As Heidorn asserts, “Curation of data is within the libraries’ mission, and libraries are among the only institutions with the capacity to curate many data types” (17). At the same time - while librarians possess the subject expertise - for some, as Heidorn also points out, becoming conversant with data management practices presents a formidable learning curve (17). It necessitates a working knowledge of not only basic
archival approaches but also metadata standards, intellectual property issues, institutional research administration policies and guidelines, and conventions in file management and storage practices.

Working with researchers on management of their data may seemingly cast librarians in roles that are different from before, but, in effect, this interaction coheres well with what librarians do and have always done. For example, liaison librarians - who typically have subject expertise - develop and manage collections driven by the research that engage their faculty and students. They are also trained in conducting reference interviews. As Witt and Carlson suggest, liaison librarians could effectively merge these approaches through an activity they call a “data interview,” during which librarians identify and characterize data sets faculty have generated as part of their research (6). In a departure from the typical targets of collection management, such as monographs and serials, the indicated sets of data in effect become “information assets to be collected, preserved, and made accessible as a function of the library’s collection development” (6). To launch the interview, Witt and Carlson recommend asking researchers, “What is the story of your data?” The process of librarians working with researchers to figure out the story of their data is an investigative activity – akin to the “data exploration” that CTO Muñoz, of the NSF, advocates is pivotal to present-day scientific inquiry. In addition, collaborating with researchers in this capacity means that acquisition of published work cannot remain the key goal in collection development; ensuring procurement of related data files, as well as of software applications that the data are dependent upon for readability and analysis, emerge as significant objectives (4).

In what other ways are librarians prepared to undertake the development and implementation of research data management services? With the diversity of domains and disciplines represented at any given institution, how will librarians, many of whom individually serve multiple departments and thus myriad faculty research interests, address data management needs through scalable, sustainable approaches? What kinds of collaborations will need to be fostered, toward building effective service models for helping researchers curate their data? What role, if any, does institutional commitment play - perhaps in the form of institutional policies that not only support research data management practices but also even require them? How can librarians prepare graduate students, who are the next generation of researchers, to be data literate, toward gaining practical experience in data curation methods? These questions are not easily answered, but paths to solutions have begun to be paved, a few of which are considered next.

**Paths to Providing Data Management Services**

As Salo advises, bringing libraries up to speed to accommodate researcher needs in data management will demand a “retooling,” toward an understanding of the following: what research data are (e.g., their extent and scale, their project-based context, their heterogeneous nature, and the non-standard formats that data can take); and whether the functionalities of digital libraries and institutional repositories - the mainstays for digital content in most libraries -
map, or not, to the life-cycle management of data (18). The goal of the retooling process is to align, as much as possible, a library’s technical infrastructure and services with what researchers require in order to manage their data efficiently and effectively. To do this, many librarians have undertaken assessment work, such as surveying their faculty, or targeting specific ones for interviews, about their research data. They have also developed pilot studies to inform the development of service models. And still others have launched self-educational, “inreach” activities for the short-term, such as reading groups and topic-based information sessions, to help gain internal traction with data management issues. Additional literature describing tiers of service and concepts of success in the research data management space - such as collaboration that is transparent and all encompassing and a commitment to learning as a community - have surfaced in the last couple of years as well (19, 20).

Below is an overview of some of these diverse paths to new service models and infrastructure for addressing researcher needs in data management and data curation. (Not all of the experiences encapsulated here occurred directly after the NSF mandate, but the earlier timing does not diminish the value or relevance of the understanding that was achieved.)

Librarians are no strangers to survey instruments and semi-structured interviews, the chief methods they have exploited for evaluation of what researchers are doing with their data. In some efforts, a team of librarians develops its own survey instrument, or it draws on existing tools - as Georgia Tech Library did with the Data Asset Framework (27), originated by the U.K.’s Digital Curation Centre (22, 23). In 2006-2007 the University of Minnesota Libraries created its own survey for a “Science Assessment” study, in which sixteen librarians with subject expertise in science met with more than 70 researchers of varying statuses - faculty members, postdoctoral fellows, and graduate students - to discuss their research practices and needs in focus-group settings and interviews (7). A significant question that the librarians asked researchers, all of them scientists, in this study was, “If you seek assistance from the library, what kinds of help are you looking for? What kind of assistance is needed? (For grants? Publishing? Data curation and preservation?)” (7). The main areas where respondents sought help were organization and manipulation of data; storage, security, and sharing of data (specifically, the apparent lack of standards and guidance to consult); and stewardship of data - including some skepticism about whether this is even possible or necessary, given the inability to know the value of one’s research data years from now.

