The Open Data Imperative:

How the Cultural Heritage Community Can Address the Federal Mandate

Suzie Allard, Christopher Lee, Nancy Y. McGovern, Alice Bishop

July 2016

Council on Library and Information Resources

ISBN 978-1-932326-56-7
CLIR Publication No. 171
Published by:

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

This project was made possible in part by the Institute of Museum and Library Services
(grant number RE-00-13-0097-13)

The views, findings, conclusions, or recommendations expressed in this publication
do not necessarily represent those of the Institute of Museum and Library Services.

Cover illustration: © Stuart Kinlough via Ikon Images

Contents

About the Authors

Suzie Allard is associate dean for research in the College of Communication and Information and professor in the School of Information Sciences at The University of Tennessee. Her research focuses on the ways in which scientists and engineers use and communicate information, particularly in the virtual environment and across distributed work teams. Current projects center on science data curation and interdisciplinary scientific work teams, with an emphasis on data sharing. Allard is a co-principal investigator and member of the leadership team for DataONE, an environmental science cyberinfrastructure project funded by the U.S. National Science Foundation (NSF). She also serves on the board of directors for the Networked Digital Library of Theses and Dissertations. She is principal or co-principal investigator on grants funded by NSF, the Institute of Museum and Library Services, the Alfred P. Sloan Foundation, and the Institute of Electrical and Electronics Engineers.

Christopher (Cal) Lee is professor at the School of Information and Library Science at the University of North Carolina, Chapel Hill. He teaches archival administration, records management, digital curation, information technology for managing digital collections, and digital forensics. He is a lead organizer and instructor for the DigCCurr Professional Institute, and he teaches professional workshops on digital forensics methods and principles. In addition, he is senior personnel on the DataNet Federation Consortium. Lee's primary area of research is digital curation. He developed "A Framework for Contextual Information in Digital Collections" (Lee 2011a) and he edited and provided several chapters for *I, Digital: Personal Collections in the Digital Era* (Lee 2011b). He is principal investigator (PI) of BitCurator Access; he was formerly PI of BitCurator and of the Digital Acquisition Learning Laboratory (DALL) project. He has also served as co-PI on several digital curation education projects.

Nancy Y. McGovern has been responsible for digital preservation at the Massachusetts Institute of Technology Libraries since 2012. She directs the Digital Preservation Management (DPM) workshops, which have been offered 50 times since 2003. In her more than 25 years of experience with preserving digital content, she has held senior positions at the Inter-university Consortium for Political and Social Research (ICPSR), Cornell University Library, the Open Society Archives, and the Center for Electronic Records of the U.S. National Archives. She chairs the Digital Records Expert Group of the International Council on Archives. In 2015, she was elected vice president/president-elect of the Society of American Archivists (SAA). She completed her Ph.D. on digital preservation at University College, London, in 2009. She is a Fellow of SAA, a National Digital Information Infrastructure and Preservation Program Digital Preservation Pioneer, and twice received the SAA Preservation Publication Award.

Alice Bishop is senior program officer at the Council on Library and Information Resources (CLIR). She oversees the Postdoctoral Fellowship Program, which provides recent Ph.D. graduates the opportunity to work on projects forging and strengthening connections among library collections, educational technologies, and current research. She is also responsible for the CLIR Chief Information Officers (CIO) group, directors of organizations that have merged their library and information technology units on the campuses of liberal arts colleges and small universities. She has more than 30 years of project management experience, including coordinating scholarly exchanges with China at the National Academy of Sciences and managing international and national media programs at The Freedom Forum/Newseum in Washington, DC.

Acknowledgments

The Council on Library and Information Resources is grateful to the Institute of Museum and Library Services for this opportunity to examine the federal public access plans and their implications for the cultural heritage community. Special thanks are due to Trevor Owens in the Office of Library Services and his predecessor, Robert Horton, for their guidance, patience, and support throughout the research process.

The authors would like to thank Fawnie Erickson for transcribing all of the interview recordings. We are also grateful to Amanda Whitmire and colleagues, who produced a table summarizing the federal public access policies resulting from the Office of Science and Technology Policy memo of February 2013, which was useful in cross-checking our own findings.

As of April 2016, when the research for this publication was completed, public access plans were still in development at several federal agencies and new curricula for data curation were becoming available. We have done our best to present an accurate picture, and take full responsibility for any errors or omissions.

We are indebted to William Kilbride, Emily Reynolds, Abby Smith Rumsey, and Thornton Staples for their invaluable input, and to Kathlin Smith for her diligent editing.

Foreword

This timely, thorough, and helpful study explores the consequences of, and responses to, a federal mandate published in 2013 that requires federal agencies with budgets of $100 million or more to develop sustainable plans to keep data created by funded projects—and the products that incorporate those data (such as original research)—open, accessible to the public, and managed over time, assuring that the data are preserved, migrated to new platforms when necessary, and can be re-used productively.

One of the key terms of this report is "capacity," as described in the opening section:

> These new requirements have significant implications for cultural heritage institutions in addressing the current deficit in the capacity to support the re-use of data over time and across generations of technology (digital curation) and in enabling collaboration based on shared infrastructure.

Because the report focuses on technology and digital data, the concept of capacity can be initially interpreted to refer to volume and magnitude: the vast, unprecedented accumulation of contemporary data, the need to develop an infrastructure to contain and make this flood of information manageable, and the sheer size of the federal government and its many agencies, perspectives, and interests that nonetheless must accommodate the open data mandate in a coherent fashion.

The richness of this report, however, is derived in part from the other meaning of capacity: the ability to understand, to master a phenomenon. In this respect, "capacity" entails both a quantitative designation and a cognitive, qualitative one. This conceptual interrelationship imbues the study, which at least tacitly acknowledges that the speed and petabytes of our infrastructure, and our acumen and ability to grasp cogently and effectively what we have produced, are inseparable.

Reflected in its title, *The Open Data Imperative* underscores the urgency of engineering the technical capacity requisite to contain and analyze our data, and the intellectual and behavioral conditions required to make this information meaningful. The report articulates a number of recommendations aimed at the behavioral. We need to collaborate across agencies, and across professional boundaries. The scale of the challenge is such that one profession cannot adequately solve it. Information technologists, librarians, archivists, curators, engineers, and scholars need to align their interests in order to create a new ecology wherein different types of data (e.g., raw, visualized, or formally published) can flourish and be susceptible to ongoing inquiry, facilitating an evolving understanding of the astonishing variety of phenomena the data represent.

Training is similarly key: new curricula and continuing education are required as the data become more complex and pervasive. Strong communication is essential in order to share best practices, identify exemplary procedures for curation, and take advantage of different perspectives that enrich the dialog. The traditional conditions we have inherited—competing institutions, siloed agencies, idiosyncratic professional lexicons—cannot address this emerging panoply of data.

It is an ancient tension, managing the technologies we create in order to purposefully advance our understanding. Federal funding has generated data that allow us to perceive the origin of the universe, the subtle mechanics of our DNA, and the sophisticated variations of manuscripts of medieval romance. *The Open Data Imperative* asks us to recognize that we have begun to develop a remarkable array of information integral to our capacity to know, and that to manage this outpouring of data our conduct needs to be organized and aligned in ways that mirror the technological interdependency so fundamental to augmenting and extending our grasp.

—Charles Henry

Executive Summary

Data are a valuable national resource for a variety of stakeholders across all sectors of society. Dramatic advances in information and communication technology have opened up unprecedented opportunities for broad public access, innovative research, and citizen engagement, but this potential can be realized only if data are properly managed and exposed over time. New U.S. government requirements for exposing and managing federally funded research data add urgency to the call for curating data so that they can be used, re-used, and exploited by future generations. These new requirements have significant implications for cultural heritage institutions in addressing the current deficit in the capacity to support the re-use of data over time and across generations of technology (digital curation) and in enabling collaboration based on shared infrastructure.

Cultural heritage encompasses various types of artifacts (analog or digital), as well as attributes and behaviors that groups or societies maintain over time to preserve our connections to the past, present, and future. Cultural heritage institutions have a mission to support, perpetuate, and provide access to essential elements of culture as a whole. There are many different types of cultural heritage institutions, but three of the most commonly recognized are libraries, archives, and museums. Materials in their care are vital to the ongoing advancement and perpetuation of the sciences, social sciences, arts, and humanities.

This report presents the implications for the cultural heritage community of the recent focus on creating public access to data and publications resulting from federal funding, and our recommendations for relevant stakeholders. The recommendations are based on a review of federal agencies' responses to new government requirements, case studies of seven digital curation projects, and an investigation of the current professional capacity for the long-term management of cultural heritage digital content, including data.

Open Access vs. Public Access

For the purposes of this report, **open access** (OA) is defined as unrestricted access and unrestricted re-use of research and scholarship. OA literature and data are digital, online, free of charge to the user, and free of most copyright and licensing restrictions.

Public access refers to the mandate for government agencies to make federally funded digital data and peer-reviewed publications fully discoverable and usable by the public.

The Mandate for Public Access to Data

In 2013, the U.S. government issued a mandate requiring federal agencies with annual research and development expenditures of more than $100 million to create plans for public access to their data. Agencies were asked to manage information as an asset, which requires a variety of professional actions to ensure the preservation and sustainability of data so that they can be re-used and interpreted in new ways. The asset management approach requires agencies to address costs and long-term sustainability for data management. Agencies were also instructed to reduce the costs of compliance through interagency cooperation.

In late 2013, the Institute of Museum and Library Services (IMLS) asked the Council on Library and Information Resources (CLIR) to conduct an analysis of the federal public access plans to help IMLS and its constituents understand what the implications of the federal mandate are and how needs and gaps in digital curation can best be addressed, and to raise awareness within the cultural heritage community of specific ways to address current needs.

Federal Agency Response to the Public Access Mandate

Understanding how federal agencies are responding to the public access mandate provides valuable insight into how other organizations can provide public access to data. Each agency's public access plan focuses on two separate, but related, components: (1) access to research data and (2) access to the products of analysis based on these research data in the form of peer-reviewed articles. Part 1 of this report reviews 21 agencies' public access plans and presents 12 key findings grouped into 3 complementary areas: open data infrastructure, roles and responsibilities, and the provision of data to the public. The findings of this review lead to a series of conclusions that provide a foundation for an organization to develop an action plan for public access to data.

Implementing Digital Curation: Project Leaders' Experience

Providing public access to data requires effective digital curation strategies. Part 2 of the report focuses on interviews with leaders of seven projects funded by IMLS. They identify skills, capabilities, and institutional arrangements that facilitate digital curation activities. Nine high-level findings derived from the experiences of these professionals are followed by consideration of the major implications for coordination and collaboration when project leaders are developing their open access practices and processes.

Capacity Building for Current and Future Digital Curation

A skilled workforce is essential if the promise of public access to data is to be fulfilled. Part 3 describes progress in competency and capacity building, and reviews the current state of continuing education for managing digital content. It explores the progress and potential role of continuing education programs in competency building, curriculum development, and support for lifelong learning as the range of requisite digital curation skills evolves. It includes an analysis of digital curation and related job postings and examines the evolution of the skills and roles involved.

Recommendations

The following recommendations are intended primarily for IMLS and its stakeholders, but many are applicable to the broader community of funders and researchers. Recommendations from each section fall under three major themes:

1. **Open data infrastructure:** Tapping the full potential of data requires ready and persistent access to usable and coherent data.
2. **Roles and responsibilities:** Building community support for open access across a vast array of stakeholders both within and outside an organization is crucial to promoting and implementing data production, use, and preservation.
3. **Public access to data:** Collaboration among government agencies, foundations, academic institutions, and other interested parties is vital to promote interdisciplinary studies, help establish a viable federal Research Data Commons, and support long-term sustainability of data.

Open Data Infrastructure

1. **The value of data lies in their use.** Just as interstate highways have improved the nation by creating access for commerce, publicly accessible data are an important component to improving economic and societal well-being (NRC 1997). They serve as a vital element of our "epistemic infrastructure" (Hedstrom and King 2006). Building the open data infrastructure should be a national priority insulated from the influence of politics and treated as a vital national asset.
2. **Exemplars can be powerful.** Agencies with successful approaches can provide leadership and vision to others. Numerous agencies have firsthand experience with challenges and solutions that can be instructive to other organizations seeking to implement open data initiatives.
3. **Changing organizational culture is difficult.** It is necessary to change organizational culture to fully implement the mandate, yet the required scale of change is challenging because it involves a variety of stakeholders, including many outside the cultural

heritage sector. Information professionals in the organization's library, archives, data center, or in other parts of the organization can help with this transition given their knowledge of the data life cycle and their understanding of the information behavior required of data producers and data consumers.

4. **Information professionals and the cultural heritage community have a vital role to play in developing a Research Data Commons.** The proposed federal Research Data Commons should be premised on the academic commons model, which has a rich tradition of facilitating vibrant new forms of scholarship.[1]
5. **IMLS can facilitate a Research Data Commons concept.** IMLS can encourage proposals and fund projects involving collaboration between the public access efforts of government agencies and the digital curation work under way at cultural institutions.
6. **Organizations should join forces to support education and training.** Government agencies, the digital curation community, and cultural heritage organizations should collaborate on joint, shared, or cooperative programs that address common educational and training needs. Developing a community-based infrastructure could help ensure that curriculum materials and related resources are broadly accessible to instructors to maximize the reach of curricula and reduce the cost of development.
7. **Funders can help support competency building.** There are opportunities for funders to encourage and fund interdisciplinary, collaborative competency-building projects. Individual researchers and practitioners, as well as data creation and digital curation programs, would benefit from collaborative projects and initiatives that include digital curators and data science researchers to leverage, extend, and refine existing competency-based models and curricula.

Roles and Responsibilities

8. **Federal agencies need ongoing support as they transition to a culture of open data.** The federal resource, Project Open Data, could help agencies make the cultural shift necessary to manage information as an asset. This site provides useful links to definitions, implementation guidance, tools of many kinds, resources, case studies, and other ties to the open data community. Expanding awareness of this resource and encouraging more community input could facilitate best practices for open data.
9. **Libraries, archives, and government data centers should be involved when public access to data is discussed and plans are implemented.** Because these entities and the information professionals who run them can provide expertise and knowledge, their role should be explicitly stated in the plans.
10. **Data sustainability needs more attention and discussion.** Fulfilling the goals of the public access mandate requires ongoing investment in infrastructure. Agency plans offer few concrete

1 For a discussion of the concept of a Research Data Commons, see Reichman and Uhlir 2003, and Halpin et al. 2006.

strategies for keeping data accessible and usable over time. This issue must soon be addressed more completely if the data are to be accessed and repurposed, and their future value retained.

11. **Relevant stakeholders should develop train-the-trainer programs and provide sustained funding for them.** These programs should include incentives to use and re-use existing curricula for continuing education programs and offerings, while increasing the scope and scale of professional development for digital curation, which is critical to address the federal mandate.
12. **Competency models should be considered when planning digital curation.** The digital curation community has devoted considerable effort to identifying and defining competencies that facilitate digital curation and continually advance and promulgate good practices. These models should be considered when planning future activities.
13. **A community-based working group should explore and monitor the digital curation workforce as it grows and evolves.** Career planning and mentoring programs for researchers and practitioners in digital curation and the development of a means to monitor the growth and potential capacity of the digital curation workforce could inform the definition of common modules to build well-formed job descriptions for digital curation and data curation positions.
14. **Support for residencies and fellowships should be expanded.** Graduates from academic programs need an established path to placement in curatorial positions in a range of repositories. One path is to increase support for residencies, fellowships, and postdoctoral programs, including the National Digital Stewardship Residency (NDSR) program, that incorporate continuing education and project-based practical experience.

