

Prospects for Research Data Management

By: [Martin Halbert](#)

“Prospects for Research Data Management,” in *Research Data Management: Principles, Practices, and Prospects*. Council on Library and Information Resources, 2013, pp. 1-15. URL: <http://www.clir.org/pubs/reports/pub160>



Research Data Management: Principles, Practices, and Prospects by the Council on Library and Information Resources is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](#).

Abstract:

The challenge of ensuring long-term preservation of and access to the outputs of scientific research, especially data sets produced by publicly funded research projects, has become a prominent topic in the United States. In 2011, the two-year DataRes Project was initiated at the University of North Texas to document perceptions and responses to this emerging challenge in U.S. higher education and to explore ways in which the library and information science (LIS) profession could best respond to the need for better research data management in universities. This chapter will highlight some of the most provocative findings of the DataRes Project on the topic of research data management in higher education and then consider possible research data management (RDM) scenarios for the future and the implications of these scenarios.

The DataRes Project sought to document and understand a critical developmental moment, when many universities were starting to articulate the conceptual foundations, roles, and responsibilities involved in research data management. The project investigated the perspectives of stakeholders (e.g., researchers, librarians, information technology [IT] professionals, sponsored research offices) throughout the research lifecycle. Because it is still too early to draw definitive conclusions about prospective roles for LIS or other professionals in research data management, the DataRes Project instead sought to document basic quantitative and qualitative information about stakeholder expectations, current institutional policies, and the preparation that information professionals will need as they take on emerging responsibilities in this area. Because the project was funded by a 21st Century Librarians grant from the Institute of Museum and Library Services, our aim was to establish a baseline study of research data management practices that institutions can use in developing new curricula and training. The greatest benefit of this baseline study may be that it brings to the surface fundamental problems in the emerging landscape of research data management responses and interventions in the United States. Our research suggests that effective institutional responses to meet the challenge of research data management may be slow in coming, but are inevitable in the long term.

Keywords: DataRes project | research data management | data management

Article:

*****Note: Full text of article below**

Research Data Management

Principles, Practices, and Prospects

November 2013

OPEN ACCESS TO RESEARCH DATA IS CRITICAL FOR ADVANCING SCIENCE, SCHOLARSHIP, AND SOCIETY. RESEARCH DATA, WHEN REPURPOSED, HAS AN **ACCRETIVE VALUE**. PUBLICLY FUNDED RESEARCH SHOULD BE PUBLICLY **AVAILABLE** FOR PUBLIC GOOD. TRANSPARENCY IN RESEARCH IS ESSENTIAL TO SUSTAIN THE PUBLIC **TRUST**.

THE VALIDATION OF RESEARCH DATA BY THE PEER COMMUNITY IS AN ESSENTIAL **FUNCTION** OF THE RESPONSIBLE CONDUCT OF RESEARCH. MANAGING THE RESPONSIBILITY OF A BROAD RANGE OF STAKEHOLDERS INCLUDING

The Denton Declaration:
An Open Data Manifesto



Research Data Management

Principles, Practices, and Prospects

November 2013



ISBN 978-1-932326-47-5
CLIR Publication No. 160
Published by:

Council on Library and Information Resources
1707 L Street NW, Suite 650
Washington, DC 20036
Web site at <http://www.clir.org>

Copyright © 2013 by Council on Library and Information Resources. This work is made available under the terms of the Creative Commons Attribution-ShareAlike 3.0 license, <http://creativecommons.org/licenses/by-sa/3.0/>.



Cover art derived from Mariette Papić's poster for The Denton Declaration. Also included are images from Shutterstock.com.

Prospects for Research Data Management

Martin Halbert

The challenge of ensuring long-term preservation of and access to the outputs of scientific research, especially data sets produced by publicly funded research projects, has become a prominent topic in the United States. In 2011, the two-year DataRes Project was initiated at the University of North Texas to document perceptions and responses to this emerging challenge in U.S. higher education and to explore ways in which the library and information science (LIS) profession could best respond to the need for better research data management in universities. This chapter will highlight some of the most provocative findings of the DataRes Project on the topic of research data management in higher education and then consider possible research data management (RDM) scenarios for the future and the implications of these scenarios.

