CHAPTER 10

Research Data Management and the Health Sciences Librarian

Andrew Creamer, Elaine R. Martin, and Donna Kafel

INTRODUCTION

As science becomes increasingly characterized by large-scale collaborations and computational **data** sets, researchers face a range of data management challenges and needs. This creates an opportunity for health sciences librarians to offer researchers at their institutions a range of data management strategies and services. By providing **research data management (RDM)** services, librarians connect and collaborate in new ways with the researcher communities within their institutions. But exactly what kinds of RDM services can health sciences librarians offer? How can health sciences librarians engage with biomedical researchers and market these services? These are the questions that current library school students concentrating in health sciences librarianship and practicing health sciences librarians are asking.

The answers are many and varied. Along with actually searching, assigning metadata elements, and curating, preserving, and archiving data sets in digital collections, health sciences librarians are teaching researchers and students about RDM fundamentals, best practices, and assisting them with writing **data management plans (DMPs)**. The librarians who fill these research roles may be called e-science librarians, scientific data curators, research librarians, data librarians, research informationists, or embedded librarians. Whatever the title, these librarians are engaging with the data needs of the health sciences research community. Some health sciences librarians are already participating in RDM activities, whereas others are still examining models to best adopt these emerging services. RDM services require the development, coordination, and synthesis of a range of library and institutional services and programs. This chapter outlines several of these health sciences librarian RDM services and roles, and addresses the following themes.

- What is RDM?
- Why manage research data?

wood_chapters.indb 252 3/5/14 6:02 AM

- Common RDM challenges
- RDM services and roles for health sciences librarians
- Librarian RDM skills and competencies
- Librarian RDM tools and resources

Health sciences librarians should explore helping researchers navigate this new data-intensive research environment, including maintaining their relevance and value for the research enterprise, for many powerful reasons. Combined with the needs of researchers to manage the overwhelming amount of data in their research labs, the National Institutes of Health (NIH) and the National Science Foundation (NSF) are requiring data management and data sharing planning to enable and optimize the use, reuse, and sharing of publicly funded research data. On February 22, 2013, the Office of the President directed federal agencies with more than $100 million in research and development expenditures to develop plans to make the published results and supporting data of federally sponsored research freely available to the public within one year of publication, and to require researchers to better account for and manage their data resulting from federally sponsored scientific research (Office of Science and Technology Policy, 2013). Federal research sponsors' prioritizing of sharing data and their DMP mandates are driving health sciences librarians' increased involvement in RDM.

BACKGROUND

The Association of Research Libraries (ARL) took the lead in identifying new roles for librarians in advancing **e-science** and supporting researchers. As early as 2006, ARL had launched its E-Science Task Force. It defined the domain of e-science as "those new methods that are large-scale, data-driven and computationally intense, and often engaging research teams across institutional boundaries" (ARL, 2010). The Task Force's first report highlighted the fundamental changes in the ways in which scientists carried out their work, the tools needed, and the nature of data documentation and data publication.

In 2010, ARL charged a Working Group with the task of raising member libraries' awareness of the professional skills librarians need to address data stewardship, policy development, and the data-publishing environment of scientists. Its report outlined the results of a survey seeking baseline data about institutional and library engagement with these issues. At that time, the data revealed a "professional community experiencing a not-unexpected diversity of approaches and investments," and "a portrait of emerging models" (ARL, 2010).

Davidoff and Florance (2000) proposed that librarians could play information specialist roles. Although many of the subsequent informationist projects were clinically focused, drawing upon the rich history of clinical librarianship as a model, the emerging informationist role is the librarian embedded on a biomedical research team (Martin, 2013). The **research informationist** reinforces the concept of "immersion" into the culture and workplace of the clinicians and scientists being supported (Florance, 2013).

In 2013, the National Library of Medicine (NLM) began sponsoring a new support program at NIH: The NLM Administrative Supplements for Informationist Services in NIH-funded Research Projects. This program specifically requested proposals from NIH-funded research teams requiring research informationists as part of their teams; the scope of the grant particularly focused on the health sciences librarian's role in data management. NIH granted eight awards, and although the librarian roles and projects varied, each application emphasized the librarian's added value to the researchers' data management needs (Martin, 2013). In this age of big data, community, clinical and translational science awards, human genomics, bioinformatics, and the electronic health record (EHR), the possibilities for health sciences librarian involvement with managing data seem boundless.

> The embedded informationist role is a new one for my library, yet it appears to be a type of service it needs to invest in for the future. Becoming integrated into a research team gives the informationist, and thus the library, opportunities to become more involved and, ultimately, more valued by the researcher community. The informationist can bring library and information services a greater depth than those offered via traditional support roles.
>
> *Sally Gore, MS, MS LIS, embedded research librarian and informationist, Lamar Soutter Library, University of Massachusetts Medical School, Worcester*

WHAT IS RDM?

The University of Edinburgh Data Library (2011) describes research **data** as unique information products that are "collected, observed, or created for purposes of analysis to produce original research results." In other words, a single datum represents information in isolation, disconnected. However, through a human or a machine's processing, contextualizing, and interpretation of data, meaning emerges. Replicating this meaning provides evidence to confirm or refute hypotheses, ask new questions, and cultivate new knowledge, new discoveries. Commonly cited types of research data are:

- data derived from observations;
- data derived from experiments, such as measurements taken by scientific instruments under varying and constant conditions;
- data derived from simulations and models, such as models used to predict epidemics and changes in climate; and
- data derived from already existing data that scientists can reexamine, repurpose, and build upon to derive original discoveries. (University of Virginia, 2012)

Because research data can take so many forms, it is logical for them to be recorded and stored in many types of digital file formats. There are competing digital file formats for spreadsheets, text documents, audio, video, and image files. Thus, many tenets of records management inform RDM: being aware of the types of data that a

project produces; following standards for labeling and appraising data; being aware of the formats that are best suited for making data discoverable, accessible, and reusable by others; and being aware of the digital file formats that are best suited for the long-term preservation of those data (Creamer, 2013a).

WHY MANAGE RESEARCH DATA?

"The activities that keep data secure, accessible, and useable comprise research data management; it is a fundamental part of good research practice" (McLeod, Childs, and Lomas, 2013: 71). DataONE (2013) describes RDM as necessary for "making data easier to find, use, analyze, share and reuse." Managing research data is not just something that may be required of a researcher by an institution or a research funding body; it is an important part of the research process. In 2011, Julie McLeod developed one of the first courses aimed specifically at teaching RDM to health sciences postgraduate researchers; she argues that RDM is an integral part of Research Methods (Creamer, 2013a).