Public Access to Data

15. **Relevant stakeholders should work together to educate the public on ways to share and re-use data.** Data sharing and re-use adds value to a resource that has already been collected. To maximize this potential, it is essential to raise awareness through education and outreach, which only a few agencies note in their public access plans. Libraries, archives, museums, and information professionals could provide essential support in this area.
16. **The role of education should be better defined in public access plans.** Some plans focus on educating federal agency staff, while others focus on educating data users and producers. It is crucial to educate all stakeholders.
17. **The issues surrounding public access to publications and data should be disambiguated.** The solutions for creating public access to data are still mostly nascent and need the greatest effort, attention, and support. The solutions for creating public access to publications are more mature and ready for implementation across agencies.

Digital Heritage

The United Nations Educational, Scientific and Cultural Organization (UNESCO) recognizes that cultural heritage can be either tangible (movable, immovable, underwater) or intangible (oral traditions, performing arts, rituals). This report focuses specifically on what UNESCO calls *digital heritage.*

"The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are 'born digital', there is no other format but the digital object.

Digital materials include texts, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained." (UNESCO 2003)

18. **Infrastructure to support ongoing data discovery, access, analysis and sensemaking is necessary for data-driven research and innovation.** Data.gov, which serves as a national catalog for open data sets, is helping to make data visible, but more tools and services are needed. Increased attention should be given to U.S. Open Data, which matches data producers and consumers to create sustainable data ecosystems but has seen relatively little use.
19. **Metadata are critical to public access; metadata creation must be improved.** The analysis of public access plans revealed important recurring themes regarding metadata[2]: (a) data management plans should identify standards used for the metadata; (b) data sets should be accompanied by formal documentation about the metadata; (c) metadata for data sets should include the common core from the schema used by the federal government;[3] and (d) metadata must be supplied for publications.
20. **There is potential for much better coordination between work on data management plans and work on access strategies and systems.** There is often a disconnect between the discussions of government public access data plans and discussions of digital curation, including the development and implementation of data management plans. We see potential for further collaboration and integration of these efforts. Professionals engaged in open access initiatives can learn from the work in developing and implementing data management plans. Similarly, experience with open access initiatives can help inform data management plans so that their provisions for access are most likely to be viable and sustainable.

2 *Metadata* refers to "data about data" and can include descriptions of data content, context, structure, interrelationships, and provenance.

3 See https://project-open-data.cio.gov.

Understanding the Background and Context

Data! Data! Data! I can't make any bricks without clay! —Sherlock Holmes

In March 2012, the Office of Science and Technology Policy (OSTP) announced its Big Data Research and Development Initiative. The initiative committed $200 million for programs over several years to "improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of data" (OSTP 2012). As Clifford Lynch notes, three groups of services must be in place and operating effectively and at scale to fulfill the most urgent and basic needs for research data management (RDM): developing credible data management plans, appropriately documenting datasets for sharing and preservation, and finding platforms (either locally developed, through consortia or disciplinary centers, or even via commercial services) to share data and guarantee preservation over the next decade (2013, 395).

The following year, in February 2013, OSTP issued a memorandum to the heads of executive departments and agencies with more than $100 million in annual research and development expenditures directing them to develop plans promoting public access to digital data sets and publications. The memorandum identified some uniform guidelines and instructed agencies to coordinate their responses and associated plans to minimize the burden and costs associated with compliance. Although the agencies have sought uniform and compatible approaches, there are discrepancies across the agency plans that have been made public to date, as well as within communities of interest and practice; these discrepancies reflect the significant variance in needs, resources, and capacities of the communities.

Data are commonly recognized as an important resource in the sciences, yet they are vital to all areas of human inquiry. It is therefore imperative to examine the implications of the federal mandate for institutions and professionals in the cultural heritage sector. The

Key Documents

February 22, 2013 Executive Directive: "Increasing Access to the Results of Federally Funded Scientific Research." This policy memo directed federal agencies with more than $100 million in research and development expenditures to create plans to provide free public access to the results of federally funded research.

May 9, 2013 Executive Order 13642: "Making Open and Machine Readable the New Default for Government Information." This executive order focused on treating information as an asset that should be managed to ensure that it remains open and freely accessible to the public when legally permissible.

May 9, 2013 Office of Management and Budget Circular: OMB M-13-13, "Open Data Policy—Managing Information as an Asset." This circular accompanied Executive Order 13642 to require agencies to collect information in a manner that encourages openness and interoperability.

mandate highlights the need for infrastructure that can support open access to data and publications. The standards, practices, and guidelines implemented by government agencies will have a notable impact on the standards, practices, and guidelines that those in the cultural heritage sector need to adopt.

IMLS as a Leader in Advancing a National Digital Platform

The Institute of Museum and Library Services (IMLS) is a key player in the development of conceptual and professional approaches to digital curation. Its mission is to "inspire libraries and museums to advance innovation, lifelong learning, and cultural and civic engagement;" it leads through research, policy development, and grant-making. IMLS serves diverse communities through libraries, including public, academic, research, special, and tribal libraries; archives; museums, including art, history, science and technology, tribal, and children's museums; historical societies; planetariums; botanic gardens; and zoos. IMLS refers to these organizations collectively as cultural heritage institutions.

Through its Laura Bush 21st Century Librarian Program (LB21), IMLS has invested heavily in helping cultural heritage organizations expand their RDM services. It has supported research in RDM, including tracking the needs of research organizations as they respond to the new federal requirements for public access. Recognizing the need for a holistic approach to the most promising digital tools, services, infrastructure, and expertise that have potential to scale, IMLS funded the creation of a national digital platform through its National Leadership Grant program.[4]

Speaking the Same Language to Facilitate Open Data

The terms *data curation*, *digital curation*, and *data management* are often used to refer to similar sets of activities, but they tend to be used in somewhat different professional or disciplinary contexts. Failure to recognize these differences and relationships can hinder professional activities related to open data, given the need to collaborate and communicate across boundaries.

Building on the LB21 Program started three years earlier, IMLS in 2006 called for grant proposals to develop educational programs in digital curation and funded several programs resulting from this call (Ray 2009); many are discussed in this report. Although *digital curation* is often used to describe activities focused on scientific data, it is used as a label for activities that span the full range of digital heritage. For example, scholars within the humanities are increasingly framing their work in terms of "data sets" as opposed to focusing solely on textual documents.

4 https://www.imls.gov/issues/national-issues/national-digital-platform

What's in the Term?

Data curation tends to be used in settings where coordinated efforts are made to care for data that have been generated from scholarly activities. The emphasis has primarily been on the products of science, though the term is increasingly applied to data generated and used in the humanities.

Digital curation is often considered more inclusive than "data curation." Use of the term is most prominent in the cultural heritage sector and within educational initiatives grounded in library and information science programs. There tends to be a relatively strong orientation toward authenticity, trustworthiness, and long-term preservation.

Data management emerged in the private sector to refer to activities focusing on the growing body of data being generated within enterprises. Within the domain sciences, it refers to the handling, manipulation, and retention of data generated within the context of the scientific process. Use of this term has become more common as funding agencies require researchers to develop and implement data management plans as part of grant-funded project activities.

Data management has arguably become a more common term in light of the recent push by many funding agencies for researchers to develop and implement data management plans as part of grant-funded project activities.

Defining Cyberinfrastructure

In 2003, the Blue Ribbon Panel on Cyberinfrastructure of the National Science Foundation (NSF) introduced a definition of *cyberinfrastructure* that included both technological and sociological aspects. Collecting, analyzing, and storing vast amounts of data requires technology to address the mechanics of data access and preservation as well as interoperability across data sets. At the same time, human processes are required in digital curation and management. In 2007, NSF noted the importance of state-of-the-art data management and distribution systems, and the need to improve services by instituting digital libraries and fostering focused education in digital curation.

As with most emerging concepts, the definition of cyberinfrastructure continues to be debated and refined. For this report, cyberinfrastructure refers to the sociotechnical framework that provides tools and services to data producers, investigators, managers, and users. With data volume expanding so rapidly, the lack of a large enough workforce with the curation skills to provide data services is a key impediment to building a robust cyberinfrastructure.

Defining Data

In this study, we use definitions from NSF's 2007 *Cyberinfrastructure Vision for 21st Century Discovery*. *Data* refers both to raw data, which may come from observations, experiments, models, or other processes, and to the documentation needed to describe and interpret the raw data. Metadata refers to "data about data" and can include descriptions of data content, context, structure, interrelationships, and provenance.

Because data are collected across disciplines, they are by nature heterogeneous. As Sayeed Choudhury noted in his testimony to the House of Representatives' Committee on Science, Space, and Technology Subcommittee on Research:

> One of the overarching issues to consider for wide-scale implementation of data sharing relates to an "ecosystem" viewpoint for infrastructure. Related to this point is the reality that all data are not alike. Scientific data comes in various levels that range from the raw, unprocessed signals generated directly by instruments (e.g., telescope, genome sequencer) to more calibrated data to highly refined, processed data cited within publications. These different levels of data possess different requirements for IT [information technology] infrastructure (Choudhury 2013).

Although there has been a significant move toward providing open access to research data created with public sector funds and considerable progress made in defining and developing professional capabilities to steward those data, neither one of these endeavors has a single clearly defined professional home. Both are undertaken by individuals with a vast array of disciplinary backgrounds, job titles, and institutional contexts. As the recent National Research Council (NRC) report, *Preparing the Workforce for Digital Curation,* states, "There is no single occupational category for digital curators and no precise mapping between the knowledge and skills needed for digital curation and existing professions, careers, or job titles" (NRC 2015, 1).

Technology for Open Data

There are numerous tools and resources supporting open data and open access publications. Examples include DataONE's Investigator Toolkit; DataCite; Creative Commons licenses; Open Researcher and Contributor ID (ORCID); institutional repositories for pre- and post-prints and aligned data repositories; repository services, such as Chronopolis, MetaArchive, and DuraCloud; re3data.org, a directory of 1,500 research data repositories; the National Institutes of Health (NIH) PubMed central repository; and the Scholarly Publishing and Academic Resources Coalition (SPARC). The collaboratively developed and customizable Data Management Planning (DMP) Tool addresses one aspect of public access to data, providing a framework through which researchers and information professionals can assess their needs and confer about ways to meet them.

Researchers and digital stewards will need to use many of the existing tools to comply with the new federal guidelines. They will have to ensure both access to and full digital re-use of the complete text of digital articles. Some university libraries have already made considerable investments in digital repositories, which have the potential to benefit professionals across cultural heritage institutions who can adopt similar tools and models.

PART I
Responding to the Mandate

Suzie Allard

In the United States, the drive to provide access to research data was invigorated when the federal government began public conversations about the value of data and issued the 2013 mandate to federal agencies creating requirements for public access to data. The executive directive that contained this mandate, "Increasing Access to the Results of Federally Funded Scientific Research," required federal agencies with annual research and development expenditures of more than $100 million to create plans for increasing access to federally funded scientific research, both as published articles and as data, and instructed the agencies to submit their public access plans within six months. (Agencies subject to the 2013 executive directive are listed in the sidebar on the next page.) The federal sequester in 2013 delayed the original timeline for the plans' release, but 20 of the agencies subject to the mandate had made their public access plans available as of April 2016.

Since the release of the executive directive, public access to data has become embedded in conversations about research, particularly research relating to science, according to OSTP Assistant Director of the Scientific Data and Information Science Division Jerry Sheehan (2015). Many disciplines receive research funding from agencies subject to the mandate.

The scientific enterprise is part of cultural heritage. For example, the Smithsonian Institution's breadth of research makes it clear that cultural heritage includes science, as well as the fields of history, art, and culture. The plans of even the primarily science-oriented agencies have implications for cultural heritage, because they contain the strategies and practices for infrastructure (i.e., skills, expertise, and technology) that these agencies need to implement the mandate. Examining these plans allows those in other disciplines to consider how the cultural heritage community might address the

Agencies Subject to 2013 OSTP Executive Directive

Agencies with public access plans that OSTP approved for public release as of April 29, 2016

Department of Agriculture
Department of Commerce*
- National Institute of Standards and Technology
- National Oceanic and Atmospheric Administration

Department of Defense
Department of Energy
Department of Health and Human Services
- Administration for Community Living *(publications only)*
- Agency for Healthcare Research and Quality
- Assistant Secretary for Preparedness and Response
- Centers for Disease Control
- Food and Drug Administration
- National Institutes of Health

Department of Transportation
Department of Veterans Affairs
National Aeronautics and Space Administration
National Science Foundation
Smithsonian Institution
U.S. Geological Survey

Source: CENDI

Agencies that had made their plans public, but had not yet been approved by OSTP as of April 29, 2016

Department of Labor
Environmental Protection Agency
Institute of Museum and Library Services

Agencies that had not yet made their plans public as of April 29, 2016

Department of Education**
Department of Homeland Security
Department of Housing and Urban Development***

*Although NIST and NOAA, agencies under the Department of Commerce, have their own public access plans, the Department of Commerce itself does not. It is, however, party to a larger Open Government document, which was included in this analysis.

**DoED created a data inventory (datainventory.ed.gov) that describes grant-funded research data and some administrative and statistical data that is being maintained.

***HUD notes the need for an open data plan in its *HUD Enterprise Roadmap* (version 6.0 May 2015), and there is a link to some documentation at the "Digital Strategy" site (http://portal.hud.gov/hudportal/HUD?src=/Digital_Strategy). However, no complete plan is available.

federal mandate.[5] In all aspects of cultural heritage, libraries, librarians, archivists, and other information professionals have an important role to play.

About the Public Access Plans for Data

As mandated, the access plans focus on two separate, but related, components: access to research data, and access to the products of analysis based on these data in the form of peer-reviewed articles. Research data are defined as "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings." Items excluded by this definition include "preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues." In addition, physical objects are excluded (OMB Circular A-110, rev.).

Together, the executive order, "Making Open and Machine Readable the New Default for Government Information," and the memo from the Office of Management and Budget (OMB), "Open Data Policy—Managing Information as an Asset," provide a well-defined approach for increasing access to federally funded scientific research and creating an open data environment. In this report, the approach put forth in these two documents will be referred to as *the framework*.

Findings

We analyzed 21 federal agency public access plans that were openly available as of late 2015.[6] Our analysis generated 12 high-level findings grouped in three areas: open data infrastructure, roles and responsibilities, and making data public. This section allows readers to negotiate the findings at different levels of detail. An explanation of methods and overview of the limitations of this research are available in Appendix 1. A list of, and links to, the 21 federal department and agency public access plans used for this report are provided in Appendix 2.

5 SPARC has created a new community resource, available at http://datasharing.sparcopen.org/, for tracking, comparing, and understanding U.S. federal funder research data sharing policies.

6 These include public access plans approved for release by OSTP; plans that had been made public but were not yet approved by OSTP; and the public access plan for USAID, although that agency is not subject to the 2013 OSTP executive directive.