The DataRes Project sought to document and understand a critical developmental moment, when many universities were starting to articulate the conceptual foundations, roles, and responsibilities involved in research data management. The project investigated the perspectives of stakeholders (e.g., researchers, librarians, information technology [IT] professionals, sponsored research offices) throughout the research lifecycle. Because it is still too early to draw definitive conclusions about prospective roles for LIS or other professionals in research data management, the DataRes Project instead sought to document basic quantitative and qualitative information about stakeholder expectations, current institutional policies, and the preparation that information professionals will need as they take on emerging responsibilities in this area. Because the project was funded by a 21st Century Librarians grant from the Institute of Museum and Library Services, our aim was to establish a baseline study of research data management practices that institutions can use in developing new curricula and training. The greatest benefit of

this baseline study may be that it brings to the surface fundamental problems in the emerging landscape of research data management responses and interventions in the United States. Our research suggests that effective institutional responses to meet the challenge of research data management may be slow in coming, but are inevitable in the long term.

Context

The DataRes Project is not the first effort to address the topic of research data management. The National Science Foundation (NSF) funds a great deal of research in the United States, and that research generates large amounts of data. In 2003, NSF issued two reports noting the growing perception of an urgent need to build up the national data management capacity. The report from a 2002 workshop, provocatively entitled *It's About Time* and sponsored by NSF, the Library of Congress, and other organizations, called for a national research initiative to “build a foundation for digital preservation practices that government agencies, cultural institutions, businesses, and others urgently require” (Hedstrom et al. 2003, 26). The 2003 report by Atkins and colleagues, in which they coined the term *cyberinfrastructure* and articulated an agenda for scientific investment based on data-intensive research, also identified the risks of not managing research data over time: “Absent systematic archiving and curation of intermediate research results (as well as the polished and reduced publications), data gathered at great expense will be lost” (Atkins et al. 2003, 11).

These and similar calls in the first years of the twenty-first century led to major collaborative efforts, such as the 10-year National Digital Information Infrastructure and Preservation Program (NDIIPP) undertaken by the Library of Congress in collaboration with NSF and many other organizations to explore and better understand the foundations of the new field of digital preservation (NDIIPP 2010). Research data management has been widely debated and discussed. Many discussions of its importance have taken place at meetings of professional groups concerned with the topic; these discussions culminated in a variety of organizational recommendations and position papers, such as those of the Association of Research Libraries (2006) and the National Academy of Sciences (2009). At the same time, those in business and society more generally were carrying on a discussion of the criticality of so-called “Big Data,” reflecting the growing recognition that computing technology in all walks of life is generating and accumulating ever more vast amounts of data that, if managed effectively, can be “used to unlock new sources of economic value, provide fresh insights into science and hold governments to account” (*The Economist* 2010).

Virtually all of these discussions agreed on two themes. First, the vast amounts of data that research organizations are accumulating are valuable in potentially game-changing ways *if the data are effectively managed*, and second, very few (if any) research organizations

are currently prepared or mandated for the effective management of such unprecedented quantities of data. The growing consensus on these two points was almost certainly a factor in NSF's decision to issue a new mandate in 2010 that all research proposals submitted to the agency after January 2011 must include a "data management plan" (NSF 2010). Such a plan is now understood to be essentially a description of how investigators will "share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants" (National Science Foundation 2013).

The NSF mandate was neither unprecedented nor an isolated intervention by one federal agency. The National Institutes of Health (NIH) had implemented the first major mandate of this kind in 2003, requiring researchers to comply with data sharing and data management practices (NIH 2003). Other federal agencies were adopting similar policies at the same time that the NSF was doing so; for example, the National Endowment for the Humanities adopted a requirement for data management plans that explicitly emulated the NSF requirement (National Endowment for the Humanities 2013).

The NSF mandate prompted a new round of discussions across the United States, especially among intermediaries such as librarians and other information professionals who devote special attention to the long-term preservation of and access to scientific research results. This attention was evident, for example, in the large number of presentations in the 2010 meetings of the Coalition of Networked Information (CNI) that mentioned either the NSF mandate or related research data management topics. It appeared that the concerns voiced in the 2003 reports cited earlier regarding the long-term survival of research data were about to be addressed. There was real hope in many of the 2010 discussions that the new federal agency mandates would lead universities and other research institutions to rapidly adopt much stronger research data management practices and policies.