Evaluating the state of research data management at one’s institution can also mean targeting even more specific populations, such as the Primary Investigators (PIs) of NSF-funded projects and projects funded by the National Institutes of Health (NIH). Peters and Dryden did just this at the University of Houston as the basis of a pilot study, contributing to the university’s overall impetus to strategize for more robust research funding (24). Besides interviewing the PIs, a chief objective of the study was to gather information about data management approaches in the context of the projects the PIs directed, totaling fourteen, which were both group- and individual-based. The team made valuable discoveries pertaining to project information, data life-cycle workflows, data characteristics
(e.g., types of data being generated across the projects), data management (e.g., methods for data storage and access), data organization, and data use. Outcomes from the study included a proposal to form a library-based “Data Working Group,” to consolidate efforts among liaison librarians and to communicate with researchers in consistent fashion. Plans also evolved to sponsor an event bringing together “data service providers” from various parts of the campus, including the libraries, IT, and research centers, to gather and share what is being accomplished across the university and try to discontinue endeavors that are redundant. More of these types of collaborations will be needed in the near future, if institutions are to align service offerings and policies for data management. In addition, the pilot study team wished to expand their investigation to include researchers in non-science domains as well as researchers in science and engineering who have not secured funding-agency support for their projects.

Goldstein and Oelker undertook an approach at Mount Holyoke College, a small liberal arts college in New England, similar to that of Peters and Dryden at the University of Houston. Beginning in summer 2010, the Library and Technology Services (LITS) Department at Mount Holyoke worked to address the NSF requirement in order to be prepared to help faculty fulfill it, engaging the College’s Sponsored Programs Office and various subject librarians. LITS liaison librarians polled faculty about the following: how much research data in digital format they have; how much they expect this data to increase by June 2012; and whether they have lost data in recent years that were not backed up (2). The librarians also enlisted the support of the Associate Dean of Faculty for Sciences to ensure as much cooperation from faculty members as possible. In addition, the head of Digital Assets and Preservation Services, which is responsible for preservation of digital content, began collaborating with a science liaison librarian. Together they formed an early “response team” to address requests from Mount Holyoke faculty for assistance with DMP development (2). Besides surveying their faculty on basic data management practices, LITS librarians sought to find out what might be happening in data management at neighboring institutions, such as the University of Massachusetts at Amherst, to suss out possible external collaborations (25). Based on their data management services start-up experience, Goldstein and Oelker recommend that librarians at small institutions “adopt a policy of cooperation and collaboration” and work proactively to address those “just-in-time” needs for help with DMPs from faculty, while also strategizing for expansion and improvement of existing tools and services, such as, in the case of Mount Holyoke, its DSpace repository instance (2). Goldstein and Oelker also emphasize keeping abreast of what peers are doing in the area of data management and of what is occurring nationally.

In addition to assessing research data management activities on campus (i.e., external to the library), librarians have also been looking within their bounds, performing gap analyses via surveys distributed among information professionals in research libraries. For example, librarians at the University of Massachusetts Medical School did a study to evaluate required competencies for providing e-science research activities, such as data curation and management services. Surveying librarians in six U.S. states (out of 141 librarians who received the survey, 63 responded), Creamer et al determined that while a small percentage
of librarians were actively providing such services, more than half the number of respondents were involved in creating a strategic plan or policy for data management (26). Some of the competencies distinguished by Creamer et al in their study were the following: technical competencies, such as providing data archiving and preservation services, working with metadata standards, and managing an institutional repository; and non-technical competencies, such as outreach and instruction in various aims of scholarly communication (data sharing, open access, intellectual property rights, data literacy), conducting data interviews, working with researchers on DMP development, and finding and locating data that their patrons need for their own scholarship. For Creamer et al, these survey findings were instrumental in the development of both the “e-Science Portal for New England Librarians” and a data curation and management curriculum as professional advancement resources (26).