Open Data Infrastructure

1. The mandate aims to create an environment for coordination of open data activities, but this has not been fully realized. The flexibility afforded to the agencies so they can best serve their communities is a considerable strength. However, this could also inhibit collaborative activities between government agencies, and more importantly, between these agencies and nongovernmental partners.

The executive order and the OMB memo provide a framework for increasing access to federally funded scientific research and creating an open data environment. The intent of the framework is to simplify client use of the data and increase opportunities for data integration across agencies. This kind of synergy across agencies can add value to the data that each agency collects by increasing the availability of multiple data streams.

Some agencies' public access plans—notably those of the Department of Defense (DOD), the Department of Energy (DOE), and the National Oceanic and Atmospheric Administration (NOAA)—adhere closely to the structure established in the framework and tend to be detailed and lengthy. Other agencies have developed plans that include some uniform elements and compatible approaches, but deviate from the framework in certain aspects, reflecting a diversity of agency missions and focus areas. In their planning, agency staff appear to be taking into account the domains most likely to use their data; there is considerable variation in how researchers from different domains use data for scientific inquiry. Agencies also have varying levels of funding for infrastructure development and research support, which has likely influenced their public access plans for data.

The following are examples of how several agencies' plans deviate from the framework:

- The Department of Transportation (DOT) has integrated its publicly available plan into its Open Government Plan web pages, where it also introduces its data inventory page for its publicly available data sets. This approach is less detailed and does not address all the items in the framework.
- The Department of Commerce addresses its public access plan for data in the Open Government Plan (version 3.5 September 2015). The plan directs the chief data officer to work with each of the bureaus and operating units (BOUs) to create plans maximizing awareness within the BOUs of the data they are creating and the ways in which those data may be used. The National Institute of Standards and Technology (NIST) is a BOU that has made its plan publicly available.
- NIST's plan provides minimal information beyond sharing the guiding principles for implementation and a brief overview of the implementation strategy. For example, NIST did not outline its intentions to use PubMed Central, although this is noted in the Department of Commerce's Open Government Plan.
- The U.S. Geological Survey (USGS) has published an instructional memo (IM OSQI 2015-01) that uses a data life cycle approach to discuss its plan for handling data. The USGS data life

cycle reflects the needs of USGS researchers (see http://www.usgs.gov/datamanagement), although it is different from the information life cycle defined by OMB Circular A-130, "Management of Federal Resources." Using this approach allows USGS to link to its active site for data management, which gives its instructional memo longevity as changes are reflected in the linked pages.

- The National Aeronautics and Space Administration (NASA) adapts and focuses the framework to reflect elements that are most relevant to NASA researchers.
- The Smithsonian Institution's plan is specific about the data that would be targeted, and its reach is not as broad as that of others; its plan covers federally funded research materials beginning October 1, 2016, and focuses on "certain" peer-reviewed scholarly publications and the associated research data. The plan includes a list of terms, such as "supporting digital research data" and "federally funded research materials," as defined by the Smithsonian to clarify what is subject to the plan and encompasses the broad community served by the agency.

2. The framework's definition of data is accepted across the agencies, making collaboration easier across myriad agencies holding diverse and heterogeneous data. The discussion of research data in the broader community often includes the question, What do we mean by data? The framework's definition answers the question for scientific data. Having a common definition is a foundation for collaborating technologically, facilitating interoperability, and aligning scientific paradigms across domains to encourage innovation and new science.

The agencies' plans suggest that scientific data have been adequately defined so that multiple agencies holding diverse and heterogeneous data can use them. Agencies are acting on the definition put forth in OMB Circular A-110.[7] Interestingly, the Smithsonian Institution does not include the term *data* in its definition list, although the Smithsonian's research includes, but is not limited to, the fields of science, history, art, and culture.

3. Some agencies' public access plans have well-defined boundaries for the scientific data to be included and specifically identify the types of data to be excluded. Well-defined boundaries can facilitate cross-agency cooperation, as they synchronize the concept of data.

OMB Circular A-110 excludes trade secrets, commercial information, and personnel and medical information from research data.

7 OMB Circular A-110 (Revised 11:19:93 Amended 9:30:99) defines research data, which is used in all the agency plans, as follows:

> Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples).

In addition, it excludes preliminary analyses, paper drafts, plans for future research, peer reviews, physical objects, and laboratory notebooks (unless the information in the notebooks facilitates data re-use). The definition allows the agencies to apply a common set of criteria to identify data that should be included in the open data collections. In addition, the list of excluded items reduces ambiguity and avoids the complexities that arise from different data types.

Because the Smithsonian Institution includes only data attached to a peer-reviewed publication in its data management plan, the peer-review process establishes the boundaries for data for the Smithsonian.

4. Collaboration is discussed across agencies, with PubMed Central emerging as a widely adopted platform for research articles. Data.gov[8] plays a vital role as a foundation for collaboration in exposing data.

The framework identifies collaboration as important for the future, and agencies are exploring collaborative activities. One of the most powerful enablers of collaboration is the PubMed Central platform. Of the 21 plans reviewed, 8 use or will use the PubMed Central platform.[9] Three of these agencies, NASA, NIST, and the Department of Veterans Affairs (VA), are not associated with the Department of Health and Human Services (HHS), which increases the scope of data available for public searchers and suggests the possibility for disciplinary cross-pollination for researchers. A tenth agency, NOAA, plans to build a data repository based on the Stacks platform of the Centers for Disease Control and Prevention (CDC). Such a repository could foster significant collaboration because, according to NOAA, all of its data are environmental data. The presence of environmental data on a common platform for health research could encourage data integration that enables researchers to address questions in new ways.

There is already a foundation for collaboration in exposing data. The office of the Assistant Secretary for Preparedness and Response (ASPR) has said that its metadata documents will be publicly available on Data.gov. CDC already has some data available on Data.gov, as well as on other sites. HHS is establishing a partnership to expose metadata in Data.gov. The PubAg model of the U.S. Department of Agriculture (USDA) includes Data.gov, and DOE says that its Public Data Listing is routinely harvested by Data.gov. DOT, the Environmental Protection Agency (EPA), the Institute of Museum and Library Services (IMLS), NIST, and the National Science Foundation (NSF) have also commented on how they will interact with Data.gov. Other agencies, such as USGS, participate in Data.gov, but their plans do not explicitly mention such participation.

8 Data.gov is the official U.S. government site providing public access to federal government data sets.

9 The eight plans are those of the Agency for Healthcare Research and Quality, the office of the Assistant Secretary for Preparedness and Response, Centers for Disease Control and Prevention, Food and Drug Administration, NASA, National Institutes of Health, NIST (as mentioned in the Department of Commerce plan), and Department of Veterans Affairs.

5. Implementation of the Open Data Policy will not occur simultaneously across agencies. Agencies are showing different levels of responsiveness to the Open Data Policy. This suggests the United States will not have a cohesive open data policy for at least several more years. This could have implications for how scientific inquiry advances and the extent to which the "grand challenges" identified by the scientific community for scientists and engineers are addressed.

Agencies are showing different levels of attention to the Open Data Policy, both in their responses and in their timetables for implementation. Several agencies have said that fiscal or other constraints could change their proposed timetable or inhibit their ability to implement their plans.

It has been more than three years since the agencies were directed to respond to the executive directive. As of April 2016, three—Department of Education (DoED), Department of Homeland Security (DHS), and Department of Housing and Urban Development (HUD)—had still not made their plans public; it is unclear if they will have feasible plans by the end of fiscal year 2016.

There is a wide range of implementation dates in the 21 plans analyzed. NIH was an early innovator in supporting public access to research data and has made published articles and data available since fiscal year 2008. Of the remaining plans that have been developed, several had already begun implementation in early 2015 (CDC, USGS, NSF, and VA). Most plans are being implemented in fiscal year 2016, with many implemented in October 2015 (Agency for Healthcare Research and Quality [AHRQ], DOE, HHS, Food and Drug Administration [FDA], NASA) and others later in that fiscal year (ASPR, USDA, DOT, and NOAA). The Department of Defense (DOD) stands out since implementation is not scheduled until late in calendar year 2016.

6. Agencies understand that their own research data management planning is part of a larger vision for the future to enable a Research Data Commons for researchers and the public. The data management planning documents suggest that most agencies see the data generated by researchers at their agency as part of a larger canvas and that the Commons would operate on the FAIR principle—Find, Access, Interoperate, Re-use. The Research Data Commons is an ambitious initiative considering the technical and sociocultural challenges surrounding its interoperability and the challenges associated with re-use, including data citation.

Seven agencies (AHRQ, ASPR, DOD, FDA, NASA, NOAA, and USDA) address the need to develop a Research Data Commons that would provide tools to facilitate the discovery, access, and use of data from across multiple agencies. Sharing through academic commons has a rich tradition that has resulted in vibrant scholarship. The Research Data Commons is being conceived on this foundation and is a promising part of the formal discussion.

Roles and Responsibilities

7. All agencies except IMLS and NSF have libraries or data centers,[10] but the role of the agency library or data center is rarely evident in these plans. The failure to cite the role of the library and established data centers suggests that some components that could serve as an important part of the infrastructure may be missing from the planning for the open data initiative.

NOAA's Central Library, the National Agriculture Library, and the National Library of Medicine (NLM) were the only agency libraries to be named specifically in any of the 21 plans. NOAA's Central Library is responsible for establishing its institutional repository; it has an active role in capturing or creating metadata for NOAA-funded, peer-reviewed publications and for developing its publications policy. The National Agriculture Library is handling petitions for changing the 12-month embargo for government-funded research publications and has provided a working capital fund to develop the PubAg system. NLM is well recognized for developing several key systems, including PubMed Central and the NIH Manuscript Submission (NIHMS) system, which are being adopted across agencies.

Data centers have even less visibility in these plans. Three data centers are mentioned by name: the AHRQ Data Center, the NOAA National Data Centers, and the CDC's National Center for Health Statistics (NCHS) Data Center. NSF mentions the general concept of disciplinary data centers. Although NASA implicitly references its Distributed Active Archive Centers (DAACs) with a brief mention of individual archives, there is no explicit reference to these robust centers, which have grown and matured over more than a decade and could serve as an important part of the infrastructure.

8. Some agencies note the role of education, but the importance of education is not prominent across plans. Many plans make no specific mention of education, and those with an education component approach it in one of two very different ways: either educating agency employees as a means of efficiently and correctly implementing the policy, or educating researchers as a means of moving science forward. Because good data management behaviors can lower the cost of managing data, adopting best practices for those behaviors requires educating both agency employees and researchers.

The framework mentions education as an important component to implement the Open Data Policy, yet eight of the plans reviewed make no specific mention of education. Thirteen plans do mention education (i.e., those of AHRQ, ASPR, CDC, DOD, DOE, FDA, HHS, NASA, NIH, NSF, NOAA, USDA, and USGS).

Four agencies (NIH, NOAA, NSF, and USDA) plan to create training programs for various stakeholders about open data, including data management. Some agencies may have this type of training already inculcated in their culture, so it is not explicitly stated in

10 The missions of IMLS and NSF are somewhat different from those of the other agencies in that IMLS and NSF are tasked with advancing knowledge boundaries primarily by funding proposals with limited-term grants.

their plans. USGS has an active data management training program that is referenced in the plan document. The DOT plan includes education-related activities, such as data challenges,[11] but does not explicitly outline education plans.

The five agencies with more developed training plans in their documents use a range of approaches. NIH already awards training grants and has outreach programs designed to familiarize researchers and librarians with NLM databases. In 2012, NLM established training to use big data as a priority and included it as a component of NIH's Big Data to Knowledge Initiative (BD2K), which focuses on training needs and the mechanisms for training researchers. In 2014, several BD2K awards were made to develop training and education approaches for scientific data analysis and management. Other initiatives are being considered, and programs are being developed to train staff and peer reviewers to evaluate better data management plans.

NOAA's National Data Centers are developing training and tools for a range of skills, including metadata creation and metadata verification. Other educational activities include an annual environmental data management workshop, free metadata training classes at one of the centers, and providing funding to the Federation of Earth Science Information Partners (ESIP) to support data management training and regular meetings. NOAA also monitors several groups developing training and uses appropriate resources from these groups. NOAA participates in interagency training activities as well.

NSF already has a robust structure of policies and solicitations regarding training and workforce development. One strategy supports programs about data and data management at information schools. NSF has launched six activities related to its Data Science Priority Goal, including developing solicitations that can be vehicles for data education and training, and conducting workshops.

USDA targets its education and training activities to four groups of major stakeholders: (1) USDA science support professionals, (2) administrative professionals, (3) leaders from scientific societies and professional organizations, and (4) USDA intramural/extramural scientists. Outreach activities include awareness presentations, collection of stakeholder input, and meetings. The three-phase approach to implementing these activities includes developing outreach and training plans, developing modules and workshops, and creating a training module that can be delivered through the USDA online learning university, AgLearn.

USGS' plan points to its data management website, which includes a section for training and resources. The training section has three interactive modules that are designed to help researchers, data managers, and the public learn about scientific data management and introduce best practices.

11 Data challenges are activities sponsored by DOT that ask the public to create a tool or use data in an innovative way. Winners are chosen by a DOT panel using a rubric that is shared with participants.

Good data management behaviors can lower the cost of managing data. The adoption of best practices in data management behavior requires educating both agency employees and data creators. Our findings suggest that agencies need to give education more attention.

9. Although there are mature solutions for access and storage of peer-reviewed publications, the solutions for access and storage of data are immature; thus, it requires considerably more effort from prospective users to obtain data. The intense effort required from the researcher creates a barrier to achieving the broader vision of the mandate and could impede the ultimate impact of the public access plans for data.

There are several well-developed strategies for making researchers' journal articles openly available. PubMed is a frequently used system that requires little effort from the researcher. A researcher can upload his or her article in a simple process, or it may be automatically ingested based on agreements with publishers. Access to peer-reviewed publications, which has been a significant challenge, is far less complex than access to research data. Providing PDF copies of papers does not require dealing with all the difficulties presented by the breadth and range of data formats. There are currently few mature strategies for making available the data underlying a journal article. Additionally, such strategies require more effort from the researcher who generates the data. For example, time must be spent on metadata creation and verification before a data set can be uploaded.

10. Discussions of the cost of open data and the recovery of this cost are not well developed at many agencies. The cost of data management is an essential consideration in designing a public access plan for data that is sustainable over time. Sustainability has been a frequent topic in the digital curation community,[12] which suggests opportunities for increased dialog between this community and federal agencies.

Even though cost is a key element, 6 of the 21 plans do not address cost at all or make only the briefest reference to the need to consider the monetary and administrative burden. Five agencies' plans state that researchers could or should include a budget item for the cost of data management.

More specific discussions of cost models or funding streams appear in plans from the USDA, DOE, FDA, NASA, NOAA, and NSF. The USDA plan has a very thorough discussion of costs. FDA notes that annual funding for data management comes from the Office of the Commissioner, and DOE has these costs in its budget. NASA not only has developed a funding model that will be included in the annual budget, but also appraises the balance between cost and value for each data set.

DOD includes data management in its current budget, but could delete this item if it is not feasible in future budgets.