Many practical benefits come from having a DMP and following it. Researchers will have no duplication of efforts; they will save time, energy, and resources. Each person on the lab team will know what he or she needs to do in order to manage, document, and protect the integrity of the project's data. Researchers will be able to locate and share data easily among team members; everyone will have the documentation to understand and interpret the data. Researchers will be able to analyze and publish data with the confidence that they are well organized, appropriately formatted, and documented. Additional reasons for managing research data are:

- comply with sponsor/institution/publisher policies;
- comply with ethical and legal requirements;
- enable reuse and new discoveries;
- increase organization and efficiency; and
- meet professional standards and ensure data integrity. (McLeod, Childs, and Lomas, 2013: 74)

Integrity of Scientific Research

In several high-profile instances, published journal articles have been retracted because of falsified or missing data. "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" (Hswe, 2012: 117). Sponsors, institutions, publishers, the scientific community, and the public should be able to verify the integrity of published scientific findings. Data managed well can be more easily stored, discovered, shared, accessed, interpreted, and scrutinized. Robin Rice and Stuart Macdonald, data librarians at the University of Edinburgh, observed that a researcher will be more willing to share well-managed data if he or she feels confident about their integrity (Creamer, 2012). **Data sharing** is important in the health sciences because **data reuse** and repurposing may help foster new research, new discoveries, and better health outcomes (Institute of Medicine, 2011).

Mechanisms are increasing for authors to share and publish the data that underlie the results of their published articles. In 2009, scientists at the National Academy of Sciences (NAS) recommended that research data be managed and made accessible to help ensure proper professional conduct and data integrity. This report put the onus for managing data on the researcher:

> Ensuring the integrity of research data is essential for advancing scientific, engineering, and medical knowledge and for maintaining public trust in the research enterprise. Although other stakeholders in the research enterprise have important roles to play, researchers themselves are ultimately responsible for ensuring the integrity of research data. (NAS, 2009: 4)

This report was followed by a series of editorials in leading scientific journals calling on research institutions to do more to provide their scientists with the appropriate support and tools for managing and sharing their research data (*Nature Neuroscience* Editorial, 2009).

In 2011, a group of these scientific journals worked with digital repositories such as Dryad to write the Joint Data Archiving Policy (JDAP) (Dryad, 2011). It required their authors to make their articles' supporting research data publicly available. Their consensus was that transparency and access to these data are crucial for the public and scientific community to scrutinize a study's overall integrity; thus, research data must be managed well in order to be shared effectively. Research integrity, professionalism, and maintaining the public's trust are powerful incentives for health sciences researchers, librarians, and their institutions to prioritize RDM.

Data as Return-on-Investment

Graham Pryor, associate director at the Digital Curation Centre (DCC) and editor of the book *Managing Research Data*, argues that data represent a return on public investment. "Surely," he writes, "data produced so expensively should not be treated as spoil, put aside like the waste materials from an intellectual mine" (Pryor, 2012: 4). Pointing out the incredible sums of public money spent on research, Pryor argues that the resulting data can no longer be ignored. He laments that the only intellectual product afforded attention by the academy and public is the resulting publications: "Justifiably, they might also expect that the fruits from such rich endeavors will be afforded the attention necessary to ensure an optimum return on investment" (Pryor, 2012: 1).

Researchers are also reaping dividends from their published data in the form of more frequent citations (Piwowar, 2013). The scientific community is increasingly starting to view data sets as just as important as the publications they underlie (DataCite, 2012). "Data should no longer be abandoned on the workbench like wood shavings in a carpenter's shop; increasingly it is expected to join the finished assembly of scholarly output as a valued and managed component with an extended life and sustained usability" (Pryor, 2012: 4). Although the traditional model of measuring researchers' impact valued only the citation of their articles, now alternative metrics (**altmetrics**) are emerging to measure the impact of a researcher's data and other online output as equally relevant scholarly products (Piwowar).

Compliance

Julie McLeod and Sue Childs found that health sciences researchers recognized that a major incentive for them to prioritize RDM was having the organized documentation and confidence to prove that they conducted their research legally, safely, and ethically, and gathered and managed their subjects' data appropriately (Creamer, 2013a). It has long been the practice of sponsors and institutions to expect scientists to conduct their research within the appropriate legal and ethical frameworks, and comply with institutional, sponsor, and legal policies. These stakeholders have reserved rights to investigate their researchers' conduct and audit materials such as laboratory notebooks, research data, consent forms and administrative documentation, lab animal protocols, and so forth. Researchers working with patient data or with other types of sensitive data have to meet a large number of legal and ethical requirements at the institutional, sponsor, state, and federal levels. From a liability standpoint, researchers have to document that they have followed the required Office for Human Research Protections (OHRP) and Institutional Review Board (IRB) policies, ethics protocols, and procedures for properly collecting, storing, and securing sensitive patient data, and that they have taken the appropriate measures to protect their participants' identities and protected health information (PHI).

Federal Data Management and Sharing Policies

A good reason to manage and share data is that more sponsors are requiring it. The NIH policy states that "all data should be considered for data sharing" and "data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data" (U.S. National Institutes of Health, 2007). The NSF policy mandates that researchers include a two-page data management plan as a supplementary document with any proposal for funding. The plan must describe how the proposal will "conform to NSF policy on the dissemination and sharing of research results" (NSF, 2011). The OSTP's memorandum on behalf of the Office of the President to the heads of executive departments and agencies within the federal government states:

> The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data. (OSTP, 2013: 1)

The U.S. Department of Health and Human Services (HHS) is currently implementing an Open Government Plan. It has developed the HealthData.gov Platform (HDP) to make health data sets available to the public. "The HDP will deliver greater potential for new data driven insights into complex interactions of health and health care services" (Challenge.gov, 2012). These sponsors' data-sharing initiatives will increase the importance of compliance, and hence the visibility of RDM services on the campuses of universities, hospitals, and research institutes.

COMMON RDM CHALLENGES

Health sciences librarians at New York University created a YouTube video titled *A Data Management Horror Story* (Hanson, Surkis, and Yacobucci, 2012). It features a comical exchange between two cartoon cancer researchers. One researcher is interested in looking at the other's data to repurpose for her research. However, the other researcher reveals that he cannot easily find the project's data files. Eventually he tracks down the files, but then he finds that he can no longer remember the meaning of the field titles that his team used, or the values and parameters. The other researchers in his lab have moved on, and the company that created the software program he used to create the files, which the other researcher would need to open and view the files, has gone bankrupt and closed.