Study of Research Data Management Responses

The two-year DataRes project was conceived amidst growing concern over research data management. The aims of the project were (1) to study and document trends in the data management plans and associated institutional policies of research institutions in response to federal requirements, and (2) to determine how the LIS profession can best respond to emerging needs of research data management in universities. In the course of the study, project personnel asked a variety of questions about the emerging research data management responses in the United States:

- What trends and patterns are observable in the data management plans and associated institutional policies now being implemented at research institutions in response to federal requirements?

- What do key stakeholders in the research community (e.g., researchers, administrative officials, librarians, funding agency officials, research equipment vendors) expect in the long-term management of research data generated in universities? What is the role of information professionals in such efforts?
- What skills, infrastructure, training, and other preparation do professionals charged with data management responsibilities need, based on both expectations of stakeholders and observed trends in data management policies now being implemented?

The detailed findings of this two-year study are provided elsewhere in this volume. There have been some undeniable quick accomplishments of the “low-hanging fruit” variety to give researchers at the local level basic advice on dealing with the new NSF mandates. For example, low-cost local university workshops have been held and tools cooperatively devised to help researchers develop data management plans (Sallans 2012). But what has most impressed the DataRes research team in the course of this work is the range of barriers to effective research data management at scale, at both the local and the national levels. Although virtually all stakeholders acknowledge the importance of effective long-term management of research data, a daunting array of barriers hamper the prospects for effective research data management practices and programs.

Barriers to Research Data Management

The barriers that hinder effective research data management are not intractable, but they are real. They must be fully understood if institutions of higher education in the United States are to make headway in overcoming them.

Lack of Funding

The most frequently identified barrier to effective research data management is lack of funding. The vast majority of stakeholders consulted in the DataRes Project believed that research data management is an important need that should be addressed, but felt that it does not receive funding at the level required to build needed infrastructure and programs. This perception is somewhat equivocal. The DataRes surveys show that *some* funding is being devoted to research data management programs, usually through a combination of sources. But the overall sentiment expressed by most DataRes survey respondents was that this funding is very modest in scale and often takes the form of incidental commitments of time by librarians who are primarily tasked with other duties. With few exceptions, it was perceived that most institutions devote an almost inconsequential amount of their budgets to research data management functions.

Research data management programs still seem to be mostly conceptual and prospective at a time when the competing demands to fund existing programs in academia are legion. DataRes discussions with stakeholders, including researchers, librarians, university

administrators, and NSF program officers, repeatedly came back to questions of how to fund these programs at scale. Researchers do not wish to allocate research funds to activities, such as research data management, that they see as occurring outside the scope of research. Librarians see a clear need for long-term preservation and access to research data, but typically are not funded to undertake such functions. University administrators do not have established frameworks to determine the relative priority of research data management in the ecology of programs for which they are expected to allocate funding. NSF program officers see the importance of research data management (hence, the new mandates for data plans), but they expect that the consensus on the relative allocation of funding in grant programs will emerge from the field, primarily from researchers. Many academic stakeholders who are not themselves researchers expect that the funding for research data management programs will come from research grants, but this approach ignores the predominant perspective of researchers that the purpose of grants is to fund research, not to maintain research outputs.

Until the fundamental issue of funding is resolved, research data management programs will not be created at any useful scale. But funding obviously follows from other preconditions, including the existence of institutional mandates, professional preparation, and organizational structures. Unfortunately, there are major deficiencies in these areas as well.

Lack of Organizational Structures

The organizational structures of academia are slow to change. They are largely based on long accepted notions of the archetypical functional parts of a university: the faculty, the administration, the library, and (most recently) business IT management. Although intramural collaboration between these groups is encouraged to advance the basic academic goals of research and teaching, these functional divisions are still largely understood as organizational silos. Research data management is among the priorities that have emerged in recent years to challenge these organizational boundaries (another is course management systems).