12 Some examples are the Blue Ribbon Task Force (see http://blueribbontaskforce.sdsc.edu/biblio/BRTF_Final_Report.pdf) and the European Union's Collaboration to Clarify the Costs of Curation (4C) project (see http://4cproject.eu/).

Making Data Public

11. The plans for public access vary in how agencies approach empowering the public with the data. Most plans note that peer-reviewed articles will be freely available in a repository no more than 12 months after publication and that data supporting the article will be available between 12 and 39 months after publication. Some agencies move beyond simple discoverability and accessibility to a discussion of the need to build an environment including tools to interact meaningfully with the data.

DOT states that it is important to engage the public with the data by creating visualization platforms for public users to manipulate and better understand the data. For example, it is likely that geospatial data will be stored in a cloud-based repository, and cloud services could be used to meet the requirement for providing public visualization capabilities usable by all levels of government, the private sector, and the public. Other agencies, such as NOAA, broadly mention the need for advanced dissemination features, but are not specific about the tools.

12. Metadata are essential for access. Nearly all the documents note the importance of metadata for discovery and access, and the approaches outlined in this area range from general to quite specific. For agencies with a broader research spectrum, metadata must meet appropriate industry standards. Some plans call for developing modules or services to manage metadata generation, acquisition, and quality control.

DOT, IMLS, and VA do not discuss metadata at all, while the U.S. Agency for International Development (USAID), which was not subject to the federal mandate, addresses metadata in its plan.

The following are recurring themes relating to metadata:

- The data management plan must identify the standards used for the metadata.
- The data set must have a formal metadata document. Many plans specify that the metadata document must identify the agency as a funding source and require that the metadata document be reviewed and approved to verify that the researcher has met agency requirements.
- Metadata for the data set must include the common core from the schema used by the federal government (found at the Project Open Data website).
- Metadata must be supplied for publications. Many plans note that publication metadata should link to the publication; many of these publications will be made available via PubMed.

Often there are requirements to make it easier to include metadata in the agency metadata catalog.

Conclusions

From the analysis of the federal public access plans and the experiences of the federal agencies creating them, several conclusions can be drawn. These conclusions are valuable to cultural heritage institutions for three reasons. First, the conclusions provide a concise overview of lessons learned by federal agencies as they developed plans for and implemented public access to the articles and data that result from federally funded research. Second, they identify topics that should be addressed as the cultural heritage community considers the challenges and opportunities of curating data. Third, they suggest that the mandate has created opportunities for cultural heritage institutions to both build upon and contribute to the infrastructure being developed by the federal agencies.

The conclusions are presented in the three groupings introduced earlier because they can be traced to the analysis discussed in each of those sections and because each of these three groupings represents a different focus. These groupings allow cultural heritage institutions to address the topic from three very different but related perspectives. The "Open Data Infrastructure," focuses on the broader context for open access to articles and data. "Roles and Responsibilities" focuses on who is engaged in the open access activities and what they might be doing. "Making Data Public" focuses on how the desired result of openness to the public is best reached.

The Open Data Infrastructure

1. **The value of data lies in their use.** Just as interstate highways have improved the nation by creating access for commerce, publicly accessible data are an important component to improving economic and societal well-being (NRC 1997). They serve as a vital element of our "epistemic infrastructure" (Hedstrom and King 2006). Building the open data infrastructure should be a national priority insulated from the influence of politics and treated as a vital national asset.

2. **Exemplars can be powerful.** Agencies with successful approaches can provide leadership and vision to others. Numerous agencies have valuable firsthand experience with challenges and solutions that can be instructive to other organizations seeking to implement open data initiatives.

3. **Changing organizational culture is difficult.** It is necessary to change the organizational culture to fully implement the mandate, yet the required scale of change is challenging because it involves a variety of stakeholders, including many outside the cultural heritage sector. Information professionals in the organization's library, archives, government data center, or other parts of the organization can help with this transition given their knowledge of the data life cycle and their understanding of the information behavior required of data producers and data consumers. Information professionals know data standards, can help with metric creation and monitoring, have experience managing digital content, and understand how to

meet the information needs of users. Therefore, they can help reduce the amount of time wasted on "reinventing the wheel."

4. **Information professionals and the cultural heritage community have a vital role to play in developing the Research Data Commons.** The proposed federal Research Data Commons is premised on the academic commons model, which has a rich tradition of facilitating vibrant new forms of scholarship.

Roles and Responsibilities

5. **Federal agencies need ongoing support as they transition to a culture of open data.** The federal resource, Project Open Data, could help agencies make the cultural shift necessary to "managing information as an asset." This website provides valuable links to definitions, implementation guidance, tools of many kinds, resources, case studies, and other ties to the open data community. When the White House developed it, the site was envisioned as a community-maintained resource. However, a review of activity shows that efforts to populate and maintain the content have been uneven and do not represent a broad community of users. Building awareness of this resource and encouraging more community input would facilitate best practices for open data.

6. **Libraries, archives, and government data centers should be involved when public access to data is discussed and plans are implemented.** Evidence from observing the broader community suggests that many information professionals are already involved in the process, but their involvement is not clearly described. The role of libraries, archives, and government data centers and the information professionals who run them should be explicitly stated in the plans for the future because of the expertise and knowledge they can provide.

7. **Data sustainability needs more attention and discussion.** Historically, government research funds have not been directed toward maintaining data management infrastructure. Successfully meeting the goals of the data mandate will require ongoing investment in infrastructure. Current agency plans offer few concrete strategies for keeping data accessible and usable in the long term. This issue must soon be addressed more completely if the data are to be accessed and repurposed, and their future value maintained.

Making Data Public

8. **Relevant stakeholders should work together to educate the public on ways to share and re-use data.** Data sharing and re-use add value to a resource that has already been collected. To maximize this potential added value, it is essential to raise awareness of ways to share and re-use data. Education and outreach efforts, which only a few agencies note in their public access plans, are required; this is an area where libraries and information professionals could provide essential support.

9. **The role of education should be better defined in the public access plans.** Some plans focus on educating federal agency staff, while others focus on educating data users and producers. It is crucial to educate all stakeholders. There is a need for educating and training individuals to take on particular roles as data curators both for those who receive research grants and for the agencies.

10. **The issues surrounding public access to publications and data should be disambiguated.** The solutions for creating public access to data are still mostly nascent and need the greatest effort, attention, and support. The solutions for creating public access to publications are more mature and ready for implementation across agencies.

11. **Infrastructure to support ongoing data discovery, access, analysis, and sensemaking is necessary for data-driven research and innovation.** Data.gov, which serves as a national catalog for open data sets, is helping to make data visible, but more tools and services are needed. Increased attention should be given to U.S. Open Data, which matches data producers and consumers to create sustainable data ecosystems but has seen relatively little use.

12. **Metadata are critical to public access; metadata creation must be improved.** The analysis of public access plans revealed important recurring themes regarding metadata: (a) data management plans should identify standards used for the metadata; (b) data sets should be accompanied by formal documentation about the metadata; (c) metadata for data sets should include the common core from the schema used by the federal government;[13] and (d) metadata must be supplied for publications.

These conclusions also provide a foundation for thinking about existing digital curation activities and successes (Part 2) and for the need to build the workforce capacity (Part 3).

13 E.g., Project Open Data.

PART II
Implementing Digital Curation: Project Interviews

Christopher Lee

A primary objective of this report is to inform the cultural heritage sector—including the Institute of Museum and Library Services (IMLS) as an essential player in this space—about the implications of the federal government's public access efforts for digital curation activities. Toward this end, it is important to have a general understanding of the requirements and needs for digital curation. We conducted interviews with leaders of seven projects previously funded by IMLS to identify lessons about skills, capabilities, and institutional arrangements that can facilitate digital curation activities. This investigation—focusing on the experiences of professionals who have engaged in digital curation work—complements the content analysis in Part 1, which focuses on the aspirations of government agencies as revealed in the text of their public access plans.

Research Design and Methods

The investigation undertaken in Part 2 was based on a case study research design. We identified seven recent (2010–2013) IMLS-funded projects that included significant digital curation objectives, which could include management, preservation, or provision of access to digital information. The sampling frame aimed for diversity of project objectives, curation functions, and data types.[14]

Our investigation was based on multiple data sources. The primary data source was a set of semi-structured interviews with key

14 As noted in our later discussion about limitations, six of the seven projects that we investigated were administered in universities, and the seventh was run at the New York Public Library, an institution that operates very much like an academic library. Therefore, this study cannot speak directly to any unique issues confronted by federal government agencies, but it does provides insights into the challenges and opportunities related to managing and providing public access to digital data.

project personnel. We conducted one interview per project for a total of seven. Six of the interviews were conducted with a single individual (usually the project's principal investigator), but one interview involved two individuals. All interviews were recorded and transcribed. They lasted between 20 and 49 minutes (average of 37 minutes). Table 2.1 summarizes the seven projects. In addition to conducting the interviews, we analyzed project documentation and (when applicable) online products of the projects.

Table 2.1 Investigated projects that were funded by the Institute of Museum and Library Services

Project	Primary Focus	Lead Institution	Interview Participant(s)
Creating a Better World by Sharing Research Online	Institutional repository (IR) to provide access to the university's research output	Southern New Hampshire University	Cathy Growney
Databib	Annotated online bibliography of research data repositories	Purdue University	Michael Witt
Datastar	Study researchers' data sharing and discovery needs and enhance a linked data platform to meet those needs	Cornell University	Huda Kahn, Mary Ochs
ETD [Electronic Theses and Dissertations] Life Cycle Management	Guidance documents and software tools for life cycle data management and preservation of ETDs	University of North Texas	Martin Halbert
Improving Data Stewardship with the DMPTool	Identification and proposal of strategies to address challenges in the adoption of the Data Management Planning Tool (DMPTool)	California Digital Library	Patricia Cruse
Virtual Archiving for Public Opinion Polls	Demonstration and promotion of streamlined workflows for getting research data into data archives	University of North Carolina	Jon Crabtree
What's on the Menu? – From Software to Funware	Support for crowdsourcing of menu transcriptions	New York Public Library	David Riordan

Findings

This part of our study generated the following nine high-level findings.

1. Successful initiatives are part of ongoing capacity-building activities.

Many successful projects were building upon lessons learned and capabilities established in previous activities, including previously funded projects.[15] In fact, interview participants often found it difficult to speak exclusively of the work they had done on the specific IMLS-funded project in question, because it was often so closely tied to work they had done in earlier projects.

In turn, IMLS-funded projects investigated in our current study have themselves often provided an important foundation for future work. For example, one interview participant noted,

15 A similar finding had emerged from a previous investigation of state digital preservation projects that had been funded through the National Digital Information Infrastructure and Preservation Program. See Lee 2012.

> We didn't really have a strong archive program on the campus. I didn't have an archivist. And, thanks to this program, I now have a Digital Initiatives Librarian who is also our archivist. And we have become known on campus as the place to send stuff. So, all of a sudden, we are getting a new building . . . and it actually has an archives space. We currently, in the old facility, have literally a coat closet for archives. So now, our archives have just exploded. . . . Our robust digital archive spawned a robust print archive.

Similarly, the What's on the Menu? project was instrumental to further research and development laboratory work at the New York Public Library.

2. Digital curation requires control over software.

Managing and providing access to digital data requires a variety of software elements. Professionals responsible for digital curation must, therefore, establish proper control over that software. In doing so, they may find it necessary to develop completely new software, customize existing code, use existing tools, and undertake various aspects of configuration management to ensure that changes to one part of the system do not adversely affect other parts of the system.

The projects investigated in this study represented this full range of software control activities. Some, for example, hired full-time programmers, while others relied entirely on existing tools to support their work. It is important to recognize that the customization of existing software can involve a substantial amount of programming. One participant, who indicated that "our IT [information technology] support did customization [of an existing system] but we didn't actually create any specific software," also said that "the need for programmer knowledge was a little bit higher than expected so we called more on IT than I would like. I would rather have someone in-house who was able to do that." Although having local development expertise can be beneficial, this does not mean that institutions must rely solely on software that has been developed in-house. According to Jon Crabtree, from the Virtual Archiving for Public Opinion Polls project, the essential thing is often to look at existing software "and say, 'I know it didn't do it for me, but I need to fix that part' and being able to write some automated tools to do that." The Virtual Archiving project took such an approach, creating enhancements that were then "rolled" into the DataVerse Network. David Riordan from What's on the Menu? also pointed to the importance of "data parsing, being able to extract information and metadata from other sources, being able to work with data that is in computable form." Databib, for many aspects of its system, "borrowed and appropriated liberally as opposed to just building everything from scratch." However, when it came to the review workflow, staff determined that "it would have been more work to incorporate one of [the existing applications] than it was to build them from scratch."

Building on existing software was a major theme from the interviews. Existing tools upon which the project staffs relied included both open source tools (e.g., Apache, ClamAV, Dataverse, DROID,

DSpace, Elasticsearch, JHOVE, Linux, MySQL, PHP, Ruby on Rails, R, Shibboleth, Solr, TomCat, VIVO), commercial tools (e.g., ABBYY Fine Reader, Acrobat, Audible, Confluence,[16] SAS, SPSS, STATA), and online services (e.g., Dropbox, GitHub, GoogleCode, SourceForge). Regardless of what combination of software is used, setup and integration are often quite resource intensive. For example, Martin Halbert pointed out that the ETD project required a substantial investment of time and effort to identify various "digital preservation tools that are out there and then actually installing them and getting them to work together, trying them out."

A concept that has gained considerable attention in the digital curation arena in recent years is that of microservices, which have been popularized by several product/service providers, including the Data-Intensive Cyberenvironments (DICE) group, Artefactual Systems, and the California Digital Library. Microservices are small, focused pieces of software that can be used to perform specific, discrete actions. Rather than creating a single, relatively monolithic system to be used by everyone, information professionals can combine microservices in various ways to meet the needs of particular institutions. Halbert from the ETD project explained this approach as one in which, rather than developing a complex set of "de novo" software, "you make things modular, you make them freestanding, you have good APIs or protocols to hand one thing to another." The project team attempted to define their products in terms of "a series of software microservices for addressing particular life cycle management functions in administering ETDs." They were responding to the particular software ecosystem that they faced, in which there were many existing "ETD software packages and environments out there," and "there was no hope" that they could develop separate modules for every one of them.

In some cases, existing tools and systems served as important models and sources of ideas, even if they were not incorporated directly into the project's own software products. For example, the ETD project looked at Vireo (an open source ETD management system) in this way.

3. Effective digital curation involves not only working with data, but also actively engaging with relevant stakeholders.

The leaders of the projects under investigation had a strong sense of who their primary stakeholders were and made concerted efforts to engage them. The primary stakeholders that they identified by the end of the project were not always the same as those expected at the outset. For example, the ETD project was originally designed primarily to meet the needs of library staff within universities. The project team did not anticipate that a significant portion of their audience would be "people associated with the graduate school," who are responsible for processing student records.

16 Confluence is proprietary, but is offered for free to qualifying open source projects.

Building an effective system requires not just technical development; it also requires marketing and outreach. In discussing the institutional repository at Southern New Hampshire University, for example, Cathy Growney emphasized the importance of "partnering with the IT department," as well as including "a lot of people from across campus" on an implementation team and a separate policy committee. More broadly, marketing "was a huge thing and getting it out there and constantly having conversations" with "key stakeholders." The compelling collection and the innovative interface contributed to the success of the What's on the Menu? project, but "getting that message out" to potentially interested populations played a big role as well.

