Mooney, H. (2016). Scholarly communication and data. In L. M. Kellam & K. Thompson (Eds.), *Databrarianship: The academic data librarian in theory and practice* (pp. 195–218). Chicago: Association of College and Research Libraries.

CHAPTER 13

Scholarly Communication and Data

Hailey Mooney

Introduction

The Internet and digital data are strong forces shaping the modern world of scholarly communication, the context within which academic librarians operate. Scholarly communication entails the ways by which scholarly and research information are created, disseminated, evaluated, and preserved.[1] Recognition of the broader forces at play in research and scholarship is imperative to keep ourselves from obsolescence and to simply function as effective librarians. The operation of the scholarly communication system is the "bedrock" of academic information literacy and forms the "sociocultural frame of reference" for understanding library research skills.[2] The purpose of this chapter is to provide foundational knowledge for the data librarian by developing an understanding of the place of data within the current paradigm of networked digital scholarly communication. This includes defining the nature of data and data publications, examining the open science movement and its effects on data sharing, and delving into the challenges inherent to the wider integration of data into the scholarly communication system and the academic library.

The Nature of Scientific Knowledge and Data

The sociology of science provides a basis from which to understand the fundamental underpinnings of the norms and values that govern the institution of science and the production of scientific knowledge. In basic information literacy instruction, librarians teach the difference between popular and scholarly information. We identify scholarly information based on a set of criteria such as a peer-reviewed publication venue and the presence of citations. These identifiers are normative practices in the production of scientific knowledge, a particular type of knowledge which requires certification based on conformance to the methods and values of the scientific community; what Robert Merton calls the scientific ethos.

According to Merton,[3] the ethos of science is based upon four institutional imperatives: universalism, communism, disinterestedness, and organized skepticism. Universalism requires that claims be objectively and impersonally supported, rather than based on the particularities of the author. Communism is the common ownership of goods. Within the institution of science, this means that scientific findings are meant to be widely communicated and shared. Disinterestedness implies lack of personal interest created by conformity to the institution of science that monitors the activity of individual scientists, as through the process of peer-review, in order to verify results. And finally, organized skepticism is the practice of suspending personal belief systems in favor of empiricism and logic. These imperatives of science as an institution support the implementation of the scientific method, which is based upon continual input of observations, or data, in order to build the body of scientific knowledge.[4]

Data, therefore, form the basis of information and knowledge[5] and the use and provision of data can be integrated into Merton's theory of the scientific ethos. For the objective claims of universalism, data provide evidence. The sharing of scientific findings, found in the imperative of communism, conceivably includes the sharing of data that form the basis of those findings. Disinterestedness and the need to validate results are strengthened by the ability to scrutinize and replicate results from data. Favoring empiricism, as necessitated by the norm of organized skepticism, rests on the scientific method and the use of data. Yet despite the intrinsic value of data to the creation of scientific knowledge, the exchange of scientific findings generally do not occur at the level of data, but through the sharing of information—explanations and interpretations of data through the vehicle of the written word in standard forms of scholarly communication such as the journal article, the conference paper, or the book.

There are two important and problematic issues with the nature of data that explain why the written publication has been prioritized over the provision of data within the scholarly record. The first is that data, by themselves, have no meaning to the outside reader. Data require context, documentation, and explanation. The second problem is that data take multiple forms and datasets lack standard structure.

Whereas the sociological frame provides a useful viewpoint to understand the institution of science, an anthropological perspective illuminates the process of data creation within the individual research projects that create the overall body of scientific knowledge. A number of studies have examined the particular practices by which data are created within various scientific projects; here in this section we look at two. Carlson and Anderson[6] selected four projects for study ranging across hard, soft, pure, and applied disciplines: an astronomy data grid, a social survey project, a digital curation project for a collection of museum artifacts, and a collection of anthropological studies of foreign cultures. They found that in all cases, data are characterized by their problematic heterogeneity and the need for careful formatting. Project materials only become usable data through unique processes that allow for materials to be "transportable" as a discrete object and "intelligible" to outsiders. For example, in the social survey project, words are converted to numbers, variables are coded, and files are restructured in order to create usable and intelligible data. The processes of conversion, formalization, and provision of documentation for the data are rarely self-explanatory or self-contained and therefore required "complementary external information" in order for the data to be understood.[7] This is particularly difficult for anthropology projects where private field data has traditionally not been made available for outside use beyond analysis and summary in books or articles.

Ribes and Jackson, in their study of ecological water stream sample data, likewise find data to be a sticky entity, a "commodity fiction," in that data are not easily made into an undifferentiated and universal form.[8] Traveling with the ecologists in field, they find that the creation of data is mutually constructed into the everyday work of scientists and that careful rituals are observed in order to ensure comparability of measurements in the creation of the stream data. Despite the presence of an online database, the stream dataset is "not singular; rather it is…distributed across databases, field sheets, and a physical archive of samples."[9] The project data are constituted of multiple parts; their production require coordination and documentation (for example, one scientist bemoans the possibility of walking into the cold storage room and finding labels have fallen off the samples: "I think the extra label tape we put on the lids will hold up."[10]). Within a single project, and indeed between different projects, data take multiple forms, are made up of multiple parts, and require documentation.

The proliferation of forms that data can take engenders debate on what exactly data are, resulting in a multiplicity of definitions. Data take as many forms as there are disciplines and methodologies. There are as many definitions of data as there are people doing research.[11] Despite this variation, there are a couple of key commonalities across definitions that provide us with a basis to understand data:

1. *Data are evidence.* Data have an evidentiary function and requires analysis and interpretation in order to derive meaning.[12]
2. *Data can take multiple forms.* Many definitions of data provide examples of data forms and formats. However, no list can be complete; note how some definitions will include a listing of research methodologies by which data can be produced in addition to specific types. Forms can include both file types and variant aggregations of data.[13]

In addition to taking multiple format types, data may be aggregated at varying levels and characterized as a *dataset* or a *data collection*. Sets or collections of data may include multiple files of data (potentially of different types) along with accompanying documentation. The dataset is generally a smaller collection of data records, such as a single social science dataset created from a social survey.[14] A data collection refers to an aggregation of data and/or datasets, and so it is larger in scope. For example, the Data Citation Index defines the dataset as "a single or coherent set of data or a data file" which may be part of a data study (generally a description of associated datasets), and at the top level of aggregation makes up the overall data collection of a data repository.[15] The National Science Board[16] identifies three levels of data collections based on functional categories. *Research data collections* include data from one or several focused research projects, serve a small specific group, and are unlikely to persist beyond the life of the research project. *Resource* or *community data collections* are the result of community-level standards and research areas and may be directly funded by stakeholder agencies, but often do not have a clear commitment to long-term maintenance. *Reference data collections* are created for the benefit of large general scientific communities, employ or set standards for data creation, and have large budgets, which often include the capacity for long-term access and preservation.[17]

Therefore, more common than the single file of data is the conception of data as compound objects that can take myriad forms, resulting in a "fuzzy conception" of data.[18] This is in part, a philosophical dilemma to define a relative concept. However, there are very real and practical implications for the integration of data into the scholarly communication system. If we cannot structure, define, and extract meaning from data, then how can we expect to systematically collect, organize, and disseminate data? In order for data to fit into a system that is largely based on dealing with discrete information commodities (e.g., the journal article, book) it must conform to a publication based paradigm. If the compound object of the dataset can be wrapped up into a distinct object, then it may appear logical to conceive of data as a publication—however, it has been proposed that conceiving of data as publication is a problematic metaphor.[19] By attempting to make data fit an old paradigm, we may limit the ability of the scholarly communication system to evolve and adapt to new models of scholarship.[20] The shift to digital technologies is an opportunity to redefine communication models, and by attempting to make data fit into the existing molds we inhibit the creative potential to share data in multiple modalities and fully embrace new forms of scholarship. Even within the realm of information sharing (as opposed to data sharing) forward-thinking scholars are experimenting with emerging formats (e.g., interactive books, blogging) that do not have a clear place within the existing formal modes of scholarly communication.[21]

Despite the rhetorical and practical issues around the fuzzy conception of data, datasets, and data collections, pragmatism dictates that fitting data into the mold of publication is the path forward for integration into the scholarly communication system as it currently stands. Publication, after all, serves an important purpose within the social system of science by legitimizing, creating a priority claim for the author, and allowing for dissemination, access, and preservation.[22] Data publication allows for data to be recognized as "first-class objects" within the scholarly record at-large through the implementation of management and curation.[23] At the micro-level, data publications provide recognition to data creators by affording the ability to bestow citations: the currency of scholarly and scientific achievement. Furthermore, *data publication* stands as a formalization of other methods of *data sharing*. For example, sharing data directly with collaborators or other vetted scholars (as through email) does not afford data with the same level of access, curation, and preservation given to an item housed within the collection of a research library.