Another tactic employed toward an improved understanding of data management service requirements has been to participate in, and promote, “inreach” activities of the “train the trainer” mode: that is, a librarian or small team of librarians with base knowledge of data management practices organizes instruction sessions, workshops, or reading groups as vehicles for distributing this knowledge among their colleagues. At the University of Virginia, Sallans and Lake formed a “Scientific Data Consulting Group,” one of whose primary aims was to conduct bi-weekly “Data Curation Brown Bag” discussions (27). The intent behind these brown bags was threefold: 1) inform subject specialists of urgent topics and trends in data curation; 2) give a brief talk and provide a one-page, so-called white paper summarizing the issue at hand, followed by informal discussion; and 3) enable subject experts to become familiar enough with data curation for them to engage in discussions about it with the departments and faculty whose interests they represent. At the aforementioned University of Houston, Peters and Dryden had plans to develop “data 101 instruction sessions” to help their colleagues conquer the data literacy learning curve (24).

Other institutions have ventured beyond assessment and internal instruction and launched pilot projects to test proof-of-concept aims. In spring 2011, at the University of California San Diego (UCSD), the Research Cyberinfrastructure (RCI) unit put out a call to faculty researchers to submit applications for participation in its Research Curation and Data Management pilot program. (RCI itself was formed following much planning, reporting, and organizing around service and infrastructure requirements in research data curation.) At UCSD eight pilot projects were approved - five emphasizing data curation needs, three steeped in storage needs. RCI is assisting the pilot projects “with the creation of metadata to make data discoverable and available for future re-use; with the ingest of data into the San Diego Supercomputer Center’s (SDSC) new Cloud Storage system, which is accessible via high-speed networks; and with the movement of data into Chronopolis, a geographically-dispersed preservation system” (28).

Efforts such as the foregoing will establish leads to answers for a host of questions librarians have been considering, such as new staffing roles and new kinds of collaborations (not to mention a better understanding of whom to collaborate with and how to collaborate effectively). These efforts will also increase knowledge about organizational capacity for improving and expanding

current infrastructure, likely leveraging it in unforeseen, innovative ways – an advancement important for both small and large institutions. As conveyed below, a few libraries have developed service models worthwhile examining in depth, particularly for discerning common characteristics or practices that others might adopt and build on.

Examples of Data Management Services in Libraries

Following the NSF mandate - and, perhaps for some institutions, even preceding it - many libraries created new websites to convey information about the DMP requirement, what it meant to researchers, and how that particular library could help. An early exemplar was MIT’s Data Planning Checklist, which posed questions that retrospectively mapped well to the five sections of the NSF’s suggested approach to DMPs (29). Another web-based resource that developed just before the NSF requirement went into effect was the Association of Research Libraries’ Guide for Research Libraries: The NSF Data Sharing Policy; it unpacks what a DMP is, states the leadership role that libraries harbor in this effort, offers guidance on how to help researchers craft a DMP, and aggregates a range of resources relevant to data curation and data management (30).

Since those heady first days of acting as early responders to the NSF mandate by assisting researchers with their questions and concerns, many libraries have arrived at service models, revising infrastructure or establishing it anew and, in some cases, creating new positions to support these promising service frameworks. There are too many to present in adequate detail here, but the examples highlighted below vary enough from each other to afford a picture of rich possibilities for other libraries to adapt for their local contexts.

Cross-campus collaborations in support of services for research data management mark one requisite for success. Besides distributing and sharing responsibility, collaborations between libraries and other entities at an institution help ensure that a diverse range of needs are investigated and met. Cornell University’s Research Data Management Service Group (RDMSG, (31)), whose members are referred to as “consultants,” brings together not only librarians and specialists in IT (such as security) but also people with experience in project management, copyright and intellectual property rights issues, high performance computing, and data management system design. The model for collaboration and consolidation endorsed by RDMSG resulted in part from extensive planning and gap analysis work, represented in their report, Meeting Funders’ Data Policies: Blueprint for a Research Data Management Service Group (32). RDMSG has sponsorship from both the Senior Vice Provost for Research and the University Librarian; a faculty advisory board helps the group discern data management needs among faculty researchers, as well as determine additional resources and services for facilitating DMPs; and a management council exercises further oversight. The group is clear about the services it makes available, one of which is to “Provide a single point of contact that puts researchers in touch with specialized assistance as the need arises.” In addition, it consolidates in a single directory the services, tools, and resources found across campus that are
relevant to research data management, including tools for collaboration, guidance in intellectual property rights and data publication, and services in storage and backup, metadata creation, and data analysis and tools for collaboration.