There are a variety of ways to engage with stakeholders, both online and face-to-face. Project staffs used methods ranging from the formal (e.g., user testing, cognitive walk-throughs, interviews, surveys, focus groups) to the informal (e.g., discussions with colleagues). Many of the projects engaged in outreach at conferences and other professional events, and they held meetings and workshops as well. Three of the interview participants commented that if they had had more resources for their projects, they would have added a face-to-face meeting with their primary data contributors.

Projects not only engaged with relevant stakeholders; they also generated resources that professionals can use to support their own engagement activities. Among the products of the DMPTool project, for example, were guidance for librarians who wished to put on their own brown bag events in order to spark discussions with campus partners; a slide deck for librarians to present to researchers, as well as other promotional materials that they could customize for local use; a startup kit for doing an environmental scan of institutional resources and services; a webinar series for librarians; and a set of case studies of institutions that are using the DMPTool. A major goal of these activities was "building confidence" so that librarians would be able to engage with relevant stakeholders and "step into the science, technology, and biomedical sectors of digital curation"—areas in which they might not have previously received much education or preparation.

One type of stakeholder is the individual involved with allied projects and initiatives. Recognizing areas of overlap and opportunities for coordination can be very important. The Datastar team, for example, drew from the work of the Australian National Data Service (ANDS), which was pursuing similar goals. Databib in the United States and re3data in Germany started at about the same time, and recognizing that they had similar goals, their leaders have actively communicated across projects and ultimately entered into an agreement to merge their platforms under the auspices of DataCite. Similarly, the DMPTool project leaders worked closely with the leaders of the Data Curation Centre in the United Kingdom, which provides a related tool called DMPOnline. Such engagement requires active monitoring of the environment for other activities that are under way elsewhere.

4. Making the case to resource allocators is a key factor in many settings.

In most digital curation initiatives, institutional leaders who make resource allocation decisions are very important stakeholders. One of the key sets of activities in the institutional repository project at Southern New Hampshire University involved "marketing and advocating" and "talking to the academic leadership." Halbert noted the importance of getting "the emerging issues with ETDs and the related issues of research data management in front of academic decision makers, especially presidents and provosts." Fundamental needs of the ETD project's audience were "a local advocacy issue. How do they advocate for support through their university administrations for a localized ETD program?" The DMPTool project staff held a two-day meeting "to identify resources that would be most helpful for [institutions] in using the DMPTool for conducting outreach," and one of the issues that repeatedly emerged was the "lack of support and education at the institutional level related to data curation." According to Patricia Cruse, "Once you have somebody at the top saying, 'this is a priority,' it can open doors." So "people need to communicate obviously with the researchers on the campus, but also with the vice-chancellors for research" and other "high-level administrators." With all of this said, the role of line staff in carrying out the work remains essential. Cruse pointed out that one danger is for a fairly high-level administrator to decide to directly take on the role of implementing something like the DMPTool and then find out that he or she does not have the time to actually carry out the work; in this event, "things peter out" unless someone else can pick up the tasks.

5. It can be beneficial to release prototypes early, so they can be tested with real data.

As discussed previously, various forms of stakeholder engagement can be essential to the success of digital curation efforts. One particularly valuable form of engagement is to have potential users interact with the intended deliverables, whether they are systems, applications, or documents. The ETD project, for example, held brown bag lunch discussions at participating institutions to collect feedback on guidance documents and other products. This also helped reduce "the variability of the actual rollout of the content" at those institutions, because they had already had a chance to discuss the content with others.

Self-reported needs (e.g., those elicited from surveys, interviews, or focus groups) can be very revealing, but they are not always accurate representations of user behaviors. One interview participant observed that "when it came down to actually interacting with the interface, that is when your feedback seemed almost diametrically opposed to what you heard earlier."

Early prototyping and testing can be an excellent way to ensure that development is moving in a direction that is likely to benefit users. For example, the Databib team attempted to "get things into

code as quickly as possible to implement them," with the expectation that "you are not going to be perfect [so] you are going to put it out there, and the people are going to give you feedback, and you are going to iterate and improve and sail forward." Once a beta version of the system was "online with a hundred records, we put out a call for editors." Having input from the editors was a valuable way to identify further development priorities.

6. Meeting user needs involves many inferences about their behaviors and expectations.

As noted earlier, analyzing user needs often involves mechanisms such as user testing, interviews, surveys, and focus groups. One tool for eliciting the needs of data creators is the Data Curation Profiles Toolkit, the enhancement of which was a primary focus of the Datastar project. Another is the DMPTool, which one interview participant characterized as "a 'gateway drug' for librarians as well as researchers." Use of such tools can be valuable in testing assumptions and in identifying design priorities and opportunities for improvement. However, even with such resources at hand, it is rarely possible to elicit data directly on all aspects of system or process.

One way to make inferences about user needs is to rely on information professionals whose experiences working with specific populations allow them to serve as proxies for users. For example, one interview participant indicated that a major source of guidance on data curation user needs in their project was "working as a librarian on the front lines" and consulting "with researchers who are putting together data management plans for the first time." Another participant stated that "one of the things that we relied on was knowledge of staff of how researchers work, what their tolerance is for reading directions and engaging with things."

Information professionals with knowledge of information practices within a domain can serve as proxies for users by providing "reality checks" on what sorts of actions users would likely be willing to perform. This question comes up frequently in terms of how much and what types of metadata users will generate, as well as what types of documentation they would be willing to read. Early in the Datastar project, for example, "one of the comments we had from librarians that were looking at the project [was] 'There is no way you can get someone to fill in all of that information.'" One of the fundamental challenges is "pinning down what would be most relevant to a user and then comparing that to what they would actually fill out." The DMPTool project benefited from the experience of a colleague who formerly worked as a researcher and would often say, "Researchers are not going to read that. Simplify it."

7. Metadata satisficing[17] is essential.

There is significant value in defining clear metadata conventions (e.g., schemas, ontologies, data dictionaries), and information professionals are well positioned to develop such conventions. Metadata enhancement, cleanup, and transformation can require substantial resources. Those working in social science data repositories, for example, must often consult the codebooks associated with studies to determine what the full labels for given values should be. A survey question could be a full paragraph or longer, and Crabtree of the Virtual Archiving project pointed out that a major challenge when dealing with quantitative data formats such as SPSS, SAS, and STATA is that they often have "truncated value labels." When working with metadata associated with their menu collection, the New York Public Library staff had to do "lots of nasty, nasty regular expression parsing" to re-use metadata that had been created previously for other purposes.

No project and no institution has unlimited resources, so it can be important to maintain the flexibility to accommodate metadata that do not fully conform to the ideal. Digital curation professionals must make numerous decisions about metadata trade-offs. One interview participant observed that there was a need within their project to avoid "overcomplicat[ing] things," adding that "the tendency for librarians is to do everything, throw everything at the problem, and help as much as you can rather than simplify."

One fundamental choice is between the following three options: (1) insist that those submitting data to their systems conform to strict metadata conventions when they submit the data; (2) accept "sloppy" metadata, but then engage in substantial cleanup activities to ensure that the metadata ultimately conform to strict metadata conventions; or (3) establish metadata conventions that are more flexible and tolerant of variance within the values. Participants in our study had adopted approaches that involved various combinations of all three options.

A common strategy is to identify a relatively limited core set of metadata elements that can then be extended in particular cases. The staff at the Databib project, for example, wanted to establish and cultivate a wide range of contributors to their system, so they decided to lower "the barrier for people to actually make submissions" rather than "implement metadata control at every turn or to ask for the sun and the moon and the stars." They aim for a minimal set of "elements [that] have proven to be the most important" (i.e., title, URL, authority that maintains the repository, and at least one subject term). The system's editors then "can expand on that and do a little

17 *Satisficing* is a term introduced by Herbert Simon in the 1950s to characterize a decision-making process that involves settling on an option that is "good enough" to meet a certain threshold of acceptability (called an "aspiration level"), rather than attempting to find a single optimal solution to a problem. It applies particularly well to decisions about metadata, because it is impossible to predict precisely which metadata elements will be most valuable in the future. However, it is possible to make educated guesses about the types of metadata that are likely to be valuable.

bit of research and flesh that out more." Similarly, the ETD project staff developed a submission drop-box approach that requires "core metadata elements and also made it extensible so that projects that might adopt it could add their own metadata fields that were important locally for administrative purposes."

In some cases, the best approach is to identify a fairly extensive set of desirable metadata elements, but then maintain the flexibility to accommodate items that provide a more minimal subset. The Datastar team found it important to recognize that "not every single part of this information is going to be something that is going to be populated." The aim was to identify "the simplest amount of metadata we can capture that would allow for the best discovery or enable discovery and access in a meaningful way."

Several of the interview participants expressed the desire to provide better systems for data producers to generate metadata. In particular, many of them suggested systems that could infer the relevant categories or elements based on attributes of the data being described. One suggested "built-in inferencing" in which "they show the user options; they don't just open and expect that you fill in everything from the start."

It is important to determine not only what metadata should be captured and created, but also what subset should be exposed to users. For example, many interview participants discussed the Semantic Web and the potential value of the relationships that can be represented using the Resource Description Framework (RDF). However, the simple presence of extensive metadata does not guarantee that it will be beneficial to end users. There can be a "gap between what can be represented and what you find the users would actually like to work with." The information that is exposed to users through an access system "can't be all the information you need to re-use the data because that would mean literally all of the information about the data that is available."

8. Open access involves not just enabling discovery of data, but also enabling new forms of interaction with and among users.

Responsibility for the provision of access to data does not end with putting the data and associated metadata on the Web (no matter how good the metadata might be). Effective data use can involve a variety of interactive mechanisms. In addition to those that allow users to search and navigate through an institution's website, interview participants cited mechanisms such as RSS feeds, Twitter (which "drives traffic to the record"), and Google spreadsheets populated with data as helpful in enabling new forms of user interaction.

User interaction can generate additional metadata and documentation. The What's on the Menu? project was successful in using gamification to encourage users to contribute metadata about their menus. The Databib project provides a "dashboard" for assigning new submissions to up to three editors. Part of the system's functionality is to ensure that "if someone else isn't getting to something, someone else can step in and do it."

Several of the interview participants pointed out the potential for facilitating further interaction *between* users. For example, the Datastar project team found "a positive response attracting user comments as long as there was some amount of moderation." Riordan from What's on the Menu? pointed out that user-to-user interactions can help build a sense of community, but the lack of an authentication process on the server prevented them from enabling such interactions as they would have liked. "Allowing them [users] to communicate with each other" is something that they still hope to address.

9. Push digital curation strategies into data producer practices and behaviors.

An essential aspect of digital curation that relates to many of the findings is the interjection of digital curation knowledge and methods into the information life cycle as early as possible. For example, in the Virtual Archiving project, the team planned to help professionals engaged in polling to "embed the process into their polling process so that it was an automatic thing." Crabtree pointed out, "if you can catch [metadata problems] while they are fresh, it is a lot easier."

Implications

In addition to the detailed findings above, we would like to highlight a further high-level implication of this part of our study for IMLS and interested stakeholders: There is potential for much better coordination between work on data management plans and work on access strategies and systems.

In Part 1 of this report, we discuss open access provisions and practices of several government agencies, and in Part 2 we discuss projects that address data management plans. Both represent important areas of professional progress. More generally, access to the data is a standard consideration in data management plans, and the need to ensure continuing access to research data has sparked many fruitful discussions among researchers, academic administrators, and funding agencies.

However, there is often a disconnect between the discussions of data management plans and discussions about government public access activities. We see potential for further collaboration and integration of these efforts. Professionals engaged in public access initiatives (most often conducted in government agencies) can learn from the work in developing and implementing data management plans (most often conducted within academic institutions). Similarly, experience with public access initiatives can help to inform the data management plans so that their provision for access is most likely to be viable and sustainable.

Limitations and Future Research

We believe that both IMLS and the U.S. federal government entities responsible for providing access to research data can learn important

lessons from this study. Even so, it is important to note its limitations as well as the possibilities that it raises for future research.

The first set of limitations relates to the study's sample. By focusing on projects recently funded by IMLS that have included significant digital curation objectives, we have been able to generate findings that can inform future IMLS programs and funding priorities. In addition, we have identified issues that professionals responsible for digital curation are likely to face. Six of the seven projects that we investigated were administered in universities; the seventh was run at the New York Public Library, an institution that operates very much like an academic library (e.g., is a member of the Association of Research Libraries). Therefore, this study cannot speak directly to any unique issues confronted by federal government agencies. Instead, it provides insights into the challenges and opportunities related to managing and providing public access to digital data.

A closely related set of limitations pertains to the scope of the study. Because we focused on grant-funded projects, issues of institutional sustainability probably did not receive as much attention as they would have if we had instead focused on ongoing programs within the respective institutions.

There are numerous opportunities for future research. One would be to investigate the experiences of those engaged in digital curation projects within federal government contexts. Another potentially fruitful avenue for research is to compare the experiences of this study's interview participants with the experiences of those engaged in digital curation activities that have been funded through operating funds (rather than project funds) to further highlight institutional issues related to ongoing coordination, integration and sustainability. The present study should serve as a useful foundation and point of comparison for such future work.

PART III
Building Capacity: Curriculum, Competencies, and Careers

Nancy Y. McGovern

Many of the relevant community reports released since or just prior to the White House issuing the "Increasing Access to the Results of Federally Funded Scientific Research" executive directive in 2013 identify the need for training and education to build the requisite organizational and individual capacity to respond to the executive directive as either a core or supporting recommendation, for example:

- *Preparing the Workforce for Digital Curation* (NRC 2015)
- *Data Curation Education: A Snapshot* (Keralis 2012)
- *The Problem of Data: Data Management and Curation Practices Among University Researchers* (Jahnke and Asher 2012)
- *A New Value Equation Challenge: The Emergence of eResearch and Roles for Research Libraries* (Luce 2008)

The findings in Part 1 of this report note that of the federal plans released by the end of 2015, eight make no mention of the need for education. Those plans that do mention the need recommend implementing either (1) programs that enable employees to comply with the plans or (2) programs that have an outreach component to inform and raise awareness among researchers and improve research practice. The findings also note that raising awareness about sharing and re-using data requires education and outreach efforts. The continuing education programs to build the skills and knowledge of digital curators that are a primary focus of this section are more aligned with the second recommendation: to inform and raise awareness.

The plans of four federal agencies—National Institutes of Health (NIH), National Oceanic and Atmospheric Administration (NOAA), National Science Foundation (NSF), and the U.S. Department of

Agriculture (USDA)—note the intention to create training programs. A fifth agency, the U.S. Geological Survey (USGS), mentions its data management training program in its plan. Although other agencies may have educational programs that they do not specifically mention, the curriculum development component of this section references training examples from these five plans.

In Part 2 of this report, the third finding—effective digital curation involves not only working with data, but also actively engaging with relevant stakeholders—refers to the lack of education and support for digital curation as a challenge facing organizations.

The importance of continuing education in advancing digital curation within the cultural heritage community is evident in the significant number of pertinent projects and reports that have been initiated and completed in the United States. There is a growing and urgent need for digital curation professionals to collaborate with stakeholders in specific domains to extend and scale up programs to reach researchers in the increasing number of disciplines engaged in data curation. Continuing education programs have a potential role in competency building, curriculum development, and options for lifelong learning as the range of requisite digital curation skills evolves.