Open Data and the Forces Behind Data Sharing and Publication

The open access movement is predicated on the value of science as a public good and leverages Internet technology to provide free and unrestricted access to scholarly literature.[24] The jump from open access for literature to open access to data is not a particularly big one given that the knowledge transferred by the scholarly literature is built from the creation and analysis of data.[25] Although certain types of data have been long published and shared, and calls for increased data sharing date back decades,[26] the recent open access movement has shifted the attention on data sharing into high gear. The reasoning behind open data rests upon two levels of social ideologies: within the scientific community as an ethical imperative, and within government as an economic and political issue.

Advances in science can benefit society. Science is conceived of as a public institution wherein it is the duty of scientists to share their findings. For example, although the United States is a capitalist economy with ongoing tension between private and public goods, knowledge is widely considered to be a public good that would otherwise be under-produced if not supported by government funding.[27] U.S. government agencies fund significant amounts of research through the provision of grants, with the federal government spending $31.0 billion on basic research in FY 2012, 50.8% of which was awarded to universities and colleges.[28] The U.S. government is itself a major data producer, with 13 major statistical agencies and over 100 total agencies that also produce statistical information as part of their mission.[29] However, the issue of transparency and access to data produced directly by government agencies is distinct from access to data funded by government grants but produced by extramural researchers within academia and non-government research institutes. Both may be referred to as *open data*, but the more distinct terms of *open government data* and *open science data* serve to clarify the difference. There are shared ideological tenets, such as right of access to work resulting from public funds, accountability, and the value of knowledge sharing, but open government data (although used in scholarly research) is not a direct component of the scholarly communication system.†

In the United States, the legislative debate over open access (as seen in bill proposals such as the Federal Research Public Access Act, Fair Copyright in Research Works Act, and the Fair Access to Science and Technology Research Act) has focused largely on traditional publications.[30] However, as recounted by the Congressional Research Service, federal debate over the right of public access to government funded research data dates back to 1980.[31] The Supreme Court ruled in *Forsham v. Harris* that a grantee's research data were not subject to a Freedom of Information Act (FOIA) request because the files were not federal agency records. In 1999, the issue was raised again with the revision of OMB Circular A-110, the regulatory document which stipulates requirements for grants. Senator Richard Shelby shepherded the inclusion of the Data Access Act of 1999 (a provision within the omnibus P.L. 105-277, also known as the Shelby Amendment) which provided for the revision Circular A-110 "to require Federal awarding agencies to ensure that all data produced under an award will be made available to the public through the procedures established under the Freedom of Information Act." The America Competes Reauthorization Act of 2010 (P.L. 111-358) stipulated that the Director of the Office of Science and Technology Policy (OSTP) coordinate agency policies related to the dissemination and stewardship of scholarly publications and digital data produced by government funded research. This mandate eventually resulted in the February 22, 2013 OSTP memo, *Increasing Access to the Results of Federally Funded Research*, which requires federal agencies with annual research and development expenditures of at least $100 million to develop plans for public access to publications and data.[32] The gap between the 1999 Shelby Amendment and the 2010 America Competes Reauthorization reflect a period of building momentum within the open access movement and the parallel attention paid to the merits of open data.

Many government research grant funders have complied with policy mandates to support the concept of public access to data by requiring data management plans. Current data management plan policies generally require discussion of whether or not data sharing is feasible and how it will be accomplished, but they are vague as to how data sharing should be accomplished and do not specifically require formal publication of data through deposit in a recommended archive or other means.[33] For example, the National Science Foundation (NSF) requires that data management plans discuss policies for sharing and access, reuse, archiving, and preservation, and also that investigators are expected to share their primary data with other researchers; however, there is no specific requirement that data

† For an introduction to open government data, see the *Open Data Barometer Global Report*, 2nd ed. (World Wide Web Foundation, 2015), http://barometer.opendataresearch.org/.

be deposited in a public database.[34] Likewise, the National Institutes of Health data sharing policy guidance offers multiple methods for data sharing. Archiving data with a repository is a possibility, but so is handling data requests individually under the auspices of the primary investigator.[35] Although the NSF and other data management plans do not actually require data publication, they are widely seen as an important step in the direction of increased data sharing.

The recent focus on increasing access to government-funded scientific research data as a matter of public good is also reflected in international policy initiatives. The Organisation for Economic Co-operation and Development endorsed a set of recommended principles and guidelines for the provision of access to publicly funded research data, noting that "the power of computers and the Internet has created new fields of application for not only the results of research, but the *sources* of research; the base material of *research data*."[36] UNESCO supports open access for wide and equitable dissemination of research and recommends that open access policies developed by governments, institutions, and funding agencies consider including data as a relevant research output for policy inclusion.[37] As part of their open access strategy to increase the impact of funded research, the European Commission explicitly includes research data in addition to publications as reflected in their *Recommendation on access to and preservation of scientific information*, which states that "Open access to scientific research data enhances data quality, reduces the need for duplication of research, speeds up scientific progress and helps to combat scientific fraud."[38] The Research Councils UK abide by a common set of principles on data wherein "publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible."[39]

A common thread throughout U.S. and international governmental policy discussions is the public's right to transparency and accountability. Data can form the basis of legislative decision-making. In 1997, a controversial air quality regulatory standard proposed by the U.S. Environmental Protection Agency was based in part on scientific research conducted by Harvard and funded by the National Institutes of Health. Questioning the validity of the research, legislators and industry groups requested access to the underlying data and were told that it was unavailable.[40] This prompted Senator Shelby's work towards the OMB Circular A-110 revision. Shelby advocated for data access arguing that "public confidence in the accuracy and reliability of information being used to drive public policy ultimately is in the best interest of scientific research. Increasing access to such data promotes the transparency and accountability that is essential to building public trust in government actions and decision-making."[41] Interest in the availability of scientific data in environmental decision-making in particular continues to this day.[42] The National Academies promote the principles of data integrity, sharing, and access, making the point that that "sharing research data enhances the data's integrity by allowing other researchers to scrutinize and verify them."[43] Legislative concern over open data is based on ensuring scientific integrity through verification and reproducibility of results in order to provide for scientific accountability and public trust in science.[44]

Another component of policy-maker and taxpayer right to access government funded research is the trend toward demanding increased "return on scientific capital."[45] Data sharing has the potential to maximize funding impact by reducing redundancy and enabling reuse for new scientific findings, providing for a higher return on investment.[46] Scientific innovation can be translated into economic benefits. When information and data are made available, knowledge can be leveraged to create new products and services.[47] This economic argument is not limited to politicians, but is ensconced within the philosophy of the scholarly open access movement, wherein public funding ought to equal public access.