The findings of the DataRes Project support the idea that effective research data management practices will require close working relationships between divisions of the university, sometimes to the point of blurring boundaries in uncomfortable ways. Although hybrid organizational structures may be required for effective research data management, there are as yet no clear models for these structures. Organizational structures exist for many reasons, including accountability, allocation of funds, and comprehensibility by those trying to interact with the organization. In the case of traditional types of research outputs (e.g., published print journal articles), stakeholders have a general understanding of how the longstanding organizational structures of academia are *supposed* to work together (whether or not they actually work well together). The functions entailed in effectively managing digital research data do not fit as

neatly into these traditional organizational divisions, although these roles are starting to blur. Libraries are not classically understood as being the primary point of management for digital information created by scholars; however, libraries are slowly being reconceived in digital terms.

Business IT is usually associated with central institution-wide functions, such as accounting and electronic mail, and is not typically considered to be deeply embedded in the work of university research teams. Nevertheless, IT functions have been a growing aspect of large research laboratories for many years. University offices of research are usually focused on the administrative aspects of applying for, receiving, and managing grant awards, not the research outputs after the grants have been expended. Yet, if federal agencies implement more stringent (read auditable) requirements for long-term preservation and access to research outputs, research offices will feel pressure to interject themselves into these longer-term aspects of research. Academia has only started groping tentatively toward an understanding of what organizational structures will best support long-term research data management; the DataRes findings show that more integrated organizational structures work better than silos. A better shared understanding of the skills and roles of the various actors in the research cycle is needed to breach these silos.

Lack of Professional Preparation

The DataRes Project identified the lack of training, certification, and other types of professional preparation as another basic deficiency in academia's readiness for research data management. This is perhaps not surprising, given that data management is still an emerging area and there is no general understanding of its requirements among the different parts of academia, but it is nevertheless a huge deficiency for effective long-term research data management. Yet, almost no one within the academic community receives systematic professional training and certification in the management of research data. Still worse on a more fundamental level, *virtually no one in academia perceives that they have a professional responsibility or mandate for research data management functions.*

The DataRes research indicates that librarians may be the closest to understanding their role in research data management, but the standard curriculum of library schools does not include preparation for managing large bodies of data. Moreover, most librarians are unsure exactly what re-training is most important for such duties. Most stakeholders (including librarians) also acknowledge that libraries cannot manage research data alone, but are not yet certain what mix of professional skills is most appropriate for cross-organizational teams working on research data management functions. There have been some LIS curriculum development activities for digital curation roles that may be relevant to research data management roles; this issue will be taken up in the section on scenarios for professional preparation.

Lack of Priority among Researchers

A recurrent theme encountered in the DataRes Project was that researchers are rewarded primarily for undertaking new research, not for managing the results of prior research. The main reason that researchers do not request grant funds for research data management is that they seek to maximize the proportion of grants devoted to research proper rather than to functions that they see (understandably) as secondary support operations. The idea that grants will increasingly be judged in terms of the quality of their data management plans is still unproven. Because researchers themselves are typically the primary agents that judge the quality of federal research proposals in peer-reviewed panels, it is unclear whether long-term management of research data will become a priority in designing future research projects.

Lack of Institutional Mandates

Finally, no generally understood institutional mandates exist for managing research data effectively. Producing data in the course of research activities has traditionally been understood as part of the task of researchers. The idea that researchers should share cumulative sets of research data to advance larger research agendas is a relatively new concept that may have developed from the experience of groups that worked together on multiyear, multi-institutional endeavors such as the Human Genome Project. But although projects like the Human Genome Project show that large-scale sharing of research data can produce major data sets of long-term significance, there is no consensus on or established expectation for long-term data management by individual researchers or institutions. This lack of consensus results in a lack of institutional mandates or policies regarding research data management.

The DataRes Project findings show that the vast majority of universities in the United States are not yet implementing research data management policies at the institutional level; it is simply too soon. After studying the current landscape of higher education, we concluded (perhaps unsurprisingly) that policies come only after practices have stabilized and become accepted, and this has not yet happened for research data management. Until there are widely shared expectations about research data management practices, the current situation will continue. Without institutional mandates, research data may or may not be preserved in accessible ways; their systematic management will definitely not be an institutional priority. There are some indications that this may change, and they will be discussed in the section on scenarios for the future.