Perspectives on Education and Training

A persistent question in the digital community has been, What skills are needed for digital curation? The answer depends on who is asking and in what context. The following is a brief review of a model that parses out several perspectives that may be present implicitly in discussions of educational and training programs and options.[18] Making the perspectives explicit offers context for the results and findings presented in the remainder of this report.

Organizational Perspectives

An organization in this context may be

- a professional association that provides educational programs for members
- a cultural heritage or other organization that is using teams and collaborating with others to achieve digital curation objectives by acquiring, managing, and retaining skills

In either case, the organization will need to consider short-term planning to meet immediate educational and training needs, as well as longer-term planning to anticipate and prepare for pending and future educational and training needs. An organization that focuses only on immediate needs and does not commit to the sustained development of its staff is likely to have difficulty retaining or possibly

18 Nancy Y. McGovern developed this previously unpublished Perspectives on Skills model based on lessons learned as the director of the Digital Preservation Management (DPM) workshop series since 2003. She presented it for the first time at the Association of Canadian Archivists @ the University of British Columbia (ACA@ UBC) Fifth Annual International Symposium, February 8, 2013.

hiring staff. Balancing short- and long-term objectives requires establishing priorities to provide the means to build necessary skills and capabilities. Meeting strategic objectives may mean that individual training goals and needs cannot be met.

Team Perspectives

A team is often assigned projects with time frames that vary in length and require team leaders to

- identify requisite skills to achieve outcomes
- allocate the skills of limited human resources for multiple teams across an organization or between collaborating organizations
- incorporate the skills of consultants and contractors

Completing projects assigned to the team requires matching needs with available resources. This need influences the perspective of teams toward skills. Project teams have to gather together immediately those with the necessary skills and are typically not able to wait for team members to develop these skills on their own.

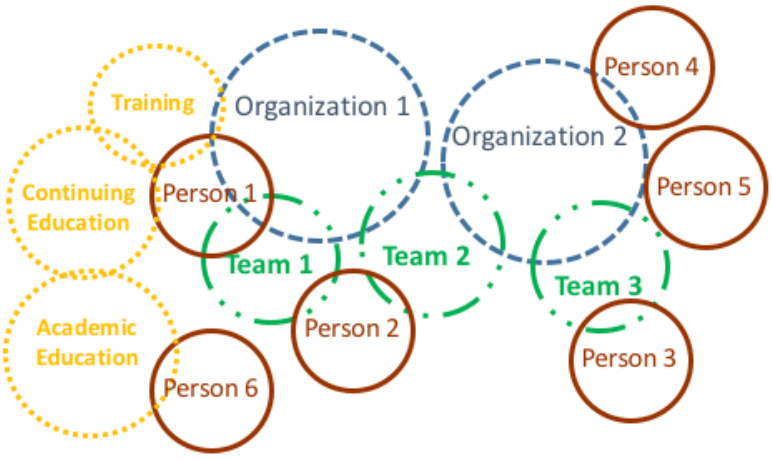
Individual Perspectives

Individuals in organizations and on teams have varying educational and training perspectives. Members may include

- individuals who are interested in expanding or deepening their skill sets through a combination of educational and training opportunities that may involve how-to training, continuing education, and academic education during their careers
- individuals who are members of multiple teams with competing needs for skills
- individuals who are specialists, possibly consultants, and are in demand based on a desired skill set
- individuals who may not be involved in current priority projects, but who require training that is suited to their responsibilities
- individuals who serve as advisors or consultants, helping organizations address strategic priorities, and who will need to maintain their skills to be effective
- individuals who decide they need or want academic education to achieve their personal objectives

Over the course of their career, individuals are likely to work in multiple organizations in a variety of roles requiring a range of skills. Community-based services are needed to guide and inform individuals about education and training that may be larger than or divergent from the needs of employing organizations. There may be connections between continuing education or training programs and organizations or teams, but typically individuals initiate or organize the courses that they take. Figure 3.1 represents these interactions.

Fig. 3.1 Perspectives on skills



Education and Training Provider Perspectives

The range of educational and training perspectives includes the following:

- *Academic education:* "education which has learning as its primary purpose" and is intended "to build a capacity to adapt and apply a 'foundation knowledge'"[19]
- *Continuing education:* "formal lectures, courses, seminars, webinars, or any other similar type of educational program designed to educate an individual and give him or her further skills or knowledge to be applied in his or her line of work"[20]
- *Training:* "the action of teaching a person … a particular skill or type of behavior"[21]

There can be significant collaboration and intersections between these three common components of education and training. Academic programs may host continuing education or training programs. Digital curation and preservation practitioners may serve as adjunct faculty members for academic programs or as instructors for continuing education and training programs. In addition, professional associations may develop and provide continuing education and training programs, and influence or inform the evolution of academic programs, especially in applied fields. For example, the Society of American Archivists (SAA) defined a curriculum for a master's program in archival studies and more recently launched the Digital Archives Specialist certification program, which is cited in table 3.1.

The needs and priorities of organizations, teams, individuals, and educational and training programs change over time. Sustainable community strategies are needed to maintain, update, and replace curriculum when necessary. The need to scale up to address

19 See http://www.acs.edu.au/info/education/trends-opinions/academic-education-explained.aspx.

20 See http://www.businessdictionary.com/definition/continuing-education-program.html#ixzz3tO5F5Pk7.

21 See http://www.oxforddictionaries.com/us/definition/american_english/training.

open access to data is an example of a shift requiring attention and resources that extend and supplement continuing education options. In considering the measures that will be required to address the need for training and skills identified in federal plans and in community reports, it is useful and necessary to keep this range of parallel and competing perspectives in mind.

Study Scope

Although investment and interest in continuing education, curriculum development, and competency frameworks for digital curation and preservation have been significant, the resources in the cultural heritage community do not yet include sufficient qualitative or quantitative data to monitor, analyze, or assess the impact of existing programs and practices. This section uses available information to consider the impact of investments in digital curation skills and training in the United States and the international programs and initiatives that informed or influenced U.S. efforts. The recommendations in this section include suggestions for filling gaps in available data that could be instrumental in encouraging and measuring efforts to build human capacity to achieve digital curation outcomes required by the 2013 executive directive, "Increasing Access to the Results of Federally Funded Scientific Research."

It will be necessary to develop new and revamped continuing education and training programs to respond to growth and transformations in data practices. Providing a curriculum that reflects recent developments will present a challenge. There are overlaps in this continuing education assessment with academic education programs that in some cases received funding to develop or offer continuing education programs on digital curation. Academic programs play an important role in developing human capacity. In addition, data curation and other training programs are emerging within research domains; examples are called out in the 2015 National Resource Council report. The digital curation community needs to develop the means to systematically identify and reach out to the providers of continuing education and training programs that are offered by research domains to encourage collaboration on common goals and interests.

The following are the three components of this review of continuing education, competencies, and capacity:

1. *Curation curriculum development and programs.* The commitment to developing training programs and building competencies is evidenced by the cumulative projects that have been funded to date in the United States and elsewhere and that have resulted in some progress in developing continuing education and academic curriculum and programs. Where are we now?
2. *Curation competencies.* There has been an extended and intensive focus on determining requisite skills and competencies for digital curation, resulting in a set of proposed frameworks for defining and developing skills. What benefits do the results offer?

3. *Curation job postings.* The range of job postings and titles in areas relating to digital curation and preservation reflect the evolution of the skills and roles involved. It is a challenge to document, study, measure, and improve the component parts of hiring and retention (e.g., job descriptions that reflect requisite skills, search strategies and outcomes, shifts and staffing, professional development, capability building). What data do we have, and what knowledge do we lack?

Together, these three components address desired digital curation skills, programs intended to develop those skills, and challenges encountered by organizations in describing, hiring, and retaining a growing range of digital curation roles and responsibilities.

Curation Curriculum Development

The review of curation curriculum development began with a list of initiatives funded by the Institute of Museum and Library Services (IMLS) and was extended to identify programs that received funding from other U.S. sources, as well as non-U.S. programs that may have informed or influenced U.S. programs. One limitation is that the list is not exhaustive. In addition, it can be difficult to trace the outcomes of projects when, for example, curriculum results are incorporated into larger academic programs. The list in Appendix 3 includes projects that produced or addressed more than curriculum development. Those additional project outcomes are outside the scope of this review.

From 2004 to 2015, IMLS funded at least 24 projects pertaining to curriculum development, continuing education, training, capability building, internships, and skills development,[22] including academic educational programming (see Appendix 3). This extensive investment in curriculum development for digital curation has resulted in a significant set of resources aimed at developing digital curation skills and competencies. The curation curriculum development review identified three categories of curriculum-related outcomes: certificate programs, workshops, and online resources and tutorials.

In addition to the programs listed in Appendix 3, IMLS funded the launch of the National Digital Stewardship Residency (NDSR) programs for the District of Columbia, Boston, New York City, and the American Archive of Public Broadcasting (AAPB). The first three programs use a proximity approach to build cohorts of residents and hosts in metropolitan areas, and the fourth uses a virtual approach to coordinate projects for a national network. CLIR is leading an IMLS-funded project to assess the outcomes of the NDSR program.[23]

There is also the Coalition to Advance Learning in Archives, Libraries and Museums,[24] which began in 2013. Sponsored by IMLS and organized by OCLC, it works across organizational boundaries

22 Total does not include one grant awarded to Kent State University that included a skills development component, but mostly focused on other issues.

23 Information on the project is available at http://www.clir.org/initiatives-partnerships/ndsr.

24 Information about the Coalition is available at http://www.coalitiontoadvancelearning.org/.

to devise and strengthen sustainable continuing education and professional development programs that transform the library, archives, and museum workforce.

Certificate Programs

With the exception of the continuing education certificate offered by the SAA, the following are U.S.-based certificate programs offering post-master's or graduate level programs. The programs focus on, include, or relate to digital curation.

Table 3.1. U.S.-based certificate programs at the graduate level relating to digital curation

John Hopkins University Zanvyl Krieger School of Arts & Sciences
Graduate Certificate in Digital Curation
http://advanced.jhu.edu/academics/certificate-programs/digital-curation-certificate/

Kent State University College of Communication and Information
Certificate of Advanced Study in Library and Information Science (Post Master's)
http://www2.kent.edu/catalog/2014/ci/certs/c837

San José State University School of Information
Post-Master's Certificate in Library and Information Science (specializations include Digital Archives and Records Management, Digital Curation, and Digital Services and Emerging Technologies)
http://ischool.sjsu.edu/programs/post-masters-certificate

Simmons College School of Library and Information Science
Digital Stewardship Certificate (Graduate Level)
http://www.simmons.edu/slis/programs/postmasters/digital-stewardship/index.php

Society of American Archivists (SAA)
Digital Archives Specialist (DAS) Curriculum and Certificate Program (Continuing Education)
http://www2.archivists.org/prof-education/das

Syracuse University School of Information Studies
Certificate of Advanced Study in Data Science (Graduate Level)
http://ischool.syr.edu/future/cas/datascience.aspx

University of Arizona School of Information Resources and Library Science
Digital Information (DigIn) Graduate Certificate
https://grad.arizona.edu/programs/programinfo/DIGCRTG

University of Illinois, Urbana–Champaign (UIUC) Graduate School of Library and Information Science Data Curation Education Program (Specialization)
Center for Informatics Research in Science and Scholarship (CIRSS)
http://cirss.lis.illinois.edu/Project/project-details.php?id=19

University of Maine
Digital Curation Certificate (Graduate Level)
http://digitalcuration.umaine.edu/

University of North Carolina at Chapel Hill School of Information and Library Science
Certificate in Digital Curation (Graduate Level)
http://sils.unc.edu/programs/certificates/digital_curation

University of Texas at Austin
Certificate of Advanced Study: Curation and Preservation (Graduate Level)
https://www.ischool.utexas.edu/programs/tailored/certificate_of_advanced_study

Wayne State University School of Library and Information Science
Graduation Certificate in Information Management: Data Analytics, Health Informatics and Data Management
http://slis.wayne.edu/certificates/library-information-science.php

Workshops and Institutes

The following U.S. programs represent existing and ongoing workshop programs and institutes that address or pertain to digital curation. For non-U.S. programs, the Digital Curation Centre and UK Data Archive programs are ongoing, but offerings vary and may not be based on a specific curriculum.

Table 3.2. Workshop programs and institutes that address or pertain to digital curation

U.S. programs

Digital Curation Curriculum (DigCCurr) Professional Institute: Curation Practices for the Digital Object Lifecycle, 2009-on
http://www.ils.unc.edu/digccurr/index.html

Digital Preservation Management (DPM)Workshops, 2003-on
http://www.dpworkshop.org/
This program received funding from the National Endowment for the Humanities (NEH).

Digital Preservation Outreach & Education (DPOE), Library of Congress, 2011-on
http://www.digitalpreservation.gov/education/

Society of American Archivists (SAA), Digital Archives Specialists Courses, 2012-on
http://www2.archivists.org/prof-education/das

Programs outside the United States

Digital Curation Centre (DCC) Digital Curation Training for All [UK]
http://www.dcc.ac.uk/training

Digital Preservation Training Programme (DPTP) [UK], 2005-on (builds on the DPM program)
http://www.dptp.org/

UK Data Service—Research [UK]
https://www.ukdataservice.ac.uk/news-and-events/events

Online Resources and Tutorials

The online resources and tutorials are more varied than the previous two categories. The U.S. programs listed continue to be maintained, though at least five of the eight non-U.S. programs have ended. The outcomes in this category may support or supplement programs listed earlier.

Our overview of current curation curriculum programs and resources leads to several observations:

- In the last several years, the number of university-based digital and data certificate programs has increased; a few were launched in the course of this study. A determination of the sustainability of the certificate programs will require more time and monitoring.
- The digital curation and preservation community has produced and has access (that may need to be negotiated) to a significant amount of curriculum material pertaining to digital curation that could be adapted, extended, or built upon to expand and scale up current educational and training offerings to address increasing needs.

Table 3.3. Online resources and tutorials related to digital curation

U.S. programs

Digital Preservation Management (DPM) Online Tutorial
http://dpworkshop.org/

Digital Preservation Education, North Carolina Department of Cultural Resources
http://digitalpreservation.ncdcr.gov/index.html

National Digital Stewardship Alliance (NDSA), Digital Preservation in a Box
https://wiki.diglib.org/NDSA:Digital_Preservation_in_a_Box

Personal Digital Archiving, Library of Congress
http://www.digitalpreservation.gov/about/presentation.html

Programs outside the United States

Digital Curation Centre (DCC) Digital Curation 101 Training Materials [UK]
http://www.dcc.ac.uk/training/train-the-trainer/dc-101-training-materials

Digital Curator Vocational Education Europe Project (DigCurV) [EU] (project ended)
http://www.digcur-education.org/

Digital Preservation Coalition (DPC) [UK]
Handbook: http://www.dpconline.org/advice/preservationhandbook - version 2
Technology Watch Reports: http://www.dpconline.org/advice/technology-watch-reports

InterPARES Educational Modules, University of British Columbia (UBC) [Canada]
http://www.interpares.org/ip3/display_file.cfm?doc=Education-Modules_Digital-Records-Pathways.zip

MANTRA Research Data Management Training [UK] (project ended)
http://datalib.edina.ac.uk/mantra/libtraining.html

RDMRose – Research Data Management for information professionals [UK] (project ended)
http://rdmrose.group.shef.ac.uk/?page_id=10

SCAlable Preservation Environments (SCAPE) [EU] (project ended)
http://www.scape-project.eu/training

Timeless Business (Timbus)[EU] (project ended)
https://www.sba-research.org/timbus/

- Online resources are plentiful, but are at risk. The results from a number of projects that have ended are still available, but the content is no longer supported.
- Content that includes organizational and technological examples is desirable and fills a demonstrated need. However, such content requires resources if it is to be updated and sustained.