Scholars recognize their role in advancing public knowledge. Scientists share the beliefs that transparency within the research process produces better science and that data sharing extends research dollars by allowing for innovative reuse. Sharing information and data is positioned as an ethical obligation to the scientific community and the public.[48] Data sharing and publication is viewed as a beneficial for multiple reasons, such as:

- *Providing access for reuse*. Maximizes the contribution of research subjects. Reduces overall research costs. Increases research impact through secondary analysis studies that ask new questions of the data or develop new analysis methodologies, or by initiating collaborations with the primary investigator.
- *Allows for replication and validation*. Reproducible science reduces risk of errors or outright fraud. Improves research practices by encouraging transparency. Enhances public trust in science.
- *Ensuring preservation*. Prevents data loss and redundant research efforts. Protects valuable resources.

Internal pressure within the scholarly community is exerted through editorials and essays promoting these data sharing benefits.[49]

The ethic of data sharing is written into the scholarly ecosystem via the adoption of policies by professional associations and scholarly journals. For example, the American Psychological Association code of ethics and the American Sociological Association ethical standards require researchers to make their data available.[50] A growing number of journals include policies for data sharing and publishing (such as the British Ecological Society journals, *The American Naturalist*, *Nature*, *Public Library Of Science (PLOS)*, American Geophysical Union journals, etc.) that serve to encourage the practice.[51] Declarations from scholarly groups in support of open data further contribute to the movement. The Panton Principles, developed by the Open Knowledge Foundation, provide guidance on licensing data publications within the public domain based on the premise that "for science to effectively function, and for society to reap the full benefits from scientific endeavours, it is crucial that science data be made open."[52] Taking a wider view, the Denton Declaration asserts the value of promoting "collaboration, transparency, and accountability across organizational and disciplinary boundaries" in best practices for data management in support of open access to data.[53]

The scholarly communication system is shifting under the weight of the open access movement. Open data places a unique strain as a non-traditional format, which substantiates the need for system-wide change and adaptation in regards not to just the economics of scholarly publications, but to the very nature of communication formats and technologies and the social ecosystem of communication.[54] Political, economic, and social concerns over access to data are evidenced by advances in policy. These are the forces pushing the scholarly communication system toward integrating data.

Data Sharing Realities

Despite the existence of strong rhetoric in favor of data sharing and publication, the status quo in most fields is low rates of data sharing and publication.[55] A disconnect exists between data sharing ideals and reality. Even when authors publish in journals with data sharing policies, most authors do not actually make their data available.[56] In a study of a group of scientists' data practices and perceptions, Tenopir et al.[57] found that 67% of respondents agree that lack of access to data is an impediment to scientific progress, yet just 36% agreed that others can easily access their data and only 6% make all of their data available for use by others. When it does happen, informal data sharing is more likely to occur than formal publication.[58] The ideal of openness and transparency in research is widely held by researchers, but normative practice does not include regular incorporation of data into formal scholarly communication outputs.

Publishing data has not been a universal feature of scholarly communication practice in the same way as publishing journal articles, books, and conference papers. Until recently, communication technologies lacked the ability to easily facilitate data storage and transfer. Widespread integration of data sharing as a standard feature of the scholarly communication process will take time as change is needed to overcome a variety of challenges. Whereas some academic communities have been early adopters and can easily see value from the open exchange of data (take for example, the adoption of the Bermuda Principle supporting the sharing of human genome sequence data[59]), others may have valid reasons for not expanding their published communications to include data. A key finding of the Research Information Network's report on data sharing is the extent to which behaviors vary by discipline; research cultures, standardization of common data types, and the concordant availability of data archives are major factors in determining the viability of open data within particular research domains.[60]

Reasons for lack of data sharing and publication range across social and technical issues. Reluctance to share and publish data includes the following considerations:[61]

- *Data sharing norms*. Many research cultures rely primarily on informal data sharing, which allows for control over who can gain access and creates collaborative opportunities for further research. In some cases, applications for data reuse may not be apparent to data creators. Published data brings a loss of control and fear that data may be misused or misinterpreted. Academic competitiveness is a concern: when rewards and recognition accrue only to traditional publications, there are no perceived benefits to sharing data.
- *Inadequate support for data management*. Preparing data for sharing and publication takes a considerable investment of time and effort. Limited knowledge or actual lack of metadata standards and other best practices for data documentation inhibit the ability of researchers to prepare data for sharing and publication.
- *Inadequate infrastructure and publication mechanisms*. Lack of publication options. Data repositories do not exist for all data types in all fields. Descriptive metadata standards may not exist to adequately make data visible.
- *Legal and ethical concerns*. Data may be of a sensitive nature and unable to be shared due to privacy and confidentiality concerns. Institutional policies governing ownership of data may not be clear.

Getting There: Data Publication Practices

Some of the data sharing obstacles are easier to solve than others. Embargoes and timelines for data publication can be set so that researchers are able to fully exploit the fruits of their labor before data are released. Sensitive data can be released through archives that handle restricted data access protocols, such as vetting research projects and enforcing data protection plans. However, these are details underlying the larger paradigm shift necessary to create a social norm of data publication within academic cultures and supplementing scholarly communication infrastructures to accommodate data.

As previously discussed, data publication is a fuzzy concept, complicating the ability to arrive at a straightforward approach to implementation. Publishing data, rather than informally sharing it, provides legitimizing factors that allow for the recognition of data contributions as first-class scholarly objects. As Kratz and Strasser explain, for data to be considered published it must be made publicly available, it must be adequately documented to support reproduction and reuse, and it must be citable.[62] These objectives are obtained through a publication process that includes deposition (as with a repository), description, and assignation of a persistent identifier. Additionally, data publications may be subject to validation (as through peer review), although the feasibility and methodology for data validation, as well as its necessity for publication remain a matter of open debate.[63] There are several models of data publication currently in use. Data may be published via deposit into a digital repository, or with journals as supplementary material to an article, or as a data paper.[64] These publication options are discussed further in the remainder of this section.

A *digital repository* provides an infrastructure for storage, management, retrieval, and curation of digital materials.[65] There is wide variation in the availability of digital repositories for data. Data may be included as one of several content types, as within an *institutional repository*. A *data repository* implies dedication to housing data as the primary content type. The scope of a data repository can be institutional, disciplinary, or multidisciplinary. The term *data archive* comes from the social sciences and refers to organizations dedicated to the long-term preservation and curation of data in order to make them available for secondary analysis; archives may be disciplinary or national in scope.[66]

Among the various options for publishing, the data archive is the most long-standing option for data publication and best fulfills the criteria for formal data publication. In addition, disciplinary data archives afford the ability to specialize in a particular type of data and can serve as a community hub by acting as a bastion for development of standards (for data formatting, metadata, and documentation) and best practices in research methodologies, data management, and sharing cultures. For example, the Inter-university Consortium for Political and Social Research was founded in 1962, curates and preserves datasets, actively supports the Data Documentation Initiative for social survey metadata standards, engages in community outreach for data management best practices, and hosts the Summer Program in Quantitative Methods of Social Research. A more recent data archiving model in the earth sciences is the Data Observation Network for Earth (DataONE), which is a unifying coordinator for a distributed network of data centers. DataONE provides a common search interface to link data archives, provides guidance on data citation and best practices for data management, and offers resources and educational opportunities to support increased community engagement in data sharing and publication activities.[67]

Whereas data archives often specialize in a disciplinary area, institutional repositories gather research outputs from across the entire range of research produced by their home universities. Institutional repositories have been built primarily to accommodate traditional publication formats, although some universities are beginning to include data collections within the institutional repository or create standalone data repositories. Heterogeneity of data types and formats presents a challenge to the ability of the institutional repository to successfully accommodate collection and curation of research data. However, when a disciplinary data archive is not available for a particular data type, the multidisciplinary nature of institutional repositories allows them to provide a home for data publication. Private data repository services, such as Figshare, also seek to fulfill the need for multidisciplinary data publication, in addition to providing a low-barrier method for researchers to make data easily available. The issue of adequately curating heterogeneous data remains problematic for multidisciplinary repositories. For example, ensuring that sufficient documentation is provided to describe the data is difficult to accomplish without subject area and research methodology expertise.[68]