Current Developments

Federal agencies made several notable announcements about research data management during the two years that the DataRes Project studied the issue. The new “Data Sharing Policy” requirements were put into effect for NSF proposals submitted on or after January

18, 2011 (NSF 2010). On March 29, 2012, six federal grant-making departments and agencies announced more than \$200 million in grant opportunities for the so-called “Big Data Research and Development Initiative” (Office of Science and Technology Policy [OSTP] 2012).

The Fair Access to Science and Technology Research Act (FASTR) was introduced in both the Senate and the House in early February 2013. If passed, this legislation will require federal agencies to develop policies that ensure rapid access to the products of federally funded research. Shortly after this legislation was introduced, on February 22, 2013, OSTP Director John Holdren issued a policy memorandum entitled “Increasing Access to the Results of Federally Funded Scientific Research,” which includes language very much like that in the FASTR bill (OSTP 2013). The OSTP memorandum “directs each Federal agency with over \$100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government. This includes any results published in peer-reviewed scholarly publications that are based on research that directly arises from Federal funds . . .” (OSTP 2013, 2). Agencies were given six months to respond, but as of this writing (mid-September 2013), the agencies to which the memorandum was directed have not issued public responses. Although much of the focus of the FASTR legislation and the OSTP memorandum is on published articles as the main category of research results, the memorandum explicitly states at the beginning that “such results include peer-reviewed publications and digital data.”

These announcements suggest that federal officials are paying a great deal of attention to research data management. The policies established by various agencies requiring researchers to submit data management plans as part of their proposals were only the first of several steps to encourage researchers and their institutions to increase their efforts to implement more effective practices for the long-term preservation of and access to research data created through federally funded grants. Most of the university responses noted by the DataRes Project were prompted to some degree by the federal announcements, but they also reflected librarians’ genuine concerns that research data are significant academic intellectual assets and parts of the scholarly record in their own regard.

Various research stakeholder groups have issued responses to the February 2013 OSTP memorandum well in advance of the deadline given to agencies. The Association of American Publishers (AAP) put forward a proposal in June 2013 titled the Clearinghouse for the Open Research of the United States (CHORUS; AAP 2013), which suggested that publishers should be the primary entities responsible for the long-term management of research results mandated in the 2013 OSTP memorandum. The CHORUS proposal was greeted with skepticism by some researchers (Eisen 2013; Neylon 2013), who questioned whether publishers would be motivated to preserve publications or make them openly accessible to the public.

A coalition of groups including the Association of Research Libraries, the Association of American Universities, and the Association of Public and Land-Grant Universities issued a draft proposal called the Shared Access Research Ecosystem (SHARE), which emphasizes the role of research universities as long-lived, mission-driven institutions focused on creating, preserving, and disseminating knowledge (Association of Research Libraries 2013). The SHARE proposal “envisions that universities will collaborate with the Federal Government and others to host cross-institutional digital repositories of public access research publications that meet federal requirements for public availability and preservation.” Other commentary on the OSTP memorandum noted that PubMed Central already provides many of the features requested, and new repositories may simply duplicate those features (Neylon 2013).

What is noteworthy about responses to the OSTP memorandum from CHORUS, SHARE, and other research stakeholders is that they were not responses from the primary audience of the memorandum, namely, the large federal grant-making agencies. Although a consensus on research data management practices has not emerged by 2013, what clearly *has* changed is that many stakeholder groups are now willing to engage in the public debate about research data management. Somewhat disheartening is that the nature of these discussions has been rather heated at times, with the positions taken resembling battle lines drawn in the sand. The DataRes Project findings highlight the need for cooperation between all stakeholders in the scholarly communication cycle, rather than strategies that emphasize the primacy of any single stakeholder group or cluster of stakeholders. The importance ascribed to research data management, not only by federal officials, but also by all stakeholders in the scholarly communication cycle, is likely to continue increasing.

Scenarios for the Future of Research Data Management

The DataRes Project sought to document basic quantitative and qualitative information about stakeholder expectations, current policies, and needed preparation for information professionals taking on emerging responsibilities in data management. This information forms a baseline for institutions as they plan new research data management infrastructures, services, policies, and training programs. Following are possible scenarios for the future in terms of the deficiencies discussed earlier.