Although entertaining, the health sciences librarians' video highlights the serious issues surrounding RDM and the consequences of mismanagement being the stymieing of data sharing, data reuse, and further discovery. RDM challenges include managing the work flows of team science: getting everyone on the team to follow a DMP, and making RDM a priority. The frequency of students, residents, and postdoctoral students (postdocs) rotating in and out of labs creates challenges as does having data stored in multiple places, and in some cases, multiple research team members spread across the globe (Ferguson, 2013). Most principal investigators (PIs) do not manage the day-to-day aspects of their labs. Graduate students and residents who come and go manage the data. Their practices for the most part are ad hoc. Multiple investigators may be collaborating on the same research but manage their data differently in each of their labs. This makes it difficult to share data within the same project and re-create results.

> As the health sciences librarian on a research team, I have helped to address several different RDM challenges. The biggest challenge we face is that the team's data exist in lots of different places and different formats. Like many research teams, they have a paper lab notebook with the results of their different experiments, as well as a lot of digital data that go with their experiments. Most of these digital data are images in proprietary systems, and these systems are not connected to each other. The challenge is to figure out how we can get all of these data into one place.
>
> *Lisa Federer, MLIS, MA, AHIP, research data informationist at NIH Library, Bethesda, MD*

In 2013, the HHS Office of Research Integrity (ORI) noted several RDM issues for improvement:

- technical data not recorded properly;
- technical data management not supervised by PI;
- data not maintained at the institution;
- financial or administrative data not maintained properly;
- data not stored properly;
- data not held in accordance with retention requirements; and
- data not retained by the institution. (Erickson and Muskavitch, 2013)

The ORI's issues highlight several problem areas where health sciences librarians are assisting researchers in addressing

- lack of assigned RDM responsibility;
- lack of a DMP;
- issues with records management;
- lack of metadata;
- lack of **data dictionaries**;
- issues with storing, backing up, and securing data;
- issues related to data ownership and intellectual property;
- lack of plans for retaining and appraising data; and
- lack of long-term plan for preserving data. (NECDMC, 2013)</bl>

Lack of Assigned RDM Responsibility

As the ORI noted, confusion exists among health sciences researchers regarding who is responsible for RDM. Over the past several years, it has become apparent from librarian surveys and interviews with students that labs do not always have a codified procedure or formal course to teach new students or team members how to manage data; students report having to learn it informally "on the job," from another graduate student, or not at all (Piorun, 2013). The ORI wants students and PIs to know that everyone in the institution bears some responsibility for managing data. By assigning RDM responsibilities and delegating tasks, health sciences researchers will increase the efficiency of their research. Laboratory notebooks, both paper and electronic, can be audited by a researcher's institution or sponsor, such as NIH, and collectively they represent a record of science; therefore, a plan is needed for managing and preserving these notebooks. A plan must also bridge the RDM knowledge of outgoing and incoming students, residents, postdocs, and staff.

Lack of a Data Management Plan (DMP)

Since the NSF began requiring PIs to submit data management and sharing plans with funding proposals, health sciences librarians have published articles, research guides, curricula, and online tools to helping researchers write these plans and understand their value for the research process. According to the NSF, these plans should describe:

- the types of data that will be created;
- who will own, have access to, and be responsible for managing these data;
- the equipment and methods that will be used to capture and process data;
- the metadata that will make these data make sense to others; and
- where the data will be stored during and after the project. (NSF, 2011)

Having a DMP will help researchers account for the day-to-day management of a project's data throughout the life of their projects, and beyond. A **data life cycle** describes when data are created, how they can be managed along the steps of the

research process, and how these data can be appraised and archived after a project is complete (Ball, 2012; DataONE, 2012). Julie McLeod and Sue Childs describe such life cycle DMPs as "living documents meant to be revisited, added to, and revised as the project unfolds and situations change" (Creamer, 2013a).

Records Management Issues

Many data management issues are related to locating and making sense of data. Common records management failures include inconsistently labeled data files; unmarked file versions; poorly structured folders; undocumented multiple media storage; undocumented file locations; and inconsistent file formats. Ferguson (2013) looked at a sample of student data files on a set of instruments in a bioscience lab. The file names she saw broke a lot of the "rules" for good file management. She saw file names such as "awesome" with just the year, and another version of the file titled "awesomer." Although humorous, these students' choices for their file names exemplified the need for improvement. The students' file naming conventions did not take into consideration how someone not involved in a project would make sense of what is in the file. Indeed, after some time, these files would probably not even make sense to the person involved in creating the file!

Lack of Metadata and Data Dictionaries

Metadata can help to address several data management related issues.

- How will researchers label, document, describe, and contextualize their data?
- How will they describe the data files to make them more discoverable by others?
- How will someone else make sense of their data during and after the project (e.g., field names, terminology, values, parameters, etc.)?

Many Clinical and Translational Science Award (CTSA) sponsored research groups use a Web-based application called REDcap to collect their data. In order to set up its data collection spreadsheet, someone on the team has to create a **data dictionary**—a guide documenting, labeling, and customizing the values and parameters used to describe the data that will be captured. Before collecting any data the researcher would need to label fields and provide adequate documentation (e.g., explain field names, field types, instrumentation, questions, responses, branching logic, etc.). Researchers have to document these details systematically and deliberately in order to record and interpret the data they have captured; these details will help them to make judgments later about which data they will want to analyze, which data will be discarded, and which will be saved or can be shared with others.

Several types of metadata can help to make sense of data. Metadata can be descriptive and give details about the nature of the information in the files: who created it, information about the subject or experiment, where was it created, and when, and so forth. Metadata can describe the file that is storing the data, such as how many bytes, the format, the software used to create the file, the hardware and software needed to open and view it, and the version. Some metadata are structural;

they can help the user to navigate the files and understand which files are associated or linked with other documentation. Metadata can be administrative and communicate if the data are licensed; it can also be technical and relay information about the instrumentation that created it. Over the years certain disciplines have come together to create metadata standards, agreed upon elements that can be used to describe their domain's data.

Storage, Backup, and Security Issues

Properly storing, backing up, and securing data are important responsibilities. Institutions and sponsors want researchers to take these responsibilities seriously to protect and ensure the integrity of their data. The IRB wants to ensure that researchers protect the identities and private information of their human subjects. Before starting a project, researchers need to understand their institutional **data security** policies and their responsibilities for storing and securing data. These policies can address several data management issues.