Rather than attempting to independently handle all possible data outputs, institutional repositories may find that building partnerships with data archives provides a better option for supporting disciplinary data cultures and taking advantage of specialized data curation knowledge and practices.[69] For example, a librarian at a college or university may be able to work with faculty to identify data to add to the library collection, but may lack the resources or expertise to recover and convert old software formats, review and clean the data, or guarantee preservation for all file types. A data archive may be able to provide tools and expert advice to assist the institutional repository staff, or add the dataset directly to their collection.[70] Additionally, Akers and Green provide a case study of a library leveraging a membership with a disciplinary data repository to provide financial assistance toward publication fees, integrating data publication with the university press, and offering local assistance and promotion.[71]

Data publishing with journals, as supplementary material or as a *data paper*, can be problematic for achieving publication of data as first-class scholarly objects. Providing data as supplementary material to a journal article does not achieve the desired effect of a standalone data publication. Furthermore, supplementary data are usually provided only for replication purposes and do not offer a complete and fully documented dataset for original reuse in a secondary analysis study. While a data paper typically provides in-depth description of a dataset (without engaging in analysis), it reinforces the standard of the journal article as the desired publication mechanism, keeping data as a second-class citizen. Data papers are essentially "data publication by proxy" because they provide a straightforward bibliographic reference and allow authors to accrue credit, but they normally do not provide access to the actual dataset being described.[72]

Although publishing data along with articles was initially a relatively popular option with the move to electronic publishing and the corresponding removal of space limitations, there have long been issues with standardization for supplementary materials.[73] Journal publishing platforms are by and large not equipped to handle the specialized and heterogeneous needs for data curation. Recent developments provide evidence that journal publishers are moving away from directly hosting data as supplementary materials.[74] Some journal publishers have contracted with digital repositories to handle the display and storage of supplemental materials, such as with the partnership between Taylor & Francis journals and Figshare.[75] Another example is the Dryad data repository, which has over 80 integrated journals and focuses on publishing data underlying peer-reviewed articles in science and medicine. In the absence of formal partnerships, journal data sharing policies may guide authors towards trusted archives for common disciplinary data types. For example, the Geoscience Data Journal provides a list of almost 20 approved repositories covering the range of expected data types.[†] PLOS journals specify that authors are required to utilize field-specific standards and data repositories where they exist. That these policies provide a multiplicity of options for data repositories reflects the steps ahead in establishing data publication norms. The advantage to working with a third-party repository is that supplementary data can be given a separate unique identifier, tracked for reuse, and directly cited.

Additionally, although it is a positive step forward that many journal publishers have turned their attention to the creation of links between articles and supporting data (as per Elsevier[76]), links may stop short of full citation. This is especially true in the case of self-citation, wherein a journal article details the analysis of the author's own data as opposed to a secondary analysis study. An argument against self-citation is that it can disproportionately inflate citation counts.[77] However, for self-citation of other published material the standard practice for an entry is the reference list. Without self-citation, data are still treated as supplementary material to the article, rather than as a discrete publication, in that the data are not listed in the article's bibliography but rather linked directly in-text, via a side-bar, or mentioned in a note. Establishing links between journal articles and datasets is an important component of integration. Those links should serve to recognize data as a publication through inclusion of a complete data citation.

The key reason to treat data as a publication is to allow for it to find a place within existing scholarly communication structures and the paradigmatic norms of the scientific ethos. Importantly, publication allows for citation. Citation is a chief characteristic of scholarly writing. It is the mechanism that establishes priority claims and enables reward. Data citation supports the parallel treatment of literature and data as equal scholarly outputs. This includes the ability to create metrics around data reuse, which has the potential to incentivize data publication by measuring usage and impact for researchers, their institutions, and funding bodies.[78]

Although there have been mechanisms for data publication and citation for decades, only recently has there been widespread acknowledgement within the publishing, library, archival, and data communities of its importance.[79] This is evidenced by the recent adoption of the 2014 Joint Declaration of Data Citation

† This list of approved repositories does include multidisciplinary options that accept all data types, such as Figshare and Zenodo.

Principles, which has been endorsed by over 90 scholarly organizations.[80] The importance of data citation is also acknowledged in legislative policy discussions[81] and government open data directives.[82] Still, there is significant change to be made in scientific culture as long-standing status-quo communication practices largely ignore formal data citation.[83] Even where data are formally published, the largely informal data sharing cultures are reinforced through informal citation practices—namely the passing in-text mention or a brief statement in an author's note or acknowledgements note.[84]

Advances in data publishing and data citation go hand-in-hand. Scientists indicate that they would be more likely to publish data if they were assured citation.[85] Implementation of data citation requires that data publication venues build systems that provide adequate metadata, unique identification, and persistence guarantees.[86] Tools such as the DataCite program to provide data repositories with Digital Object Identifiers and establish metadata standards for datasets are helping to create the infrastructure necessary to support citation.[87] The Data Citation Index (DCI) essentially provides the application of bibliographic citation indexing and metrics to data, supporting the integration of data into the scholarly record. The DCI enables data discovery, data citation, and examination of current data publication and citation practices.[88] By indexing data from repositories, it recognizes the primacy of the data repository as the preferred method of data publication (rather than supplementary materials and data papers). Although the DCI may be somewhat premature given the current state of citation practices, the database was developed in part to serve as a driver for change.[89] Although research data publication and citation infrastructure and practices are still in early stages, there are sufficient tools and models to continue moving forward.

Steps for Libraries

The place of data within the scholarly record is in flux. We are in a period of rapid development that will likely continue for some time before infrastructures, policies, scientific norms, and cultural practices are settled. Next steps in supporting the full integration of data into the scholarly communication system include sustained and increased attention to constructing infrastructures (both technical and human), including positioning libraries to build data collections.

Policies in support of data sharing and publication require clear and accessible routes to data publication. However, robust and established data publication options are not available across all domains. The current system is in the unfortunate position of embodying the chicken/egg conundrum: people are not going to publish data until sufficient mechanisms for data publication are in place.[90] Existing policies tend towards vagueness in acknowledgment of the present reality that not all fields have established norms for the publication of data. For example, the

NSF offers this guidance in response to the question, "Am I required to deposit my data in a public database?":

> What constitutes reasonable data management and access will be determined by the community of interest through the process of peer review and program management. In many cases, these standards already exist, but are likely to evolve as new technologies and resources become available.[91]

The *PLOS* journal policy on data sharing is similarly open-ended, requiring that "authors comply with field-specific standards for preparation and recording of data and select repositories appropriate to their field" and encourages authors to "select repositories that meet accepted criteria as trustworthy digital repositories."[92] Although these are laudable policies, the difficulty of navigating the data publication landscape make compliance a struggle when even the policy-makers cannot specify how to best accomplish data publication. In some cases, data librarians may be better informed regarding the range of data sharing and publication options than faculty researchers, putting librarians in a valuable advisory role.

The nascent data publication systems require investment, both inside and outside of libraries. In order to support data management and open data policy directives, there must be investment in both the technical and human infrastructures required to facilitate data sharing and publication. The focus on cyberinfrastructure development in the United States by the NSF includes the vision to create a national digital data framework that would support development of community standards for data management and build architectures for data collections across many different stakeholders, including recognition of the role of digital libraries within colleges and universities to house data produced by their faculty.[93] This has manifested in the NSF's data management plan requirement for grant proposals and funding for programs through the Sustainable Digital Data Preservation and Access Network Partners program (better known as DataNet).[94] For example, the Data Conservancy at the Sheridan Libraries at Johns Hopkins University is a DataNet funded program that provides data management and curation services. Vision, however, has not yet caught up to reality. There is a perception that data management and curation is an unfunded mandate as many researchers still lack institutional support.[95]

Viable economic models for setting up and maintaining data publishing infrastructures are yet to be fully developed. The U.S. federal mandate from the OSTP for government agencies to provide public access to publications and data did not come with any additional funding to accomplish this measure.[96] Research libraries also have competing demands on the allocation of their resources, although many see involvement in research data as a growth area and endorse the NSF value proposition for the role of digital libraries in creating a digital data framework.