Another necessity for data management services in libraries, especially as suggested by the University of Virginia model discussed above, is the training component, whether in the form of workshops, instruction sessions, or information sessions. The University of Minnesota (UMN) Libraries have crafted a suite of training opportunities to meet the needs of their faculty, students, and staff. Three librarians with expertise in the sciences and the social sciences form the foundation of support in the UMN Libraries for data management planning services. The workshop offerings range from “Creating a Data Management Plan for Your Grant Application,” to “Introduction to Data Management for Scientists and Engineers,” to “Data 101: Best Practices throughout the Data Life Cycle” (33). Perhaps most important, the workshop on creating a DMP meets UMN’s requirements for continuing education in responsible conduct of research, thereby providing another incentive for faculty and students to enroll in it. At the UMN Libraries’ website, online tutorials and workshops given by other departments and units on campus are also listed - these include sessions on technology training, intellectual property, quantitative data management, qualitative data management, and HIPAA data security training, better data searching, and help with grant proposal development (e.g., tools and resources for it, as well as guidance geared to graduate students seeking grant funding).

A chief requirement in NSF’s DMP guidance is that data generated by a research project be preserved to ensure continued sharing and access. While quite a number of data repositories, or disciplinary repositories accepting data, exist, they tend to suit “big science” projects producing data at a far larger scale than many university-based researchers with NSF funding contend with, in reality. Libraries that run or manage institutional repositories (IRs) may be in a satisfactory position to accept “small science” data - although, as Salo asserts, most repository software applications are suited for finished scholarly publications such as journal articles and book chapters, rather than versions of data sets: “For data, permitting only the immutable is unacceptable [. . .] much of the value of data is precisely its mutability in the face of new evidence or new processes.” (18). In addition, IRs that do accept data sets can usually accommodate raw data files but have few or no additional tools for data visualization (18).

Some institutions have started addressing the challenge of data deposits in a repository context. One is Purdue University, through its Purdue University Research Repository or PURR, and another is Rutgers University, through its RUresearch Data Portal, a part of its Rutgers University Community Repository, also known as RUcore. PURR is an instance of the Purdue-developed Hub Zero platform, devised for collaboration and for sharing of data and tools needed for research data in the sciences: “PURR provides workflows and tools for ingestion, identification and dissemination of data as well as services to ensure data security, fidelity, backup, and mirroring. Purdue Libraries will consult with investigators to facilitate selection and ingestion of data with the application of appropriate descriptive metadata and data standards as well as to provide appraisal of data for long-term preservation and stewardship” (34). RUresearch Data Portal allows a
broad range of “research genres” that are defined largely by the type of data, or data container, germane to that particular genre (35). These include, but are not limited to, codebooks, experimental data, multivariate data, quantitative discrete data, and quantitative continuous data. The research domains represented in the RUResearch Data Portal, as of spring 2012, are cognitive science, computer science, environmental engineering, political science, and statistics. One of the service points in the RUResearch Data Portal is customization of a portal for searching and retrieval of one’s data, made possible largely by a “sophisticated, flexible metadata strategy that can customize metadata to support your primary audience yet still be compatible with prevailing metadata standards” (35).

When it comes to supporting data management for grant-funded projects, some institutions are implementing cost-based models, particularly for storage and archiving services. The aforementioned PURR charges for extra project space (for the life of the project) and extra publication space (for ten years). To archive research data for a minimum five-year period, Johns Hopkins University charges a fee that is 2% of the direct total cost on an NSF grant, “with the option for an extension, and our expert support helping you prepare data for preservation and sharing” (36). Johns Hopkins is explicit about what its Data Archive offers—and thus what NSF PIs would be paying for, which is an archive that accepts discipline-agnostic data; a “data integration framework” enabling querying across the archived data; and a “preservation-ready system” (36).