- How often should research data files be backed up?
- How many copies of the data should researchers have?
- What storage options do they have within the institution?
- How much institutional server space can they get and how much does it cost?
- Are they allowed to use their personal hard drives, portable storage such as USBs, or commercial cloud storage vendors? (NECDMC, 2013)

Data Ownership Issues

Determining the ownership of data deserves attention before anyone begins to collect data. Data ownership can be very situation-dependent. However, a growing number of institutions have data ownership and intellectual property policies. Typically, institutions come from the perspective that they provide the employment and facilities of researchers, and therefore they own any data produced by researchers who are employed by the institution and using institutional resources. Similarly, federal sponsors view the institution as the owner of research data and ultimately responsible for its management. The institution sees the PI and students as stewards of the data. If a

For us, performing outreach in the area of data management has involved leveraging existing liaison relationships while reaching across traditional organizational divisions. A request from a biomedical researcher for information on storage and computing resources at UF led us to meet with the director of UF's High Performance Computing Center. Resultant collaborative campus-wide outreach activities have included an invitation to present on research data management best principles/practice as part of Research Computing Day, membership on the Data Life Cycle Subcommittee, creation and teaching of workshops on best practices in RDM, and formation of the Data Management/Curation Task Force.

Rolando Garcia-Milian, MLS, AHIP, basic biomedical sciences librarian; and Hannah F. Norton, MSIS, AHIP, reference and liaison librarian, UF Health Science Center Library, Gainesville, FL

PI working on a federally sponsored research project were to leave to go to a new institution, the ORI would still expect the PI's former institution to be able to account for the project data even after the PI had left (Erickson and Muskavitch, 2013).

No Plan for the Long Term

Before researchers embark on a research project, it may prove difficult for them to appraise the value of their data or think about **data preservation** before any are produced and recorded. However, regardless of how valuable they might think these data may prove to be, researchers should locate their sponsor's data retention policies and policies within their institution issued by stakeholders such as the IRB, Office of Responsible Conduct of Research or Intellectual Property, and the library, which could one day be the archive for these data. The regulatory framework for retaining and destroying data is complex, and researchers or their institution may be called upon to produce data, even if it has been years after a project has ended.

Regulations for retention can overlap, depending on the type of research, the nature of the data, and the sponsors of the research. For example, with human subjects' data or other sensitive data, the regulations are specific regarding the collection, storing, sharing, publishing, and archiving (or destroying) of these data. On top of the institutional and sponsor retention guidelines, state and federal laws might apply as well. Thus, many health sciences researchers are interested in locating information about retention because they want to be in compliance. If questions or allegations arise concerning the validity of a health sciences researcher's data or conduct, then he or she would want to retain all data and supporting documents until the investigations have been resolved.

RDM SERVICES AND ROLES FOR HEALTH SCIENCES LIBRARIANS

In 2012, the Association of College and Research Libraries (ACRL) identified **data curation** as a top trend and highlighted academic libraries' research data services models. These models depend on the following services:

- information and researcher services (e.g., conducting **data interviews** and consulting on researchers' DMPs);
- metadata services (e.g., **annotating and describing data**, creating records for data, **linking publications with data sets**, etc.);
- reference support services (e.g., finding and citing data);
- educational services (e.g., developing subject guides or websites, teaching **data literacy**, RDM best practices);
- technical support for **data repositories** (e.g., **ingesting data sets** and building digital collections, and **assigning persistent identifiers**); and
- **digital curation** services (adding value by preserving data and access to data for the long term). (ACRL Research Planning and Review Committee, 2012)

In the face of fiscal constraints, many academic health sciences libraries cannot afford to hire and designate a single librarian to offer RDM instruction and consultation services; the trend has been to create a working group of librarians from various library departments to share their expertise and cooperate to provide RDM services. Within smaller health sciences and hospital libraries, clinical and subject liaison librarians have chosen to upskill and add RDM services to their support services.

Overall, health sciences librarians' involvement in RDM services is quite idiosyncratic. It is a reflection of the librarians' personal RDM competencies, the level of library and institutional administrative support they receive, as well as the resource and budget capacity of their libraries to support their RDM services (Creamer et al., 2012). In addition, they reflect the nature of their researcher communities' RDM requests made in the context of their institutions and the frequency of these demands. The librarians and their administrators must also envision how they see these circumstances and trends evolving in the future (Piorun, 2013). It may mean that for a library or librarian to add or strengthen a RDM service, they will have to reduce or end other services (Cox, Verbaan, and Sen, 2012).

The emerging RDM roles for health sciences librarians blend a variety of traditional and new skill sets and incorporate aspects from a range of the library's services: public access, reference and instruction, research and scholarly communications, technical services, archives and records management, and collections. These roles include:

- participating in institutional data policy making;
- providing expertise on scholarly communications;
- managing an institutional data repository;
- teaching RDM best practices;
- providing instruction on using RDM-related tools and resources;
- providing RDM reference and consulting;
- offering advice on disciplinary metadata standards and repositories;
- helping researchers to annotate and apply metadata to their data sets;
- offering advice on data ownership, copyright, licensing, and intellectual property;
- helping researchers to locate and cite research data sets;
- creating RDM guidance websites, subject guides, and learning materials;
- archiving and preserving data sets in open formats; and
- partnering with researchers and embedding on research teams.

Instruction and Access

From an instructional and reference perspective, health sciences librarians offer instruction to student and faculty researchers about the value of RDM, and the best practices, tools, and resources for managing and curating their research data (Federer, 2013). Health sciences librarians are stepping outside of the library to offer RDM consulting. They conduct reference-style data interviews with their institutions' researchers to assist with planning and managing their data throughout the life cycle of their research projects, and to write DMPs required or encouraged by a research funder.

Collection Development

From a collection development and archives perspective, health sciences librarians create and manage collections of research data (Kochi, Chatterjee, and Rizk-Jackson, 2013). They help researchers, even in the planning phases of their projects, to appraise and think about the archival and preservation options for their data, as well as the potential for sharing their data. Librarians help researchers to take the appropriate steps to make their data sets more accessible and to provide an adequate amount of metadata to make their data both discoverable and meaningful to other researchers (Higgins, 2012). Health sciences librarians assist researchers at the end of their projects to locate, evaluate, and ingest their data into an appropriate institutional or disciplinary repository. These roles involve being aware of a variety of factors affecting the sharing of data such as institutional policies related to intellectual property, and institutional and research funders' open access and open data sharing policies.

It is never too soon for researchers to consult health sciences librarians about their plans for publishing and sharing data, and appraising and preserving data. Many serendipitous scientific discoveries are out there waiting to happen, and the federal government wants to make publicly funded research data publicly available in the hope that it can spur future innovation. Repositories and data archives play important roles in sharing and preserving data, but they are not the only options. Researchers can self-archive or self-publish their data, place their data in their institution's data archive or repository, place their data in an open data repository, or place them in a national or international data archive or repository.