Setting up and maintaining data publishing infrastructures is a challenge that requires participation from stakeholders across public and private sectors. Cooperative efforts like the Research Data Alliance are working toward solutions to create and enhance technical and social infrastructures in support of data sharing. Areas of concentration include development of guidance around data citation, persistent identifiers, shared metadata frameworks, common organizational policies and practice, data standards, and shared best practices for data access and preservation.[97] Academic libraries can both extend the reach of their impact and gain support in building new programs by participating in relevant interest groups and associations (e.g., International Association for Social Science Information Services & Technology), as well as partnering with data archives.

Libraries can support data publication through expanding the scope of their services to engage with our community members not just as readers, but as authors and producers of data and information.[98] Publishing data is the final step at the end of the data management process for researchers. Efforts towards data management education, integrating data management into research ethics instruction (e.g., Responsible Conduct of Research programs), and embedding with research teams are all in the service of facilitating the progression of behavioral norms in scholarly communication to include data sharing and publication.

Data librarians can work to build collections of faculty data and reposition existing collections to support data intensive research. Libraries may find that they already hold significant collections of data as part of their legacy collections. For example, humanists consider the library their laboratory and our collections as their data. Digitizing special collections and serving them in formats amenable to computational analysis is one way for libraries to actively engage in data publication.[99] With regard to collecting faculty data, institutional repository infrastructures can be adapted or supplemented to include data.

Librarians should be aware that repository models that work for publications may not always work for data given the unique and varied nature of data publications.[100] In addition, flexible definitions of data can confound attempts at defining the scope of data collections.[101] Offering data publication services, as through an institutional repository, is likely to be a high-touch proposition which may require the development of additional infrastructure, processes, procedure, and expertise. Reports of pilot projects indicate that collection of faculty data involves a curation process which requires a significant time investment and challenges existing digital collection workflows.[102]

Library administrators and librarians should take heed from lessons learned in the development of scholarly communication and article-based institutional repository programs, as they speak to issues that will be faced in developing data collections and data management services. To point, the open access movement for literature was not immediately translated into wholesale use of institutional repositories; under-resourcing and a lack of supportive policies have been an impediment to success.[103] Libraries seeking to develop data collections should ensure an ongoing commitment to fund resources both human and technical as data librarians cannot succeed in isolation. This includes both the back-of-house repository support as well as outward facing interactions with faculty researchers. Subject area liaison librarians are important partners in vocalizing the centrality of library resources and services to all aspects of scholarship across the research lifecycle and integrating data into the library's overall collection building.[104]

Library-based data repositories should seek to track the development of standards for data publication. Starting points in the development of standards for data repositories are evident within the digital preservation community, where assessment measures for trusted digital repositories (i.e., Data Seal of Approval, Trusted Repositories Audit & Certification, and DRAMBORA) are now available.[105] These benchmarks provide the opportunity to elucidate best practices and certify the ability to provide long-term preservation and access to quality data. The implementation guidelines for repositories around metadata and identifiers stemming from the Joint Declaration on Data Citation Principles[106] demonstrate how ongoing discussion and initiatives from the scholarly community are also contributing to the development of best practices. Although disciplinary data archives are best suited to aid the creation of community-based standards within a particular niche area, library repositories will likely fill a need given the large diffusion of data types and practices and can apply broad-based standards to promote data publication and citation.

Conclusion

Data librarianship is a field that requires its members to keep on the forefront of current policies and practices. There are multiple stakeholders within the system of scholarly communication (e.g., authors, publishers, research funders), with libraries serving as an important part of the overall infrastructure, allowing librarians the opportunity to actively contribute to shape developments in scholarly communication concerning the role of data. The open data movement is pushing researchers toward data publication as they attempt to comply with policy mandates. The library can position itself as an integral part of the institutional ecosystem as universities work to develop programs and policies to meet researcher needs for grant funder and publishing requirements. An understanding of the complexities of the scholarly communication system, including scientific knowledge and data sharing practices, the nature of research data and data publication, and funder and journal policy requirements, will aid the academic data librarian as they navigate local institutional environments and endeavor to support their research communities. These are the broader forces that impact our day-to-day work.

1. Joan M. Reitz, *Online Dictionary for Library and Information Science* (Westport, CT: Libraries Unlimited, 2004), http://www.abc-clio.com/ODLIS/odlis_A.aspx.
2. Kim Duckett and Scott Warren, "Exploring the Intersections of Information Literacy and Scholarly Communication: Two Frames of Reference for Undergraduate Instruction," in *Common Ground at the Nexus of Information Literacy and Scholarly Communication*, ed. Stephanie Davis-Kahl and Merinda Kaye Hensley (Chicago: Association of College and Research Libraries, 2013), 41.
3. Robert K. Merton, "The Normative Structure of Science," in *The Sociology of Science: Theoretical and Empirical Investigations* (Chicago: University of Chicago Press, 1973), 267–78.
4. Randall Frost, "Scientific Method," in *The Gale Encyclopedia of Science*, ed. K. Lee Lerner and Brenda Wilmoth Lerner, 5th ed., vol. 7 (Farmington Hills, MI: Gale, 2014), 3862–65.
5. Jennifer Rowley, "The Wisdom Hierarchy: Representations of the DIKW Hierarchy," *Journal of Information Science* 33, no. 2 (2007): 163–80, doi:10.1177/0165551506070706.
6. Samuelle Carlson and Ben Anderson, "What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use," *Journal of Computer-Mediated Communication* 12, no. 2 (2007): 635–51, doi:10.1111/j.1083-6101.2007.00342.x.
7. Ibid., 647.
8. David Ribes and Steven J. Jackson, "Data Bite Man: The Work of Sustaining a Long-Term Study," in *"Raw Data" Is an Oxymoron*, ed. Lisa Gitelman (Cambridge, MA: MIT Press, 2013), 147, http://ieeexplore.ieee.org/servlet/opac?bknumber=6451327.
9. Ibid., 164.
10. Ibid., 162.
11. Chaim Zins, "Conceptual Approaches for Defining Data, Information, and Knowledge," *Journal of the American Society for Information Science and Technology* 58, no. 4 (2007): 479–93, doi:10.1002/asi.20508.
12. R. L. Ackoff, "From Data to Wisdom," *Journal of Applied Systems Analysis* 16, no. 1 (1989): 3; Laura Wynholds, "Linking to Scientific Data: Identity Problems of Unruly and Poorly Bounded Digital Objects," *International Journal of Digital Curation* 6, no. 1 (2011): 218, doi:10.2218/ijdc.v6i1.183; National Academy of Sciences (U.S.), National Academy of Engineering, and Institute of Medicine (U.S.), *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age* (Washington, D.C: National Academies Press, 2009), 22; Office of Management and Budget, "Circular A-110 Revised 11/19/93 As Further Amended 9/30/99," *The White House*, September 30, 1999, http://www.whitehouse.gov/omb/circulars_a110/.
13. National Academy of Sciences (U.S.), National Academy of Engineering, and Institute of Medicine (U.S.), *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*, 22; Consultative Committee for Space Data Systems, "Reference Model for an Open Archival Information System (OAIS)" (National Aeronautics and Space Administration, June 2012), http://public.ccsds.org/publications/archive/650x0m2.pdf; National Science Board (U.S.);National Science Foundation (U.S.), *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century* (Washington, D.C. : National Science Foundation, 2005), 13, www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf.
14. Inter-university Consortium for Political and Social Research, "Glossary of Social Science Terms," *Inter-University Consortium for Political and Social Research*, accessed January 21, 2015, http://www.icpsr.umich.edu/icpsrweb/ICPSR/support/glossary; Reitz, *Online Dictionary for Library and Information Science*.
15. Thomson Reuters, "Repository Evaluation, Selection, and Coverage Policies for the Data Citation Index within the Thomson Reuters Web of Science," *Web of Science*, accessed January 21, 2015, http://wokinfo.com//products_tools/multidisciplinary/dci/selection_essay/.
16. *Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century*.
17. Ibid., Appendix D.