Funding Scenarios

Much of the future progress on research data management programs will depend on the availability of funding. The DataRes survey of administrators indicates that the most common practice now is to fund research data management programs through a mixed revenue stream model in which funds from several sources are combined. If this hybrid funding model continues to be the most common means

of funding RDM programs, then the main question is how much funding overall will be achievable for such programs through a combination of sources. One scenario is that the status quo will continue. The early research data management programs now in place, consisting primarily of advisory services for faculty seeking to write data management plans, do not receive significant dedicated funds. The incidental time commitments of those providing advisory services are not much above the level of administrative “noise” and could continue indefinitely without significantly advancing the status of research data management nationally. If the status quo continues in regard to funding, it seems likely that researchers will continue to manage data (if at all) through informal mechanisms, such as USB drive backups in desk drawers. Different scenarios may occur in which one or more of the sources of funding devoted to research data management increases, but the likelihood that new funds will be allocated to research data management depends to some degree on how the other deficiencies are or are not addressed.

Scenarios for Professional Preparation

In at least nine U.S. LIS programs, new curricula and associated certificate programs have been or are being developed to address the new data curation responsibilities of information professionals (Keralis 2012). The well-known DigCCur curriculum development project at the University of North Carolina at Chapel Hill has carefully examined a range of new competencies needed by information professionals tasked with managing digital collections (Hank et al. 2010). The DigCCur program and data curation certificates at other LIS programs around the United States are now beginning to produce graduates who are entering the field, but at a time when (as the findings of the DataRes Project make clear) the future of research data management programs is very uncertain. The real question for scenario analysis comes back to the relative level of priority and funding that research data management programs will receive on university campuses. Sustaining and refining professional preparation programs will require that libraries and other academic employers hire and reward professionals with these skills.

Many library directors consulted in the course of the DataRes Project hope to create research data management programs that will employ new graduates to manage large corpora of data sets. If the number of these programs does increase significantly and the demand for individuals with these skills continues to expand, there is likely to be a national blossoming of professional curricula and certification programs for data curation. If, instead, a perception spreads that librarians with these skills are not in demand, these professional preparation programs will come to be seen as a passing fad. A scenario in which this might occur would be if libraries are largely bypassed in the landscape of emerging responses to research data management. If other stakeholders in the research landscape (especially the growing body of IT managers specializing in operational support of research laboratories) become the primary actors in

establishing research data management programs, there is likely to be less demand for research data management curricula in LIS programs. There could also be a hybrid scenario in which professionals from other disciplinary fields enroll in certificate programs for data curation established by LIS programs. What will drive the demand for professional preparation programs in data curation is a rise in the perceived priority of research data management functions among researchers and institutional mandates for research data management functions.

Research Data Management Priority Scenarios

For long-term research data management to become a higher priority for researchers, they must see clear benefit to be derived from devoting time, attention, and funds to these purposes. It is easy to understand a status quo scenario in which research data management continues to be seen as a low priority or simply as an activity outside the scope of research proposals, but what might a more progressive scenario look like?

There are at least two ways that data management may be assigned a higher priority in research proposals. One possibility is that universities that have been early adopters of strong research data management practices (e.g., Purdue University, University of California, San Diego) will be able to demonstrate the added value of these services prominently enough for researchers at most other institutions to see a compelling competitive need for such services at their own institutions. When research grants regularly begin to feature requests for funds to support local data management, significant progress will start to occur in research data management program development.

The other possibility is that political pressures will build to the point that federal agencies mandate more robust and specific requirements for long-term preservation and access for data produced by grant-funded research, including explicit guidance on requests for research data management funding in applications. This second scenario provides the clearest path to funding research data management programs on a regular basis in the future, but it is also highly speculative because it would entail federal agencies specifying far more prescriptive guidelines for the use of awarded project funds.

Scenarios for Institutional Responses and Organizational Structures

Research data management programs will become a prominent part of the research landscape when they become an expected part of the institutional organization of most universities. The need for research data management is unlikely to go away and will likely continue to grow more prominent over time given that academia and society in general are rapidly becoming more data-driven. The response to the need for research data management can be primarily *reactive* or primarily *proactive*, and these two tendencies will produce quite different outcomes.