Health sciences libraries support data sharing and the building of digital collections by providing data repositories, guidance on open file formats, assigning persistent identifiers to digital objects, and by teaching **data citation** practices. Depositing in a repository is not always the same as preserving data; although some data archives' mission is to maintain and preserve data for the future, other repositories may simply be a place to access and store data—no active preservation over time will maintain its integrity. Health sciences librarians help researchers to understand this and the scope and policies of a repository or archive before they submit data.

Scholarly Communications

The stewardship of data described above also has become increasingly important to scholarly communications as the U.S. government and its research funders such as NIH and NSF have continued to mandate and encourage researchers receiving a certain level of public funds to make their resulting publications, and soon their underlying data, publicly available. Simultaneously, the academy is also evolving to see scholars' products of research, including their research data and online output, as equally valuable contributions alongside their publications. Health sciences librarians involved in RDM recognize that they and their libraries have a role in helping to implement and bring about these paradigm shifts in scholarly communications within the cultures of their institutions (Palmer, 2013).

Health sciences librarians help researchers to locate and use open file formats and to obtain persistent identifiers for their data sets such as DOIs. Persistent identifiers

make data sets or an associated publication easier to find online, and they are used to measure the citation impact of research data sets. Papers that published data sets are cited more often (Piwowar, 2013). Thus health sciences librarians teach researchers about the emerging conventions for citing the data they use or repurpose for their own research.

RDM Guidance

RDM services from health sciences librarians seem to expand daily, but one of the core services is still providing their researcher communities with the best and most timely information about managing their data. The most visible ways to accomplish this are by creating an online presence for RDM at the library, and by teaching and creating educational materials to support student and faculty researchers (Crummet, 2013). Over the past few years, health sciences librarians have begun to write RDM guidance Web pages, and to create online subject and research guides on a variety of RDM topics. Health sciences libraries provide their researcher communities with an online overview of RDM, tutorials, information about the research and data life cycles, and subject guides for tools and resources that support their RDM planning. These may include the conventions for naming their files and structuring folders to house these files, or guidance on writing DMPs. Health sciences librarians have created checklists, templates, and authoring tools to help their researchers write DMPs, and some librarians have even compiled the different data management and data sharing policies and requirements from major research funding bodies so their researchers can have all of this information at hand. Lastly, these library tutorials point researchers to the support services available at their own institutions. A common set of guidance topics includes:

- selecting open file formats, file and folder naming conventions, versioning control, and folder structuring;
- best practices for data security, storage and backup;
- information on intellectual property, copyright, and data licensing;
- publishing data;
- citing data;
- disciplinary and generic metadata for data sets;
- data analysis tools;
- anonymizing/de-identifying patient data;
- sharing data; and
- information on ethical and legal issues.

RDM Instruction

Another highly visible role and important RDM service that health sciences librarians provide is teaching (Crummet, 2013; Hanson and Surkis, 2013; Norton et al., 2013). The library can be a logical venue for researcher communities to learn about RDM best practices, DMP resources and tools, and institution-specific research support services, policies, and resources. These formal in-person or online courses complement many of the courses that librarians are already providing, such as

locating research funding and writing grant proposals, support for writing a dissertation, copyright and publishing, and tutorials on searching PubMed, Evidence-Based Medicine, and using citation management tools.

Librarians in academic health sciences libraries have had to customize RDM teaching models to meet the needs and time demands of their researcher communities: integrating RDM into their lab management curricula and courses on research methods; collaborating with other instructors to present RDM lectures in their courses, or to the members of a research group; offering RDM library workshops for students during postgraduate induction periods; and creating online course modules (Kafel et al., 2013; Creamer, 2012; Creamer, 2013a).

Data Management Consulting

Writing DMPs also requires that health sciences librarians build up a certain level of knowledge about the research being conducted in their institution, the institution's data management and sharing policies, and the support services available to researchers. This equips the health sciences librarians with the knowledge to identify strategies for the types of data being produced, and they will be able to offer institutional-level advice for managing, storing, securing, backing up, and describing these data, and the local or disciplinary repository options available for archiving the data after the project is finished.

Data management consulting requires health sciences librarians to keep current on the policies affecting health sciences research. Although best practices for RDM apply across all disciplines, some disciplines, such as the health sciences, have unique regulatory considerations when it comes to helping researchers manage and share their data. For example, some policies have an impact on how data can be managed and shared in a clinical trial (ClinicalTrials.gov, 2013). Federal laws such as HIPAA regulate the collection, retention, and use of patient data, www.hhs.gov/ocr/privacy/. Institutional polices such as those overseen by the IRB regulate research using human subjects. These policies affect the types and amount of identifying information that can be used in a data set; the amount of time that data and laboratory records are required to be maintained; the procedures for storing, securing or destroying data; and which data, if any, can be shared or made publicly available.

Creating Data Repositories

Many large academic health sciences libraries have their own repositories for researchers to submit their publications and their data sets (Kochi et al., 2013; Palmer, 2013). When it comes to ingesting data into a repository, unique considerations include the size and formats of the files, core metadata, and the appraisal methods necessary to determine which data should be preserved. Health sciences librarians consult with researchers interested in placing their data into a repository about the types and amount of core metadata to use that will make their data meaningful to others.

Health sciences librarians develop and curate repositories and collections of research data. In addition to institutional repositories, health sciences librarians assist researchers with uploading their data to disciplinary repositories required by sponsors. For example, NIH funds a number of data-sharing repositories. Health sciences

librarians create persistent identifiers for data sets and link publications to their underlying data, and keep abreast of the evolving conventions for creating data citations.

The ability to appropriately cite data is professionally important for health sciences researchers for a number of reasons. They may have reused data and want to responsibly cite others' data in their own research. They also want to receive credit from other researchers who have reused their data sets. The ability to document the number of times others have downloaded and cited their data could be of use in defense of their promotion and tenure by demonstrating how their data have increased the impact of their research (Piwowar, 2013). Organizations such as DataCite, www.datacite.org; Impact Story, www.impactstory.org; and Orcid, www.orcid.org, are working to increase the acceptance of a researcher's data and online works as equally legitimate scholarly contributions alongside their publications. Health sciences librarians are an important part of this paradigm shift because they help researchers to locate and cite data sets, and to publish their own data (Palmer, 2013).

Championing Public Access to Data

Health sciences librarians help researchers to see the value in publicly and openly sharing their data. Some types of health sciences data certainly cannot and should not be shared, such as sensitive or patient identifying data. But within the health sciences library community it is accepted that transparency in the research process (open science) and universal access (open access) to health information such as research data will benefit the health of a society by leading to better research and better health outcomes (HealthData.gov, 2013). The research community's inability to share, access, or reuse data results in "missed opportunities in economic, social and scientific advancements" (Pryor, 2012: 1). Health sciences librarians create guidance Web pages and research guides to help their researchers learn about open access, open science, and open data; how to avoid the use of proprietary file formats; how to prepare and submit their data to open repositories; how to locate and search for data sets within open repositories; and how appropriately to repurpose and cite the data they find (Hanson et al., 2013).