Recommendations: Curation Curriculum Development

These recommendations for curation curriculum development are intended to leverage and build on the base of available curricula, instructors, community interest, and lessons learned.

- Develop community-based infrastructure to ensure that curriculum materials and related resources are broadly accessible to instructors to maximize the reach of the curriculum and the impact of the cost of development.

- Support the development of advanced courses that will build on introductory and foundational courses, and will address lifelong learning objectives to keep up with technology, trends, and needs.
- Develop and provide sustained funding for train-the-trainer programs to permit an increase in the scope and scale of programs and to provide incentives that will encourage train-the-trainer programs to use and re-use existing curricula.
- Identify incentives to encourage collaboration between existing and emerging educational and training programs for digital curation.
- Provide funding and support the development of cost models to ensure the sustainability and expansion of existing curation curriculum programs.
- Conduct periodic surveys of program attendees to assess impacts and share data for broad review and use by instructors, curriculum developers, and funders.
- Adopt or adapt success measures for educational programs to determine impact, adjust programs, and use metrics to extend and improve programs.

As noted earlier, several federal public access plans referenced planned or existing education and training, including these examples:

- **NIH:** programs to familiarize researchers and librarians with the National Library of Medicine (NLM) databases and offerings on scientific data analysis and management
- **NOAA and national data centers:** annual environmental data management workshop, free metadata training classes, data management training
- **USDA:** outreach and training plans, developing modules and workshops, and an online training module
- **USGS:** training modules to inform and introduce scientific data management best practice to researchers, data managers, and the public

These examples suggest opportunities for the digital curation community and cultural heritage organizations to address shared educational and training needs by collaborating on joint, shared, or cooperative programs.

Curation Competencies

Within the cultural heritage community, there is a deep and sustained interest in determining requisite or desirable digital curation competencies as evidenced by the number of funded and local projects in the United States and beyond, as well as by the many conference sessions and discussions that have focused on the definition of competency building, the programs designed to develop those competencies, and the curricula to develop those skills. And yet, "there is no single occupational category for digital curators and no precise

mapping between the knowledge and skills needed for digital curation and existing professions, careers, or job titles" (NRC 2015, 1).

This review identified four models that have been presented in the digital curation community and at digital curation and preservation meetings and conferences, are openly available for study and use, and were produced within the last five years.

1. **Digital Curation Curriculum (DigCCurr) Matrix.** An early effort developed at the University of North Carolina (UNC), Chapel Hill, to specify and address curation competencies—possibly the earliest in the United States or elsewhere—the matrix has been implemented as a frame for the School of Information and Library Science program at UNC, Chapel Hill. It has informed other programs and developments, including those noted in the NRC report, *Preparing the Workforce for Digital Curation*.
2. **Digital Curator Vocational (DigCurV) Curriculum Framework.** Developed by a European Commission-funded project, this framework has influenced work in the United States and elsewhere because it is well documented, openly available as a Web resource, and is easy to navigate and cite.
3. **Staffing for Effective Digital Preservation.** This report from the National Digital Stewardship Alliance (NDSA) is based on the results of a 2013 skills survey instrument that aligns curation competencies with current practices as reported by organizations responding to the survey.
4. **Preparing the Workforce for Digital Curation.** This high-profile and significant NRC report released in 2015 includes a section identifying distinct and essential curation knowledge and skill areas that are informing discussions in the United States and beyond pertaining to digital curation, data curation, curriculum development, and funding priorities.

The approach used to produce the findings for this section has the following limitations:

- It does not include other models and approaches that are less well-known and not easily available.
- The scope is mostly on U.S. efforts because of the report's intended audience, although we did consider or include models developed in the United Kingdom and Europe.
- The review reflects only results available in English.
- The four models were developed within the digital curation and preservation community and do not include examples from the broad array of disciplines and domains that are engaged in studying and addressing similar issues.

The recommendations reflect these limitations.

DigCCurr Matrix

Source: *Matrix of Digital Curation Knowledge and Competencies (Overview)*, Digital Curation Curriculum (DigCCurr) Project, Christopher (Cal) Lee, version 13, June 17, 2009; available at http://www.ils.unc.edu/digccurr/digccurr-matrix.html [IMLS Grant # RE-05-06-0044]

Outcome: Curriculum framework

Scope: The DigCCurr Matrix helps to identify and manage course material for digital curation curriculum.

Characteristics:

Number of categories	7
Number of skills	162
Specificity	Extensive description, sections vary in detail

Strengths: The DigCCurr Matrix is comprehensive and manages complexity by defining competencies using six dimensions; the matrix itself is a great resource for understanding digital curation.

Limitations: Some of the more detailed components (e.g., Prerequisite Knowledge Categories and Elements) will need to be updated as technology and practice evolve. (The current online draft was completed in 2009.) The comprehensiveness of the matrix also makes it fairly complex to navigate and use.

DigCurV

Source: *DigCurV Curriculum Framework*, Digital Curator Vocational Education Europe Project, funded by the European Commission's Leonardo da Vinci program, 2013; available at http://www.digcurv.gla.ac.uk/skills.html

Outcome: Curriculum framework

Scope: DigCurV was designed to identify, evaluate, and plan training to meet the skill requirements of staff engaged in digital curation, both now and in the future. Rooted in the experience of curators, the model identifies three lenses: executive, manager, and practitioner. The lenses echo the layers of the Digital Preservation Outreach and Education (DPOE) Pyramid that addresses the continuing education needed at different levels: basic (is aware of), intermediate (understands), and advanced (is able to).

Characteristics:

Number of categories	14
Number of skills	110
Specificity	Detailed, consistent throughout

Strengths: The DigCurV model is adaptable to contexts (lenses and levels) to support a broad range of educational offerings and delivery methods. The examples provided to explain the competencies contribute to its usability. The consistency of the model structure and the identifiers assigned to each competency make it easy to navigate and reference.

Limitations: The examples require updates; the project has ended, and the framework is not current and will not be maintained.

Model Profiles

Following are profiles of the competency model produced by each project and a data set that includes the categories, competencies, and skills defined by each project. Appendix 4 identifies the competency categories and skills defined for each of the four models.

A comparison of the four models reveals commonalities as well as some informative differences:

- There is enough commonality or complementarity across the categories in the models to enable the consolidation of efforts in developing and extending requisite curriculum.

NDSA Staffing Report

Source: *Staffing for Effective Digital Preservation*, NDSA Standards and Practices Working Group, National Digital Stewardship Alliance, 2013; available at http://www.digitalpreservation.gov/ndsa/documents/NDSA-Staffing-Survey-Report-Final122013.pdf

Outcome: Digital preservation competency framework

Scope: The National Digital Stewardship Alliance digital preservation skills survey was conducted to better understand the staffing and organization of institutions that are responsible for the long-term management and preservation of digital content. The competencies were addressed in one question of the survey.

Characteristics:

Number of categories	1
Number of skills	20
Specificity	Competency category labels, no descriptions

Strengths: The focus on digital preservation makes the NDSA Staffing Report a useful supplement that can plug into a cumulative model or overview. The survey includes a useful and broader set of digital preservation competencies than other models.

Limitations: The survey was not developed to be used as a curriculum model, so it does not lend itself to comparison and use as the other models do. Ideally, the results would inform a longitudinal approach to monitoring relevant institutional behaviors as they evolve.

NRC Report

Source: *Preparing the Workforce for Digital Curation*, Committee on Future Career Opportunities and Educational Requirements for Digital Curation; Board on Research Data and Information; Policy and Global Affairs; National Research Council, 2015; available at http://www.nap.edu/catalog/18590/preparing-the-workforce-for-digital-curation

Outcome: Curation competency framework

Scope: The knowledge and skill areas included are essential to the education of professionals in the field of digital curation.

Characteristics:

Number of categories	11
Number of skills	76
Specificity	Brief descriptions in one section of the report

Strengths: The framework takes a broad approach that reflects feedback study authors received from numerous experts during the course of their investigation. The model takes a step toward consolidation that the others do not, although only the DigCCurr model of the models reviewed here is referenced.

Limitations: The NRC model includes competencies that are desirable (but not necessarily specific to digital curation and preservation) to a greater degree than the other models.

- The differences in scale, scope, specificity, and approach among the models contribute to the difficulty of comparing and applying the models.
- The availability of the models for comparison and use provides a valuable resource for defining the next steps in the extension of current curation training and education to meet the growing needs identified in the community reports and documents cited earlier in this report.

- The range of perspectives and methods represented in the models cumulatively address the problem of curriculum for competency building in an extremely useful way.

Although the cumulative categories in the competency-based models reviewed provide a comprehensive scope and definition of competencies, a gap will emerge and grow (1) if supplementary work does not address and build on current models and (2) if technology-specific and practice-specific curriculum content is not updated.

Connecting the Models

As a step toward integrating and enabling broader use of existing curation competency models, table 3.4 presents a four-level competency framework that maps to the four models. The first three aggregate categories are applicable to any type of digital content management, although each can and would be specialized to address specific data issues and concepts. The fourth focuses more on data-specific skills.

The fourth aggregate category requires the most resources to update and maintain relevant curriculum and courses as technologies and techniques evolve. This category is the most directly linked to the online resources category of the curation curriculum development review.

Table 3.4 Four-level competency framework, mapped to the four models

Aggregate Categories	Mapping Aggregate Categories to Four Models
Contexts: Addresses the need to develop a deep familiarity with cultural, disciplinary, organizational contexts to enable long-term curation.	*DigCCurr:* Mandates, values, principles; transition point of digital objects; prerequisite knowledge *DigCurV:* Knowledge and intellectual abilities/subject knowledge [KIA1] *NRC:* General background and abilities; values and principles *NDSA:* N/A
Management and Administration: Recognizes the need to ensure that curators are able managers. Topics include advocacy and outreach, policy development, standards implementation, values and principles, ethics, strategies, evaluation, audit, collaboration, contracting, technical infrastructure investment, education and training, staffing, needs assessment, cost models, project management.	*DigCCurr:* Functions and skills/metalevel functions *DigCurV:* Knowledge and intellectual abilities/selection and appraisal [KIA2]; evaluation studies [KIA3]; management and quality assurance [MQA]; personal qualities [PQ]; professional conduct [PC] *NRC:* Management and administration; policy and planning *NDSA:* N/A
Functions: Defines functional areas of curation that are essential for long-term curation, including appraisal, acquisition, identification, research, development, forensics, accessibility, monitoring, destruction, tools and workflows, preservation planning, digitization, storage management, records management, rights management.	*DigCCurr:* Functions and skills *DigCurV:* Knowledge and intellectual abilities/information skills [KIA4] *NRC:* Preservation and archiving; data collections and management; presentation and visualization; services and support; technologies, tools, and infrastructure *NDSA:* Digital preservation skills
Data Skills: Identifies specific skills across functions at which curators must become adept, including analytics, practices, formats, metadata, databases, vocabularies, provenance, linkages, citation, identifiers.	*DigCCurr:* Type of resource *DigCurV:* Knowledge and intellectual abilities/data skills [KIA5] *NRC:* Data analytics; data practices *NDSA:* N/A

Recommendations: Curation Competencies

The curation competency review produced these recommendations:

- Urge future projects on curation competencies to identify and consider the relevancy of existing models—not only these four, but also relevant models from research domains. With existing competency models in place as a foundation for understanding and building digital curation competencies, future work on competencies should be expected, at a minimum, to use the available foundation as a starting point.
- Encourage and fund collaborative competency-building projects that are interdisciplinary. Efforts by individual researchers and practitioners, as well as by the staffs of data creation and curation programs, would benefit from collaborative projects and initiatives that include digital curation and data science researchers to leverage, extend, and refine existing competency-based models and curricula.

To enable practitioners to understand and carry out digital curation, it was necessary and important for the digital curation community to identify and define requisite competencies. Those efforts have resulted in competency models from multiple perspectives and for diverse purposes that cumulatively and effectively address the question of skills needed for digital curation.

Curation Job Descriptions

The *Preparing the Workforce for Digital Curation* report delineates indicators of the growing need for curators to respond to the explosion of data being produced. It includes an analysis of digital curation and related jobs using indeed.com that projects an increase in data-related jobs and an insufficient number of applicants prepared to fill them.

As with any hiring process, hiring practices for digital curation jobs reflect the regulations, preferences, and culture of the institutions seeking to fill positions. As a result, differences in the format and structure of job descriptions, the means used to advertise, and the application submission mechanism can be substantial.

The digital curation community may be shifting toward the use of services like indeed.com to post positions and there are relevant vacancy listings such as the one provided by the Digital Preservation Coalition (DPC) in the United Kingdom, but there is no single, known source of postings for digital curation jobs. Therefore, this review of job descriptions covers job postings to the American Library Association (ALA) digital preservation (digipres) listserv from 2013 to 2015. The listserv is a common place to post digital curation and preservation jobs; it is also the largest known and available source of potentially relevant job descriptions. This approach has a number of limitations, including the heavy overrepresentation of academic libraries in the job postings on the listserv and the inclusion on the listserv of unrelated job postings.

Analysis of Postings to the ALA Digital Preservation Listserv

Appendix 5 lists 110 job postings from the ALA digipres listserv from 2013 to 2015. The columns in the table identify month and year of posting, job title, type of organization posting the job, availability of a detailed job description, and an indicator of whether the job is in or outside the United States.

This limited analysis revealed a number of challenges in using job postings to increase our understanding of the workforce:

- Comparing and analyzing job descriptions is complicated by substantial variations in the level of detail provided about postings, in the title and other terminology used to describe positions, and in the ways in which full job descriptions are provided in the postings.
- If the e-mail job postings do not include an attachment to the e-mail or a sufficient description of the job, links to external systems in the e-mail postings typically (and not surprisingly) become inactive when the associated job search is closed.
- Information in the job descriptions about the reporting lines for the position, level of position, and scope of responsibilities may be missing or incomplete enough to make comparisons between postings difficult or impossible.
- There is no way to determine from job postings alone if the job search was successful or, if successful, how long the employee remained in that position.

If information on job postings is to become a useful and ongoing measure of change in the digital curation workforce, it would be helpful to have a common community-based approach for collecting, comparing, and analyzing job postings.