In a scenario in which institutional responses are primarily reactive, universities would grudgingly adhere to the stricter compliance measures required by federal agencies and implement the measures only in response to threatened penalties by federal auditors. Standards for research data management might come to be understood as similar to other required compliance standards of performance mandated by the U.S. Office of Management and Budget (such as standards for financial reporting). Universities might be forced to comply with legal strictures by reluctantly creating research data management programs that meet the letter of the law rather than embracing the intent and promise of effective research data management programs.

In contrast, universities could respond proactively by establishing new cross-divisional (perhaps interinstitutional) organizations charged with a strong mandate to preserve and provide access to research data. These organizations could be funded at a level robust enough to develop effectively scaled infrastructure and services in support of this goal. The leadership of many or most universities in the United States would have to be convinced to make a strong commitment to research data management for this proactive scenario to come about, but it could certainly happen. The vision and leadership of individuals in positions of authority will ultimately drive this scenario (and by extension, most of the other positive scenarios discussed). If leaders embrace the concept of research data management in coming years, a proactive scenario could have far-reaching effects across the entire landscape of higher education and research in the United States. Are there reasons to believe that such a scenario could come about?

Conclusions

The DataRes Project has noted several events that may constitute reasons for cautious optimism about the future of research data management. Politicians and federal agency officials are paying more attention to research data management. Federal agencies will soon be required to respond to the OSTP directive with agency plans “to support increased public access to the results of research funded by the Federal Government” (OSTP 2013, 2). Whatever form these individual agency plans may take, they should be understood as incremental steps in guiding institutions and individual researchers toward better stewardship of research data. The actual responsibility for long-term stewardship of research data will fall upon the institutional actors who are tasked with sustaining the various parts of the research endeavor. Are these institutions responding to this challenge?

The CHORUS and SHARE proposals by stakeholder communities demonstrate that those in the field are taking the research data management challenge seriously and that stakeholder groups are engaging in efforts to find solutions to the problems of research data management. Both of these proposals (as well as suggestions to extend existing services such as PubMed Central) offer realistic

approaches that would significantly improve the overall capacity of researchers to manage their data in the future. Each proposal has distinct pros and cons, and a healthy debate is warranted about the relative advantages of these and other new proposals that will no doubt emerge over time.

There are signs that stakeholder groups are coming together to hold constructive debates and discussions. For example, the Research Data Alliance is an international collaboration of many different research stakeholder groups that are addressing research data management as a grand challenge of the same scale as mapping the human genome (Research Data Alliance 2013). This collaboration is a relatively rapid, grassroots community response to the perceived need for multiple institutions to advance the understanding of research data management. Another promising sign of confluence is a September 2013 announcement jointly made by 25 organizations that archive scientific data calling for the creation of models for sustaining and coordinating research data management activities across subject domain repositories (Inter-university Consortium for Political and Social Research 2013).

Finally, DataRes interviews conducted with university administrators reveal that research data management planning efforts are going on at many universities across the United States. During the two years in which the DataRes Project was conducted, the status of these planning efforts has evolved from conceptual debates about whether research data management is a good idea to more practical and specific discussions of who will undertake what efforts with what resources. Although the specific outlines of these programs are still emerging, the overall prospects for research data management are encouraging. The second decade of the twenty-first century will inevitably be a time when the foundations for long-term research data management practices will be established. The shape, scope, and success of these practices will make up the next stage of this developmental process.

Works Cited

Association of American Publishers. 2013. Understanding CHORUS. Press release, June 5, 2013.

Association of Research Libraries. 2006. *To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering*. Washington, DC: Association of Research Libraries.

Association of Research Libraries. 2013. *Shared Access Research Ecosystem (SHARE)*. Draft document, June 7, 2013. Last accessed September 1, 2013 at <http://www.arl.org/storage/documents/publications/share-proposal-07june13.pdf>.

Atkins, Daniel, et al. 2003. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Washington, DC: National Science Foundation.

The Economist. 2010. Data, Data Everywhere: Special Report on Information Management. February 25, 2010, Special Issue. Last accessed September 1, 2013 at <http://www.economist.com/node/15557443>.