In recent years, NYU Health Sciences Libraries has been building a suite of data management services. To help evaluate potential new data services, my colleague and I are partnering with a research team seeking to acquire new data sets to merge with grant-generated data from their previous studies. In light of the team's need to support the introduction of external data sets into its core data set, we are evaluating and restructuring their data model and data entry tools.

Karen Hanson, MLIS, knowledge systems librarian, NYU Health Sciences Libraries, New York

Data Management and Sharing Policies

Health sciences librarians shape their institutions' policies regulating the management and sharing of their researchers' data. Data management and data sharing policies not only outline the standards for data management, the data ownership,

and intellectual property rights of the researchers and institutions, they also set the bar for each department's RDM roles and responsibilities, including the library (Creamer, 2013b). Health sciences librarians are on the committees that craft and implement these policies because they understand that these policies can officially recognize the responsibility of the library to support the institution's research enterprise. This recognition formalizes the role and the many efforts that libraries have pursued to offer RDM services. These policies directly impact the work of libraries to organize, build, and disseminate local collections of data and their mission to ensure that information is preserved and accessible to their users.

LIBRARIAN RDM SKILLS AND COMPETENCIES

Some health sciences librarians want to know if they need to have a background in science or have conducted scientific research to get involved with RDM (Creamer et al., 2012; Gore, 2012). Although this question has no definitive answer, health sciences librarians offering RDM services seek out professional development opportunities to learn more about science and the research process (Kafel, 2012). Corral (2012) found that the library and information science schools that have data curation programs use the data life cycle and **research life cycle** as a foundation for their curricula and many require practical field experience to engage students with the research process. Higgins's (2012) data life cycle outlines the relevant data management and preservation considerations that may arise at each of these steps.

Health sciences librarians embedded into research teams help to manage and preserve their projects' data from the outset, coordinating database design and knowledge mapping (Martin, 2013). Librarians need to build on their core information science knowledge base and expand that base to include technology, informatics, and outreach skills (Corral, 2012; Cox, Verbaan, and Sen, 2012; Creamer et al., 2012; Gore, 2012). Specific skills such as database management, database design, information searching, research methods, and digital preservation are traditional librarian skill sets. Data literacy, being able to communicate and solve problems, having an entrepreneurial spirit, a willingness to take risks, confidence, and working as part of a team are skills that librarians may be less familiar with (Gore, 2012).

Job postings for librarians involving some aspect of data management appear on a regular basis (Martin, 2012). Sample job titles include director of research data management, scientific data curator, data services manager, and associate director for data science. Selected job duties include:

- being the primary liaison for scientific data management
- developing library services and guidelines for supporting scientific research
- representing the library in institutional policy development
- facilitating the collection, preservation, and access to scientific data
- serving as a consultant for data management plans and practices for campus constituencies

Required qualifications are extensive and may include some or all of the following:

- advanced degree in physical or life sciences, data curation, or related disciplines
- recent graduates of data curation programs
- an understanding of the research process
- demonstrated technical knowledge
- familiarity with metadata
- understanding of institutional repositories
- grant-writing experience (E-science Portal for New England Librarians, 2013)

Creamer et al. (2012) surveyed health sciences librarians about their data management roles and competencies and the types of professional development instruction they felt were needed to offer RDM services. In addition to knowledge about RDM best practices, they needed domain-specific knowledge, data literacy instruction, and additional professional development opportunities to increase their technical competencies. Corrall (2012) found LIS programs teaching RDM skills to their students focused instruction on data literacy and technical skills. However, Cox, Verbaan, and Sen (2012) reported that the challenges to librarians' adopting RDM roles were not technical or data-related, but rather their lack of direct personal experience with research and difficulty balancing the provision of RDM services with their other priorities.

LIBRARIAN RDM TOOLS AND RESOURCES

The Data Interview

The data interview has become the essential tool for librarians to engage a researcher about his or her research practices and RDM needs (Witt and Carlson, 2007). Librarians use this type of interview to help researchers thoughtfully review their project and communicate the essential information about their data needs. This includes information on how the projects produce data, the nature of these data, as well as any challenges their teams may encounter managing these data. The data interview also provides the librarian with a tool to help researchers explore their intentions to preserve and share their data post-project.

The Data Curation Profiles

One of the first tools to build and expand upon the data interview was the Data Curation Profiles Toolkit (Purdue University and University of Illinois Urbana-Champaign, 2010). Librarians use the toolkit to interview researchers and explore the life cycle of a research project and requirements and issues regarding the resulting data. The completed **Data Curation Profiles** also comprise a digital reference collection with detailed information about a specific research project, its methods, and its data.

The DMPTool

Today, online data management planning tools such as the **DMPTool** have become key resources for librarians helping researchers to write DMPs for funders

and to plan for their data management over the life cycle of their projects. A service of the University of California Curation Center of the California Digital Library, the DMPTool is an online tool that researchers can use to write their DMPs using customized templates for specific research funders and their data management and sharing requirements (California Digital Library, 2013). Any researcher at any institution can use the DMPTool; however, institutions that become partners of the DMPTool can customize the templates with guidance, resources, and policies specific to their institution.

RDM Professional Development Resources

In 2009, the Lamar Soutter Library at the University of Massachusetts Medical School (UMMS) and the National Network of Libraries of Medicine, New England Region (NN/LM, NER), created a continuing education program to develop RDM roles for health sciences librarians. The program has expanded to include a cadre of data-related professional development and continuing education events, courses, tools, and resources. UMMS hosts a data-related librarian Community of Interest (COI), an annual symposium, an annual boot camp, and an annual librarian professional development workshop, as well as a Web portal for librarians, http://esciencelibrary.umassmed.edu/. In 2011, UMMS published a framework for a data management curriculum that included lesson plans, readings, and activities; and, in 2012, it launched a journal to publish health sciences librarians' data-related scholarship (Kafel, 2012).

In 2013, UMMS partnered with several libraries to publish the New England Collaborative Data Management Curriculum (NECDMC), http://library.umassmed.edu/necdmc/index, an online, case-based RDM course intended for use by health sciences librarians to teach RDM best practices to students and researchers. Each of the course modules aligns with NSF's DMP recommendations. The course's growing collection of actual research teaching cases provides discipline-specific context for the content presented in the modules. The teaching cases come from a range of research settings including clinical research, biomedical labs, clinical trials, and a qualitative behavioral health studies.