Finally, a key component in data management service models is dedicated referencing of institutional policies and guidelines related to research activities, if not a distinct institutional policy for research data. In this particular service area, institutions in the U.K. are strides ahead of those in the U.S. The Digital Curation Centre, based in Edinburgh, Scotland, maintains a evolving online list of institutions with agreed-upon data management policies as well as a list of institutions that have completed policy drafts. The definition and implementation of an institution-wide policy for research data management has many dependencies, not least of which is obtaining buy-in from a spectrum of campus stakeholders. Such a commitment requires a common understanding among stakeholders (who, at the outset, may encompass librarians, IT staff, researchers, and administrators) of the issues, needs, and goals for management of research data throughout their life cycle. One place with an institutional policy regarding research data is Johns Hopkins. Its “Policy on Access and Retention of Research Data and Materials” (36) defines what is meant by research data and by the role of the “primary responsible investigator” and specifies how long the university will retain the data (for five years). Another institution with a research data policy is the University of Tennessee, which addresses responsibility, control, retention, and ownership of data, as well as the rights of the University to research data (37). Other considerations for an institutional policy on management of research data could include - but are not limited to - the following: a commitment to offer training and support opportunities, as well as guidelines and templates, to help researchers create DMPs; provision of tools and services for conducting preservation actions on data and retaining them to ensure their uninterrupted access; and a resolve not to relinquish rights to commercial entities to reuse or
publish research data without making certain the institution continues to have the right to make the data publicly available (38).

Conclusion

The story of data management services in libraries reflects an abundance of still developing plot points and characters - ones that have begun to yield promising service frameworks, collaborations, tools, and new roles. An example of a tool that emerged a year after the May 2010 press release from the NSF is the DMP Tool, which walks researchers through a data management plan, allowing them to input the relevant information for each section, and then generates the plan (39). (It is up to the researcher to make sure the plan, after it is generated, does not exceed the two-page maximum length.) The tool also provides additional resources, such as DMP guidelines from not only the NSF but also the National Endowment for the Humanities and other funding agencies with similar requirements. The DMP Tool is a collaborative effort of seven institutions and organizations: California Digital Library, the Digital Curation Centre (U.K.), Smithsonian Institution, UCLA Library, University of Illinois at Urbana-Champaign, and University of Virginia Library. Developers of the tool are also enabling federated access to it for a growing number of universities and colleges.

Since the activation of the NSF requirement, several academic libraries have also created positions in which data curation or data management (or both) is the primary responsibility. Examples of some of these positions are “Data Services Librarian” (Kansas State University Libraries, (40)), “Data Curator” (Simon Fraser University Library, (41)), “Data Management Specialist” (Emory University Libraries, (42)), and “Manager, Data Management Services” (Johns Hopkins University Libraries, (43)). While these positions are based in an academic library, they are highly collaborative roles that work across library and, often, campus departments. They frequently require experience with project management and with curation of scientific data; knowledge of intellectual property rights issues, repository infrastructure, and metadata standards, as well as an awareness of rising trends, both locally and globally, in data curation and data management; excellent communication and interpersonal skills; and a dedication to providing the best possible patron service and support. In addition, in spring 2012, the Council on Library and Information Resources (CLIR), working with the Digital Library Federation (DLF), announced a new postdoctoral fellowship program in data curation, which the two organizations launched with funding from the Alfred P. Sloan Foundation. A main intent behind this effort is “to raise awareness and build capacity for sound data management practice throughout the academy” (12). In summer 2012, CLIR and DLF announced the first cohort of Data Curation Postdoctoral Fellows.

The ultimate aim in telling the story of data as thoroughly and properly as possible, drawing on some of the paths and examples relayed above, is so that one’s data will be able to be found, preserved, accessed, shared, used, and reused - not just in the current century but also beyond it. Additionally, it is so that
researchers themselves will have a clear idea of where their own data are. As this chapter suggests, traction is gaining in favor of libraries, IT units, research administration, faculty, graduate students, and others working together to flesh out the story of data at their campuses. They are encouraging best practices and standards, reconfiguring roles and responsibilities strategically to meet demands in data management, and planning for infrastructure that aligns, rather than duplicates, services across an institution. In other words, the story of data in the 21st century and further is the story of a brave new world.

References