Eisen, Michael. 2013. A CHORUS of Boos: Publishers Offer Their "Solution" to Public Access. Blog posting, June 4, 2013. Last accessed September 1, 2013 at <http://www.michael Eisen.org/blog/?p=1382>.

Hank, Carolyn, Helen Tibbo, and Christopher Lee. 2010. DigCCurr I Final Report, 2006-09: Results and Recommendations from the Digital Curation Curriculum Development Project and the Carolina Digital Curation Fellowship Program. March 1, 2010. Last accessed September 1, 2013 at http://www.ils.unc.edu/digccurr/digccurr_I_final_report_031810.pdf.

Hedstrom, Margaret, et al. 2003. *It's About Time: Research Challenges in Digital Archiving and Long-Term Preservation*. Washington, DC: National Science Foundation and Library of Congress.

Inter-university Consortium for Political and Social Research (ICPSR). 2013. Sustaining Domain Repositories for Digital Data: A Call for Change from an Interdisciplinary Working Group of Domain Repositories. Press release, September 16, 2013. Last accessed September 17, 2013 at <http://www.icpsr.umich.edu/icpsrweb/ICPSR/support/announcements/2013/09/sustaining-domain-repositories-for>.

Keralis, Spencer. 2012. "Data Curation Education: A Snapshot." In *The Problem of Data*, by Lori Jahnke, Andrew Asher, and Spencer D. C. Keralis. Washington, DC: Council on Library and Information Resources. Available at <http://www.clir.org/pubs/reports/pub154>.

National Academy of Sciences, Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. 2009. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age*. Washington, DC: National Academies Press. 2009.

National Digital Information Infrastructure and Preservation Program (NDIIPP). 2010. *Preserving Our Digital Heritage: The National Digital Information Infrastructure and Preservation Program 2010 Report. A Collaborative Initiative of The Library of Congress*. Washington, DC: Library of Congress.

National Endowment for the Humanities. 2013. *Data Management Plans for NEH Office of Digital Humanities Proposals and Awards*. National Endowment for the Humanities. Last accessed September 1, 2013 at http://www.neh.gov/files/grants/data_management_plans_2013.pdf

National Institutes of Health. 2003. Final NIH Statement on Sharing Research Data: NOT-OD-03-032. February 26, 2003. Last accessed September 1, 2013 at <http://grants.nih.gov/grants/guide/notice-files/not-od-03-032.html>.

National Science Foundation. 2010. Dissemination and Sharing of Research Results: NSF Data Sharing Policy. Washington, DC: National Science Foundation. Last accessed September 1, 2013 at <http://www.nsf.gov/bfa/dias/policy/dmp.jsp>.

National Science Foundation. 2013. *Grant Proposal Guide (GPG): Chapter VI, Other Post Award Requirements and Considerations*. Washington, DC: National Science Foundation. Last accessed September 1, 2013 at http://www.nsf.gov/pubs/policydocs/pappguide/nsf13001/aag_6.jsp.

Neylon, Cameron. 2013. Chapter, Verse, and CHORUS: A First Pass Critique. Blog posting, June 7, 2013. Last accessed September 1, 2013 at <http://cameronneylon.net/blog/chapter-verse-and-chorus-a-first-pass-critique/>.

Office of Science and Technology Policy. 2012. Obama Administration Unveils “Big Data” Initiative: Announces \$200 Million in New R&D Investments. Press release, March 29, 2012. Last accessed September 1, 2013 at http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf.

Office of Science and Technology Policy. 2013. Memorandum for the Heads of Executive Departments and Agencies, Subject: Increasing Access to the Results of Federally Funded Scientific Research, February 22, 2013. Last accessed September 1, 2013 at http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Research Data Alliance. Website, “About” page. Last accessed September 1, 2013 at <http://rd-alliance.org/about.html>.

Sallans, Andrew. 2012. DMP Online and DMPTool: Different Strategies Towards a Shared Goal. In *The International Journal of Digital Curation* 7(2):123–129. Last accessed September 1, 2013 at <http://www.ijdc.net/index.php/ijdc/article/view/225>.