SUMMARY

RDM will increasingly become a part of health sciences librarianship. Today's health-care delivery and research environments use a range of digital technologies to track measurements of patient health indicators and store clinical data. Tools such as EHRs, patient-monitoring devices, automated lab instrumentation, and smartphones have become invaluable sources for locating and collecting health-care data. The unprecedented volume of clinical and research data is rapidly growing. To ensure that researchers, clinicians, and patients can access these data, they must be appropriately managed, stored, and preserved. The stewardship of data requires a coordination of efforts by multiple players in the research environment—PIs, researchers, clinicians, lab assistants, IT staff, statisticians, funding agencies, and health sciences librarians. The successful use of these data in research and patient

care has the potential to transform health care and spur treatments that will drastically improve the quality of human lives (Institute of Medicine, 2011).

The Institute of Medicine (2011) noted: "The successful development of clinical data as an engine for real-time knowledge generation has the potential to transform health and health care in America." However, broader access to and use of health-care data requires not only fostering reliable and accessible data systems, but also addressing the issues such as individual data ownership and patient and public perception of clinical data as a carefully stewarded public good. The adaptation of digital technologies in health-care delivery holds great promise to spur innovations in detection and treatment of disease and improve patient care. Thus, as operating components of health-care and research environments, health sciences libraries are adopting and developing roles in RDM.

This chapter has highlighted emerging RDM roles for health sciences librarians: writing RDM guidance pages, teaching RDM best practices, demonstrating RDM tools and resources, offering RDM consulting and assistance with DMPs, building and managing collections of research data, and embedding within clinical and biomedical research teams. The RDM services offered vary by institution and by librarian. They depend on the number of staff, time, resources, and funds available, as well as the individual experience, skills, and competencies of the librarian. Offering RDM services is an evolving and emerging role for health sciences librarians. No one health sciences librarian can offer the full range of all data services, but he or she can observe best practices, share curricula, cooperate, and learn from colleagues. Retooling and learning new skills (upskilling) are important as few libraries have the funds to hire additional staff. The health sciences librarians undertaking these roles need continued administrative support and professional development. Investing in RDM services is important for health sciences libraries because they will ensure their future relevance.

STUDY QUESTIONS

1. What activities comprise RDM?
2. Why are health sciences librarians teaching researchers how to manage their data?
3. How have the federal government research sponsors' data management and sharing policies affected the roles of librarians?
4. Why do librarians use the research and data life cycles to teach data management?
5. What is the benefit of having a data management plan?
6. What RDM skills would health sciences librarians need to partner with/support researchers?

REFERENCES

ACRL Research Planning and Review Committee. 2012. "2012 Top Ten Trends in Academic Libraries." *College & Research Libraries News* 73, no. 6 (June): 311–20. http://crln.acrl.org/content/73/6/311.full#sec-3.

ARL (Association of Research Libraries). 2010. "E-Science and Data Support Services." August. www.arl.org/storage/documents/publications/escience-report-2010.pdf.
Ball, Alex. 2012. "Review of Data Management Lifecycle Models." February. http://opus.bath.ac.uk/28587/1/redm1rep120110ab10.pdf.
California Digital Library. 2013. "About the DMPTool." DMPTool. Accessed May 21. https://dmp.cdlib.org/about/dmp_about.
Challenge.gov. 2012. "Health Data Platform Metadata Challenge." Office of the National Coordinator for Health Information Technology. Accessed May 21. http://challenge.gov/ONC/374-health-data-platform-metadata-challenge.
ClinicalTrials.gov. 2013. "Protocol Data Element Definitions." Accessed May 21. http://prsinfo.clinicaltrials.gov/definitions.html.
Corral, Sheila. 2012. "Roles and Responsibilities: Libraries, librarians and data." In *Managing Research Data*, edited by Graham Pryor, 105–34. London: Facet.
Cox, Andrew, Eddy Verbaan, and Barbara Sen. 2012. "Upskilling Liaison Librarians for Research Data Management." *Ariadne*. December 6. www.ariadne.ac.uk/issue70/cox-et-al.
Creamer, Andrew. 2012. "Creating an Online Research Data Management Course: A Conversation with Data Librarians Robin Rice and Stuart Macdonald." *E-Science Community*. October 9. http://esciencelibrary.umassmed.edu/rice_macdonald_interview.pdf.
———. 2013a. "Crossing That Bridge We Have Come to: Teaching Students How to Manage Qualitative Data. A Conversation with Professor Julie McLeod and Susan Childs about DATUM for Health and DATUM: Research Data Management at Northumbria University." *E-Science Community*. March 19. http://esciencelibrary.umassmed.edu/esci_community_creamer_20130312.pdf.
———. 2013b. "A Seat at the Table: Involving the Library in the Planning of Institutional Research Data Management Policies and Responsibilities. A Conversation with the iridium Project Team at Newcastle University." *E-Science Community*. May 22. http://esciencelibrary.umassmed.edu/esci_community_creamer_20130312.pdf.
Creamer, Andrew, M. E. Morales, J. Crespo, D. Kafel, and E. R. Martin. 2012. "An Assessment of Needed Competencies to Promote the Data Curation and Management Librarianship of Health Sciences and Science and Technology Librarians in New England." *Journal of eScience Librarianship* 1, no. 1. http://escholarship.umassmed.edu/jeslib/vol1/iss1/4/.
Crummet, Courtney. 2013. "One Academic Library's Response to the Data Dilemma." Contributed paper presented at the annual meeting of the Medical Library Association, Boston, May 6.
DataCite. 2012. "Why Cite Data?" Accessed May 21. www.datacite.org/whycitedata.
DataONE. 2012. "Best Practices." The Data Life Cycle. Accessed May 21. www.dataone.org/best-practices.
———. 2013. "Data Management Guide for Public Participation in Scientific Research." DataONE Public Participation in Scientific Research Working Group. February. www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf.
Davidoff, Frank, and Valerie Florance. 2000. "The Informationist: A New Health Profession?" *Annals of Internal Medicine* 132, no. 12 (June): 996–98.
Dryad. 2011. "Joint Data Archiving Policy." Dryad. Reaffirmed April 7. http://datadryad.org/pages/jdap.
Erickson, Stephen, and Karen M.T. Muskavitch. 2013. "Data Management." U.S. Department of Health and Human Services. Office of Research Integrity. http://ori.hhs.gov/education/products/rcradmin/topics/data/open.shtml.
E-science Portal for New England Librarians. 2013. "Career Opportunities." University of Massachusetts. May 28. http://esciencecommunity.umassmed.edu/category/jobs/.
Federer, Lisa. 2013. "All It Takes Is One: Single-Session Data Literacy Instruction." Contributed poster presented at the annual meeting of the Medical Library Association, Boston, May 5.