18. Wynholds, "Linking to Scientific Data," 218.
19. M. A. Parsons and P. A. Fox, "Is Data Publication the Right Metaphor?," *Data Science Journal* 12 (2013): WDS32–46, doi:10.2481/dsj.WDS-042.
20. Christine L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World* (Cambridge, MA: MIT Press, 2015).
21. Kathleen Fitzpatrick, *Planned Obsolescence: Publishing, Technology, and the Future of the Academy* (New York: New York University Press, 2011).
22. Christine L Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* (Cambridge, MA: MIT Press, 2007).
23. Clifford Lynch, "Jim Gray's Fourth Paradigm and the Construction of the Scientific Record," in *The Fourth Paradigm: Data-Intensive Scientific Discovery*, ed. Tony Hey, Stewart Tansley, and Kristin Tolle (Microsoft Research, 2009), http://research.microsoft.com/en-us/collaboration/fourthparadigm/contents.aspx.
24. "Budapest Open Access Initiative," February 14, 2002, http://www.budapestopenaccessinitiative.org/read.
25. J. Klump, R. Bertelmann, J. Brase, M. Diepenbroek, H. Grobe, H. Höck, M. Lautenschlager, U. Schindler, I. Sens, and J. Wächter, "Data Publication in the Open Access Initiative," *Data Science Journal* 5 (2006): 79–83, doi:10.2481/dsj.5.79.
26. Stephen E. Fienberg, Margaret E. Martin, and Miron L. Straf, eds., *Sharing Research Data* (Washington, D.C: National Academy Press, 1985).
27. Paula E. Stephan, "Robert K. Merton's Perspective on Priority and the Provision of the Public Good Knowledge," *Scientometrics* 60, no. 1 (2004): 81–87, doi:10.1023/B:SCIE.0000027311.17226.70.
28. Michael Yamaner, "Federal Funding for Basic Research at Universities and Colleges Essentially Unchanged in FY 2012," InfoBrief (Arlington, VA: National Science Foundation, National Center for Science and Engineering Statistics, September 2014), http://www.nsf.gov/statistics/infbrief/nsf14318/.
29. "FedStats," accessed January 27, 2015, http://fedstats.sites.usa.gov/.
30. Bart Ragon, "The Political Economy of Federally Sponsored Data," *Journal of eScience Librarianship* 2, no. 2 (2013), doi:10.7191/jeslib.2013.1050.
31. Eric A. Fischer, "Public Access to Data from Federally Funded Research: Provisions in OMB Circular A-110," Congressional Research Service, (2013), HTTP://congressional.proquest.com.proxy2.cl.msu.edu/congressional/docview/t21.d22.crs-2013-rsi-0116?accountid=12598.
32. John Holdren, "Increasing Access to the Results of Federally Funded Scientific Research" (Office of Science and Technology Policy, February 22, 2013), http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
33. Dianne Dietrich, Trisha Adamus, Alison Miner, and Gail Steinhart, "De-Mystifying the Data Management Requirements of Research Funders," *Issues in Science and Technology Librarianship* 70, no. 1 (2012), doi:10.5062/F44M92G2.
34. National Science Foundation, "Grant Proposal Guide" (National Science Foundation, January 2013), http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp.
35. National Institutes of Health, Office of Extramural Research, "NIH Data Sharing Policy and Implementation Guidance," March 5, 2003, http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
36. Organisation for Economic Co-operation and Development, *OECD Principles and Guidelines for Access to Research Data from Public Funding* (Paris: OECD Publishing, 2007), 9, http://www.oecd-ilibrary.org/content/book/9789264034020-en-fr.
37. Alma Swan, "Policy Guidelines for the Development and Promotion of Open Access" (United Nataions Educational, Scientific and Cultural Organization, 2012), http://unesdoc.unesco.org/images/0021/002158/215863e.pdf.

38. European Commission, “Commission Recommendation of 17.7.2012 on Access to and Preservation of Scientific Information” (European Commission, July 2012), 3, https://ec.europa.eu/digital-agenda/node/66216.
39. Research Councils UK, “RCUK Common Principles on Data Policy,” accessed January 28, 2015, http://www.rcuk.ac.uk/research/datapolicy/.
40. Fischer, “Public Access to Data from Federally Funded Research: Provisions in OMB Circular A-110”; Richard Shelby, “Accountability and Transparency: Public Access to Federally Funded Research Data,” *Harvard Journal on Legislation* 37 (2000): 369–89.
41. Shelby, “Accountability and Transparency,” 379.
42. Government Accountability Office, “Climate Change Research: Agencies Have Data-Sharing Policies but Could Do More To Enhance the Availability of Data from Federally Funded Research,” 2007, http://www.gao.gov/products/GAO-07-1172; David Goldston, “Big Data: Data Wrangling,” *Nature News* 455, no. 7209 (September 3, 2008): 15–15, doi:10.1038/455015a.
43. National Academy of Sciences (U.S.), National Academy of Engineering, and Institute of Medicine (U.S.), *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age* (Washington, D.C: National Academies Press, 2009), 63.
44. Subcommittee on Research; Committee on Science, Space, and Technology. House, *Scientific Integrity and Transparency*, 2013, http://www.gpo.gov/fdsys/pkg/CHRG-113hhrg79929/pdf/CHRG-113hhrg79929.pdf.
45. Hans E. Roosendall and Peter A. Th. M. Geurts, “Forces and Functions in Scientific Communication: An Analysis of Their Interplay” (CRISP ’97, Cooperative Research Information Systems in Physics, Oldenburg, Germany, 1997), http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html.
46. J. Klump et al., “Data Publication in the Open Access Initiative,” *Data Science Journal* 5 (2006): 79–83, doi:10.2481/dsj.5.79; Heinz Pampel and Sünje Dallmeier-Tiessen, “Open Research Data: From Vision to Practice,” in *Opening Science*, ed. Sönke Bartling and Sascha Friesike (Springer International Publishing, 2014), 213–24, http://link.springer.com/chapter/10.1007/978-3-319-00026-8_14.
47. National Academy of Sciences (U.S.), National Academy of Engineering, and Institute of Medicine (U.S.), *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*; Subcommittee on Research; Committee on Science, Space, and Technology. House, *Scientific Integrity and Transparency*.
48. Beth A. Fischer and Michael J. Zigmond, “The Essential Nature of Sharing in Science,” *Science and Engineering Ethics* 16, no. 4 (November 2010): 783–99, doi:10.1007/s11948-010-9239-x; Patricia A. Soranno, Kendra S. Cheruvelil, Kevin C. Elliott, and Georgina M. Montgomery, “It’s Good to Share: Why Environmental Scientists’ Ethics Are out of Date,” *BioScience* 65, no. 1 (2014): 69–73, doi:10.1093/biosci/biu169; John Willinsky and Juan Pablo Alperin, “The Academic Ethics of Open Access to Research and Scholarship,” *Ethics and Education* 6, no. 3 (2011): 217–23, doi:10.1080/17449642.2011.632716.
49. see for example Daniel S. Caetano and Anita Aisenberg, “Forgotten Treasures: The Fate of Data in Animal Behaviour Studies,” *Animal Behaviour* 98 (2014): 1–5, doi:10.1016/j.anbehav.2014.09.025; Fischer and Zigmond, “The Essential Nature of Sharing in Science”; Russell A. Poldrack and Krzysztof J. Gorgolewski, “Making Big Data Open: Data Sharing in Neuroimaging,” *Nature Neuroscience* 17, no. 11 (November 2014): 1510–17, doi:10.1038/nn.3818.
50. Barbara Schneider, “Building a Scientific Community: The Need for Replication,” *The Teachers College Record* 106, no. 7 (2004): 1471–83; Joan E. Sieber, “Will the New Code Help Researchers to Be More Ethical?,” *Professional Psychology: Research and Practice* 25, no. 4 (1994): 369, doi:10.1037/0735-7028.25.4.369.
51. Katherine W. McCain, “Mandating Sharing Journal Policies in the Natural Sciences,” *Science Communication* 16, no. 4 (June 1, 1995): 403–31, doi:10.1177/1075547095016004003; Fiona Murphy, “Data and Scholarly Publishing: The Transforming Landscape,” *Learned Publishing* 27, no. 5 (2014): 3–7, doi:10.1087/20140502; Hazel Norman, “Mandating Data Archiving: Experi-

ences from the Frontline," *Learned Publishing* 27, no. 5 (2014): 35–38, doi:10.1087/20140507; Michael C. Whitlock, Mark A. McPeek, Mark D. Rausher, Loren Rieseberg, and Allen J. Moore, "Data Archiving," *The American Naturalist* 175, no. 2 (2010): 145–46, doi:10.1086/650340.

52. Peter Murray-Rust, Cameron Neylon, Rufus Pollock, and John Wilbanks, "Panton Principles: Principles for Open Data in Science," February 19, 2010, http://pantonprinciples.org/.
53. Spencer D. C. Keralis, "Denton Declaration: An Open Data Manifesto," May 22, 2012, http://openaccess.unt.edu/denton-declaration.
54. Force11, "Force11 White Paper: Improving the Future of Research Communication and E-Scholarship," ed. Phil E. Bourne, Tim Clark, Robert Dale, Anita de Waard, Ivan Herman, Eduard Hovy, and David Shotton, October 28, 2011, https://www.force11.org/white_paper.
55. Patrick Andreoli-Versbach and Frank Mueller-Langer, "Open Access to Data: An Ideal Professed but Not Practised," *Research Policy* 43, no. 9 (November 2014): 1621–33, doi:10.1016/j.respol.2014.04.008; Bryn Nelson, "Data Sharing: Empty Archives," *Nature News* 461, no. 7261 (September 9, 2009): 160–63, doi:10.1038/461160a; Maggie Puniewska, "Scientists Have a Sharing Problem," *The Atlantic*, December 15, 2014, http://www.theatlantic.com/health/archive/2014/12/scientists-have-a-sharing-problem/383061/.
56. Alawi A. Alsheikh-Ali et al., "Public Availability of Published Research Data in High-Impact Journals," *PLoS ONE* 6, no. 9 (September 7, 2011): e24357, doi:10.1371/journal.pone.0024357; Caroline J. Savage and Andrew J. Vickers, "Empirical Study of Data Sharing by Authors Publishing in PLoS Journals," *PLoS ONE* 4, no. 9 (September 18, 2009): e7078, doi:10.1371/journal.pone.0007078; J.M. Wicherts, D. Borsboom, J. Kats, and D. Molenaar, "The Poor Availability of Psychological Research Data for Reanalysis.," *American Psychologist* 61, no. 7 (2006): 726, doi:10.1037/0003-066X.61.7.726.
57. Carol Tenopir, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame, "Data Sharing by Scientists: Practices and Perceptions," *PLoS ONE* 6, no. 6 (2011): e21101, doi:10.1371/journal.pone.0021101.
58. Christine L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World* (Cambridge, MA: MIT Press, 2015); Amy M. Pienta, George C. Alter, and Jared A. Lyle, "The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data," November 22, 2010, http://hdl.handle.net/2027.42/78307; Jillian C. Wallis, Elizabeth Rolando, and Christine L. Borgman, "If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology," *PLoS ONE* 8, no. 7 (2013): e67332, doi:10.1371/journal.pone.0067332.
59. Eliot Marshall, "Bermuda Rules: Community Spirit, With Teeth," *Science* 291, no. 5507 (February 16, 2001): 1192–1192, doi:10.1126/science.291.5507.1192.
60. Aaron Griffiths, "The Publication of Research Data: Researcher Attitudes and Behaviour," *International Journal of Digital Curation* 4, no. 1 (2009), doi:10.2218/ijdc.v4i1.77; Research Information Network, "To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs," June 2008, http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf.
61. Djoko Sigit Sayogo and Theresa A. Pardo, "Exploring the Determinants of Scientific Data Sharing: Understanding the Motivation to Publish Research Data," *Government Information Quarterly* 30 (2013): S19–31, doi:10.1016/j.giq.2012.06.011; Suenje Dallmeier-Tiessen, Robert Darby, Kathrin Gitmans, Simon Lambert, Brian Matthews, Salvatore Mele, Jari Suhonen, and Michael Wilson, "Enabling Sharing and Reuse of Scientific Data," *New Review of Information Networking* 19, no. 1 (2014): 16–43, doi:10.1080/13614576.2014.883936; Research Information Network, "To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs"; Poldrack and Gorgolewski, "Making Big Data Open"; Benedikt Fecher, Sascha Friesike, and Marcel Hebing, "What Drives Academic Data Sharing?," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, May 1, 2014), http://papers.ssrn.com/abstract=2439645.
62. John Kratz and Carly Strasser, "Data Publication Consensus and Controversies," *F1000Research*, October 16, 2014, doi:10.12688/f1000research.3979.3.