Ferguson, Jen. 2013. "Lurking in the Lab: Analysis of Data from Molecular Biology Laboratory Instruments." *Journal of eScience Librarianship* 1, no. 3 (March): 148–58. http://escholarship.umassmed.edu/jeslib/vol1/iss3/5/.

Florance, Valerie. 2013. "Informationist Careers for Librarians: A Brief History of NLM's Involvement." *Journal of eScience Librarianship* 2, no. 1 (May): 3–5. http://escholarship.umassmed.edu/jeslib/vol2/iss1/2/.

Gore, Sally. 2012. "Failure to Communicate." *A Librarian by Any Other Name*. September 25. http://librarianhats.net/2012/09/.

Hanson, Karen, Alisa Surkis, and Karen Yacobucci. 2012. "Data Management Snafu in 3 Short Acts." *Youtube*. December 12. www.youtube.com/watch?v=N2zK3sAtr-4.

Hanson, Karen, Theodora A. Bakker, Mario A. Svirsky, Mario A. Neuman, Arlene C. Neuman, and Neil Rambo. 2013. "Informationist Role: Clinical Data Management in Auditory Research." *Journal of eScience Librarianship* 2, no. 1 (May): 24–29. http://escholarship.umassmed.edu/jeslib/vol2/iss1/7/ /.**[AQ1]**

Hanson, Karen, and Surkis, Alisa. 2013. "Introducing Researchers to Data Management: Pedagogy and Strategy." Contributed paper presented at the annual meeting of the Medical Library Association, Boston, May 6.

HealthData.gov. 2013. "Unleashing the Power of Data and Innovation To Improve Health." U.S. Department of Health and Human Services. Accessed May 21. www.healthdata.gov/unleashing-power-data-and-innovation-improve-health.

Higgins, Sarah. 2012. "The Lifecycle of Data Management." In *Managing Research Data*, edited by Graham Pryor, 17–45. London: Facet.

Hswe, Patricia. 2012. "Data Management Services in Libraries." In *Special Issues in Data Management*, edited by Norah Xiao and Leah Rae McEwen, 115–28. Washington, DC: American Chemical Society.

Institute of Medicine. 2011. "Clinical Data as the Basic Staple of Health Learning." National Academies. February 3. www.iom.edu/Reports/2011/Clinical-Data-as-the-Basic-Staple-for-Health-Learning.aspx.

Kafel, Donna. 2012. "Activities of Regional Consortia in Planning e-Science Continuing Education Programs for Librarians in New England." In *Special Issues in Data Management*, edited by Norah Xiao and Leah Rae McEwen, 69–96. Washington, DC: American Chemical Society.

Kafel, Donna, M. Piorun, S. Najafi, T. LegerHornby, and E. R. Martin. 2013. "Frameworks for a Data Management Curriculum for Science, Health Sciences, and Engineering Students." Contributed poster presented at the annual meeting of the Medical Library Association, Boston, May 5.

Kochi, Julia K., K. Cameron, A. Chatterjee, and A. Rizk-Jackson. 2013. "DataShare: Facilitating Scientific Data Sharing." Contributed paper presented at the annual meeting of the Medical Library Association, Boston, May 6.

Martin, Elaine R. 2012. "What Do Data Services Librarians Do?" *Journal of eScience Librarianship* 1, no. 3 (March): 1–2. http://escholarship.umassmed.edu/jeslib/vol1/iss3/3/.

———. 2013. "Highlighting the Informationist as a Data Librarian Embedded in a Research Team." *Journal of eScience Librarianship* 2, no. 1 (May): 1–2. http://escholarship.umassmed.edu/jeslib/vol2/iss1/1/.

McLeod, Julie, Sue Childs, and Elizabeth Lomas. 2013. "Research Data Management." In *Research Methods in Information*, 2nd ed., edited by Alison Jane Pickard, 71–86. Chicago: Neal-Schuman.

(NAS) National Academies of Sciences. 2009. "Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age." Committee on Science, Engineering, and Public Policy (COSEPUP). http://books.nap.edu/openbook.php?record_id=12615.

Nature Neuroscience Editorial. 2009. "Ensuring Data Integrity." *Nature Neuroscience* 12, no. 1205. www.nature.com/neuro/journal/v12/n10/full/nn1009-1205.html. doi: 10.1038/nn1009-1205.

NECDMC (New England Collaborative Data Management Curriculum). 2013. Lamar Soutter Library. University of Massachusetts Medical School. November. http://library.umassmed.edu/necdmc/index.

Norton, H. F., R. Garcia-Milian, M. Tennant, and C. E. Botero. 2013. "Supporting the Local Research Data Environment via Cross-Campus Collaboration and Leveraging of National Expertise." Contributed paper presented at the annual meeting of the Medical Library Association, Boston, May 6.

NSF (National Science Foundation). 2011. "Dissemination and Sharing of Research Results." Accessed May 21. www.nsf.gov/bfa/dias/policy/dmp.jsp.

OSTP (Office of Science and Technology Policy). 2013. "Expanding Public Access to the Results of Federally Funded Research." February 22. www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research.

Palmer, Lisa. 2013. "Article Level Metrics." Contributed paper presented at the annual meeting of the Medical Library Association, Boston, May 5.

Piorun, M. 2013. "The Library's Role in E-Science Programs in Research Universities." Contributed paper presented at the annual meeting of the Medical Library Association, Boston, May 5.

Piwowar, Heather. 2013. "Altmetrics: Value All Research Products." *Nature* 493, no. 159 (January). www.nature.com/nature/journal/v493/n7431/full/493159a.html.

Pryor, Graham. 2012. "Why Manage Research Data?" In *Managing Research Data*, edited by Graham Pryor, 1–16. London: Facet.

Purdue University and University of Illinois, Urbana-Champaign. 2010. "Purpose and Use of the Profiles." Data Curation Profiles. Accessed May 21. http://datacurationprofiles.org/purpose.

U.S. National Institutes of Health. 2007. "NIH Data Sharing Policy." Office of Extramural Research Grants & Funding. April 17. http://grants.nih.gov/grants/policy/data_sharing/.

University of Edinburgh Data Library. 2011. "Research Data Explained." *MANTRA*. June 20. http://datalib.edina.ac.uk/xerte/play.php?template_id=9.

University of Virginia. 2012. "What is Data: An Introduction to Data Types." University of Virginia Graduate Student Portal. July 9. http://pages.shanti.virginia.edu/SciDaC_Grad_Training/2012/07/09/observational-data/.

Witt, Michael, and Jake R. Carlson. 2007. "Conducting a Data Interview." Purdue University Libraries Research Publications. December. http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1092&context=lib_research.