63. Ibid.; Mark A. Parsons, Ruth Duerr, and Jean-Bernard Minster, "Data Citation and Peer Review," *Eos, Transactions American Geophysical Union* 91, no. 34 (2010): 297–98, doi:10.1029/2010EO340001.
64. Don MacMillan, "Data Sharing and Discovery: What Librarians Need to Know," *Journal of Academic Librarianship* 40, no. 5 (2014): 541–49, doi:10.1016/j.acalib.2014.06.011.
65. Najla Semple, "Digital Repositories," DCC Briefing Papers: Introduction to Curation (Edinburgh: Digital Curation Centre, 2006), http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/digital-repositories.
66. Louise Corti, "Data Archive," in *The SAGE Encyclopedia of Qualitative Research Methods*, by Lisa Given (Thousand Oaks, CA: SAGE Publications, 2008), http://dx.doi.org/10.4135/9781412963909.
67. Suzie Allard, "DataONE: Facilitating eScience through Collaboration," *Journal of eScience Librarianship* 1, no. 1 (February 2012), doi:10.7191/jeslib.2012.1004.
68. Kratz and Strasser, "Data Publication Consensus and Controversies"; Gail Steinhart, "Partnerships between Institutional Repositories, Domain Repositories and Publishers," *Bulletin of the American Society for Information Science and Technology* 39, no. 6 (2013): 19–22, doi:10.1002/bult.2013.1720390608.
69. Ann G. Green and Myron P. Gutmann, "Building Partnerships among Social Science Researchers, Institution-based Repositories and Domain Specific Data Archives," *OCLC Systems & Services: International Digital Library Perspectives* 23, no. 1 (2007): 35–53, doi:10.1108/10650750710720757.
70. Lyle, Alter, and Green, "Partnering to Curate and Archive Social Science Data."
71. Katherine G. Akers and Jennifer A. Green, "Towards a Symbiotic Relationship Between Academic Libraries and Disciplinary Data Repositories: A Dryad and University of Michigan Case Study," *International Journal of Digital Curation* 9, no. 1 (2014), doi:10.2218/ijdc.v9i1.306.
72. Kratz and Strasser, "Data Publication Consensus and Controversies."
73. Todd Carpenter, "Standards Column—Journal Article Supplementary Materials: A Pandora's Box of Issues Needing Best Practices," *Against the Grain* 21, no. 6 (November 1, 2013), http://docs.lib.purdue.edu/atg/vol21/iss6/6.
74. Kratz and Strasser, "Data Publication Consensus and Controversies"; MacMillan, "Data Sharing and Discovery."
75. Taylor & Francis, "Taylor & Francis and Figshare: Making Research More Discoverable," January 2014, http://newsroom.taylorandfrancisgroup.com/news/press-release/taylor-francis-and-figshare-making-research-more-discoverable#.VNFEKy558mE.
76. Elsevier, "Database Linking Tool," accessed February 4, 2015, http://www.elsevier.com/about/content-innovation/database-linking.
77. Rodrigo Costas et al., "The Value of Research Data: Metrics for Datasets from a Cultural and Technical Point of View" (Knowledge Exchange, April 2013), http://www.knowledge-exchange.info/datametrics.
78. Ibid.
79. Micah Altman and Mercè Crosas, "The Evolution of Data Citation: From Principles to Implementation," *IASSIST Quarterly* 37, no. 1–4 (March 2013): 62–70, http://www.iassistdata.org/downloads/iqvol371_4_altman.pdf.
80. Data Citation Synthesis Group, "Joint Declaration of Data Citation Principles," *Force11*, 2014, http://www.force11.org/datacitation.
81. Subcommittee on Research; Committee on Science, Space, and Technology. House, *Scientific Integrity and Transparency*.
82. Holdren, "Increasing Access to the Results of Federally Funded Scientific Research."
83. Joan E. Sieber and Bruce E. Trumbo, "(Not) Giving Credit Where Credit Is Due: Citation of Data Sets," *Science and Engineering Ethics* 1, no. 1 (1995): 11–20, doi:10.1007/BF02628694.
84. Hailey Mooney and Mark Newton, "The Anatomy of a Data Citation: Discovery, Reuse, and Credit," *Journal of Librarianship and Scholarly Communication* 1, no. 1 (2012), doi:10.7710/2162-3309.1035.

85. Tenopir et al., "Data Sharing by Scientists."
86. Joan Starr, Eleni Castro, Merce Crosas, Michel Dumontier, Robert R. Downs, Ruth Duerr, Laurel L. Haak, Melissa Haendel, and Ivan Herman, "Achieving Human and Machine Accessibility of Cited Data in Scholarly Publications," *PeerJ PrePrints*, 2015, https://peerj.com/preprints/697/.
87. Joan Starr and Angela Gastl, "isCitedBy: A Metadata Scheme for DataCite," *D-Lib Magazine* 17, no. 1/2 (2011), doi:10.1045/january2011-starr.
88. Megan M. Force and Nigel J. Robinson, "Encouraging Data Citation and Discovery with the Data Citation Index," *Journal of Computer-Aided Molecular Design* 28, no. 10 (October 1, 2014): 1043–48, doi:10.1007/s10822-014-9768-5; Nicolas Robinson-García, Evaristo Jiménez-Contreras, and Daniel Torres-Salinas, "Analyzing Data Citation Practices Using the Data Citation Index," *Journal of the Association for Information Science and Technology*, 2015, doi:10.1002/asi.23529.
89. Force and Robinson, "Encouraging Data Citation and Discovery with the Data Citation Index."
90. Nelson, "Data Sharing."
91. National Science Foundation, Office of Budget, Finance and Award Management, "Data Management & Sharing Frequently Asked Questions."
92. PLOS, "PLOS Editorial and Publishing Policies," accessed February 5, 2015, http://www.ploscompbiol.org/static/policies#sharing.
93. National Science Foundation, Cyberinfrastructure Council, "Cyberinfrastructure Vision for 21st Century Discovery," March 2007, http://www.nsf.gov/pubs/2007/nsf0728/.
94. Dharma Akmon, "NSF DataNet Partners Update," *Bulletin of the Association for Information Science and Technology* 40, no. 6 (September 2014): 22–25. http://www.asis.org/Bulletin/Aug-14/AugSep14_Akmon.html.
95. Science Staff, "Challenges and Opportunities," *Science* 331, no. 6018 (February 11, 2011): 692–93, doi:10.1126/science.331.6018.692.
96. Francine Berman and Vint Cerf, "Who Will Pay for Public Access to Research Data?," *Science* 341, no. 6146 (August 9, 2013): 616–17, doi:10.1126/science.1241625.
97. Fran Berman, Ross Wilkinson, and John Wood, "Building Global Infrastructure for Data Sharing and Exchange Through the Research Data Alliance," *D-Lib Magazine* 20, no. 1/2 (January 2014), doi:10.1045/january2014-berman.
98. Kara J. Malenfant, "Leading Change in the System of Scholarly Communication: A Case Study of Engaging Liaison Librarians for Outreach to Faculty," *College & Research Libraries* 71, no. 1 (2010): 63–76, doi:10.5860/crl.71.1.63; Todd Bruns, Steve Brantley, and Kirstin Duffin, "Scholarly Communication Coaching: Liaison Librarians' Shifting Roles," in *The 21st Century Library: Partnerships and New Roles*, ed. Brad Eden (Lanham, Maryland: Littlefield and Rowman/Scarecrow Publishing, 2015), http://works.bepress.com/steve_brantley/22.
99. Thomas G. Padilla and Devin Higgins, "Library Collections as Humanities Data: The Facet Effect," *Public Services Quarterly* 10, no. 4 (2014): 324–35, doi:10.1080/15228959.2014.963780.
100. Dorothea Salo, "Retooling Libraries for the Data Challenge," *Ariadne*, no. 64 (July 2010), http://www.ariadne.ac.uk/issue64/salo/.
101. Hans Jørn Nielsen and Birger Hjørland, "Curating Research Data: The Potential Roles of Libraries and Information Professionals," *Journal of Documentation* 70, no. 2 (2014): 221–40, doi:10.1108/JD-03-2013-0034; Borgman, *Big Data, Little Data, No Data*.
102. David Minor, Matt Critchlow, Arwen Hutt, Declan Fleming, Mary Linn Bergstrom, and Don Sutton, "Research Data Curation Pilots: Lessons Learned," *International Journal of Digital Curation* 9, no. 1 (June 17, 2014): 220–30, doi:10.2218/ijdc.v9i1.313; Tyler O. Walters, "Data Curation Program Development in US Universities: The Georgia Institute of Technology Example," *International Journal of Digital Curation* 4, no. 3 (2009): 83–92; Lisa R. Johnston, "A Workflow Model for Curating Research Data in the University of Minnesota Libraries: Report from the 2013 Data Curation Pilot" (University of Minnesota Digital Conservancy, January 19, 2014), http://hdl.handle.net/11299/162338; Mark P Newton, C. C. Miller, and Marianne Stowell Bracke, "Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a

Research Library Task Force," *Collection Management* 36, no. 1 (2010), doi:10.1080/01462679.2011.530546.
103. Dorothea Salo, "Innkeeper at the Roach Motel," *Library Trends* 57, no. 2 (2008): 98–123, doi:10.1353/lib.0.0031; Dorothea Salo, "How to Scuttle a Scholarly Communication Initiative," *Journal of Librarianship and Scholarly Communication* 1, no. 4 (2013), doi:10.7710/2162-3309.1075.
104. Bruns, Brantley, and Duffin, "Scholarly Communication Coaching."
105. Inter-university Consortium for Political and Social Research, "Trusted Digital Repositories," accessed February 5, 2015, http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/trust.html.
106. Starr et al., "Achieving Human and Machine Accessibility of Cited Data in Scholarly Publications."