Recommendations: Curation Job Descriptions

The following recommendations build on the results and findings of this review of job postings and related workforce issues:

- Encourage community-based initiatives to support career planning and mentoring programs for researchers and practitioners in digital curation.
- Complete an in-depth data curation and digital curation job study, possibly using a study of digital archivist job postings by Jane Zhang, an archival educator at Catholic University.
- Develop a means for a systematic study to better understand and monitor the growth and potential capacity of the digital curation workforce; the results could inform the definition of common modules to build well-formed job descriptions for digital curation and data curation positions.
- Develop a framework to encourage digital curation capacity building that defines levels of skills, possibly using the DigCurV model reviewed in the competencies component

Analysis of Postings to the ALA Digital Preservation Listserv

Within the job titles:

- 4 positions include data in the title: Data Librarian, 2 Data Management Services Librarians, and Data Services Librarian/Specialist
- 8 Digital Archivists, 2 AV Archivist positions, and 8 other positions have archives or archivist in the title
- 39 positions have the word digital in the title, most of which have only 1 position per title except:
  - » Digital Archivist....8
  - » Digital Initiatives Librarian....4
  - » Digital Projects Librarian....4
  - » Digital Preservation Manager....3
  - » Digital Preservation Analyst....2
  - » Digital Preservation Librarian....2
- 13 positions have preservation in the title, of which 3 specify digital in the title
- 9 positions relate to audiovisual content
- 9 positions have metadata in the title
- 8 information technology (IT)–related positions: Project Developer DMP Tool, Systems Librarian, Applications Analyst, Senior Software Developer, Software Developer, Repository Technical Project Manager, Production Systems Architect and Administrator, and IT Analyst III [digital library program]
- 8 are higher level management positions
- 7 are higher level management positions that have not already been mentioned in another category
- 7 are digitization and related library positions
- 4 job titles pertain to special collections, of which 1 also mentions archives
- 4 positions specifically address only physical collections
- 3 positions may suggest efforts to look to the future: Strategic Program Specialist, Head of Research and Development Department, and Web Archiving and Emerging Formats Librarian

In addition to Academic libraries, other types of institutions that posted positions included:

- » Art museum....2
- » City archives....2
- » Curation services provider....3
- » DP membership organization (non-U.S.)....1
- » Historical society....1
- » Library professional association....2
- » Library service provider....17
- » National U.S. collection....8
- » National library or archives (non-U.S.)....4
- » Preservation or media service provider....3
- » Public library....2
- » University....3

As for the means of posting the job descriptions: 5 were included as attachments to the e-mail messages, 6 provided a link to the job description (all 6 are 2015 postings that were still active when checked), and 16 provided the job description through a link that is no longer active. Of the remaining postings that included a description in the message, 15 were too brief to be informative.

Of the 110 job postings, 13 were non-U.S. postings; the remaining postings were by U.S. institutions.

of this section, and use the framework to develop and assess the impact of subsequent continuing education and training programs.
- Devise measures to monitor digital curation hiring and retention, including trends in titles and responsibilities, expectations for new or advanced skills, and indicators of emerging or fading specializations.

These recommendations could inform the charge for a community-based working group to explore and monitor the digital curation workforce as it grows and evolves.

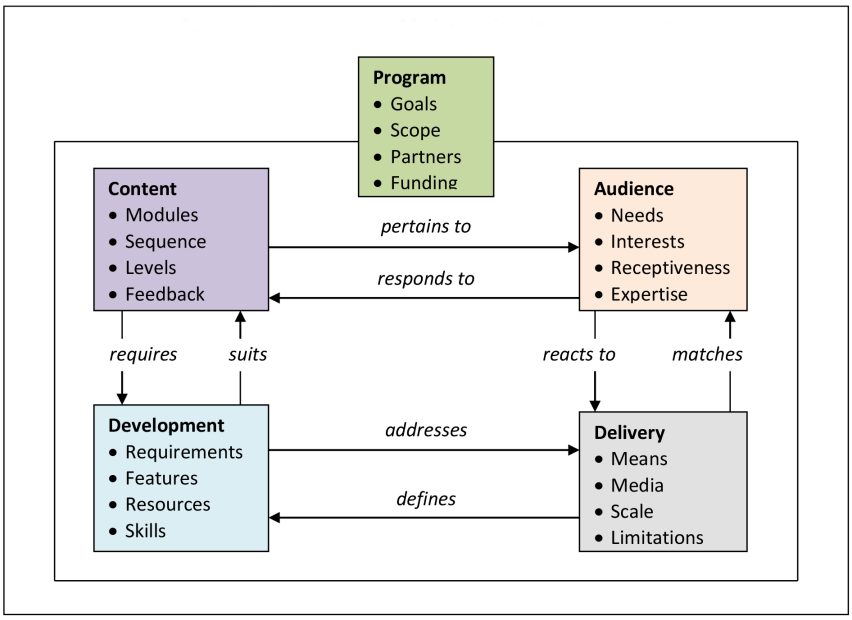
Conclusion

Capacity building for digital curation requires programs and other initiatives that will support and enable the successful implementation of the federal public access plans. The subsections on curriculum development, competencies, and job postings as one element of career tracking and development each contain recommendations specific to those topics.

Continuing education is needed to address skill development from basic to advanced levels of expertise with support for continual growth. The curriculum development model presented here provides a conceptual framework to inform the development of an organizational and technological infrastructure that will ensure sustainable support for continuing education programs. The curriculum development model suggests the following components for a curriculum-based program and the relationships among those components:

- **Program:** The goals and scope of the program should be clearly defined and current; the partners and funding should be sufficient to develop and maintain the program.
- **Content:** The curriculum content should be modular for easier expansion and updates; options or requirements for sequencing content should be specific to support levels and allow flexible delivery; and feedback loops should be in place to drive updates.
- **Audience:** Herding participants of all levels and varying interests into courses is not a successful approach for continuing education or training; it is essential to match audience needs and intended outcomes.
- **Development:** Curriculum development is costly, so establishing requirements-based projects to develop, extend, and update curriculum works well.
- **Delivery:** Modular development of curriculum content with well documented materials enables the curriculum to be delivered and adapted as needed for in-person and online, as well as synchronous and asynchronous, options.

Fig. 3.2. Curriculum development model for continuing education



The curriculum development model in figure 3.2 informs this proposed package of next steps to extend and expand continuing education and training capacity:[25]

- Expand the role of continuing education and training to enable compliance with federal plans and enhance the role of education in outreach.
- Work collaboratively across communities that are interested in and responsible for the life cycle management of research data to develop educational programming that uses existing curriculum materials as a foundation for building skills for new and emerging curation roles.
- Determine a foundational curriculum that builds on commonalities across curation curriculum development programs and competency models.
- Develop advanced courses, timely advanced modules, and topical content that, as the base of curators grows, can supplement foundational curriculum and provide lifelong learning that responds to technological change, emerging trends, and evolving needs.
- Adopt and adapt success measures for educational programs to determine impact and make it possible to adjust programs as needed; use metrics to extend and improve programs.
- Convene a community initiative to define a framework of curatorial skills that map to existing and emerging curatorial roles and provide common modules for job descriptions.

25 Nancy Y. McGovern developed this previously unpublished curriculum development model, which has informed the development of curriculum for the Digital Preservation Management workshop program.

- Use the resulting curatorial skills framework to define levels of development and map to current and future educational and training offerings. Invest in train-the-trainer programs for data curation and digital curation to build the base of instructors outside of academic library contexts and expand the reach of programs.
- Continue and expand support for residencies, fellowships, and postdoctoral programs, including the National Digital Stewardship Residency (NDSR) program, to provide a bridge for graduates from academic programs to curatorial positions in a range of repositories that receive tangible benefits from hosting residents.
- Sustain the development of the Coalition to Advance Learning in Archives, Libraries and Museums, and emphasize the potential for building on curatorial expertise within its membership.
- Study past and current curatorial practices and lessons learned to inform curriculum and provide resources for courses to promulgate good curatorial practices.

A community-supported base that promotes sustainable continuing education and training development would facilitate these activities. The Coalition to Advance Learning in Archives, Libraries and Museums and other cross-community initiatives could help address the growing need to educate and train a cohort of curators to meet growing research data management needs.

Looking Ahead

The cultural heritage community captures and preserves data that tell the story of our world. Providing public access to those data presents unique challenges and opportunities for cultural heritage institutions and professionals. This report identifies lessons that can be extrapolated through three different approaches: (1) the plans of agencies subject to the federal mandate for open data, even though that focus is on federally funded research primarily in the sciences and social sciences; (2) projects supported by the Institute of Museum and Library Services (IMLS) that have developed model services and tools supporting data management; and (3) efforts to build capacity through continuing education programs and comprehensive workforce development.

In addition to the findings, implications, and recommendations identified in our report, we note the following actions as deserving priority:

- Track federal agencies' responses to the federal mandate to understand how they have implemented activities outlined in their public access plans and the outcomes from those activities, including the effects on data sharing and sustainability. Examine how this can inform the cultural heritage community's approaches to digital curation.
- Monitor the development of the Research Data Commons, and develop a strategy for involving the cultural heritage community.
- Explore ways to identify and possibly quantify how open access to cultural heritage data is creating new knowledge and is facilitating cross-disciplinary research.
- Devise measures and means to assess the effectiveness of continuing education as programs are scaled up across domains to address the open data imperative.

- Develop a data model to monitor and evaluate the growth and evolution of the digital curation workforce.

The cumulative results, findings, and recommendations of this report provide a holistic view of data stewardship and the infrastructure required to support data-driven research and innovation. As digital curation roles and responsibilities emerge and change, new opportunities to engage and collaborate with data will open fresh frontiers across the range of research domains. These frontiers will yield increasingly interdisciplinary approaches to digital curation that will encourage and enable innovation of a kind we now only imagine.

References

Works Cited

Berelson, Bernard. 1952. *Content Analysis in Communication Research*. New York: Free Press.

Burwell, Sylvia M., et al. 2013. Open Data Policy—Managing Information as an Asset. Memorandum for the Heads of Executive Departments and Agencies. Executive Office of the President, Office of Management and Budget, May 9. Available at https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.

Choudhury, G. Sayeed. 2013. Written Testimony of G. Sayeed Choudhury given before the Testimony to the Committee on Science, Space, and Technology Subcommittee on Research. House of Representatives Hearing on Scientific Integrity and Transparency, March 5. Available at http://dataconservancy.org/wp-content/uploads/2013/05/WrittenTestimony_Sayeed_CongressHearing.pdf.

Halpin, Patrick N., et al. 2006. OBIS-SEAMAP: Developing a Biogeographic Research Data Commons for the Ecological Studies of Marine Mammals, Seabirds, and Sea Turtles. *Marine Ecology Progress* Series 316: 239–246.

Hedstrom, Margaret, and John Leslie King. 2006. Epistemic Infrastructure in the Rise of the Knowledge Economy. In *Advancing Knowledge and the Knowledge Economy*, edited by Brian Kahin and Dominique Foray, 113–134. Cambridge, MA: MIT Press.

Holdren, John. 2013. Increasing Access to the Results of Federally Funded Scientific Research. Office of Science and Technology Policy, Memorandum for the Head of Executive Departments and Agencies, February 22.

Jahnke, Lori, and Andrew Asher. 2012. The Problem of Data: Data Management and Curation Practices Among University Researchers. Pp. 3-31 in *The Problem of Data.* Washington, DC: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/pub154/pub154.pdf.

Keralis, Spencer. 2012. Data Curation: A Snapshot. Pp. 32-40 in *The Problem of Data*. Washington, DC: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/pub154/pub154.pdf.

Lee, Christopher A. 2012. *States of Sustainability: A Review of State Projects Funded by the National Digital Information Infrastructure and Preservation Program (NDIIPP).* Report to the U.S. Library of Congress, March. Available at http://www.ils.unc.edu/callee/ndiipp-states-report.pdf.

Lee, Christopher A. 2011a. A Framework for Contextual Information in Digital Collections. *Journal of Documentation* 67(1): 95–143.

Lee, Christopher A., ed. 2011b. I, *Digital: Personal Collections in the Digital Era*. Chicago, IL: Society of American Archivists.

Luce, Richard E. 2008. A New Value Equation Challenge: The Emergence of eResearch and Roles for Research Libraries. In *No Brief Candle: Reconceiving Research Libraries for the 21st Century*. Washington, DC: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/pub142/luce.html.

Lynch, Clifford. 2013. The Next Generation of Challenges in the Curation of Scholarly Data. In *Research Data Management: Practical Strategies for Information Professionals*, edited by Joyce M. Ray. West Lafayette, IN: Purdue University Press.

National Science Foundation. 2007. *Cyberinfrastructure Vision for 21st Century Discovery*. Available at http://www.nsf.gov/pubs/2007/nsf0728/nsf0728_1.pdf.

NDSA (National Digital Stewardship Alliance). 2013. *Staffing for Effective Digital Preservation*. Available at http://www.digitalpreservation.gov/ndsa/documents/NDSA-Staffing-Survey-Report-Final122013.pdf.

NRC (National Research Council). 2015. *Preparing the Workforce for Digital Curation.* Committee on Future Career Opportunities and Educational Requirements for Digital Curation. Washington, DC: National Academies Press. Available at http://sites.nationalacademies.org/pga/brdi/PGA_069853.

NRC (National Research Council). 1997. *Bits of Power: Issues in Global Access to Scientific Data*. Committee on Issues in the Transborder Flow of Scientific Data. Washington, DC: National Academies Press. Available at http://www.nap.edu/read/5504/chapter/1.

OMB (Office of Management and Budget). 2000. Memorandum for Heads of Executive Departments and Agencies: Management of Federal Information Resources. OMB Circular A-130, Revised. November 28, 2000. Available at https://www.whitehouse.gov/omb/circulars_a130_a130trans4/.

OMB (Office of Management and Budget). Rev. 1993; 1999. Circular A-110, to the Heads of Executive Departments and Establishments: Uniform Administrative Requirements for Grants and Agreements with Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations. Revised 11/19/93 as Further Amended 9/30/1999. Available at https://www.whitehouse.gov/omb/circulars_a110/.

OSTP (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. Available at https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf.

Ray, Joyce. 2009. Sharks, Digital Curation, and the Education of Information Professionals. *Museum Management and Curatorship* 24(4). Available at doi: 10.1080/09647770903314720.

Reichman, J. H., and Paul F. Uhlir. 2003. A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment. *Law and Contemporary Problems* 66(1/2): 315–462.

Sheehan, Jerry. 2015. Policy Implications: Policy and Legal Issues Associated with the Commons. Presentation at "The Future of the Commons: Data, Software, and Beyond," Washington, DC, November 18, 2015.

UNESCO (United Nations Educational, Scientific and Cultural Organization). 2003. Charter on the Preservation of Digital Heritage, October 15.

The White House Office of the Press Secretary. 2013. Executive Order: Making Open and Machine Readable the New Default for Government Information, May 9, 2013.

Whitmire, Amanda, Kristin Briney, Amy Nurnberger, Margaret Henderson, Thea Atwood, Margaret Janz, Wendy Kozlowski, Sherry Lake, Micah Vandegrift, Lisa Zilinski, Lisa. 2015. A Table Summarizing the Federal Public Access Policies Resulting from the U.S. Office of Science and Technology Policy Memorandum of February 2013. Available at https://figshare.com/articles/A_table_summarizing_the_Federal_public_access_policies_resulting_from_the_US_Office_of_Science_and_Technology_Policy_memorandum_of_February_2013/1372041.

Websites Cited

Data Curation Profiles Toolkit: http://datacurationprofiles.org/

Data.gov: https://www.data.gov/

National Digital Stewardship Residency Program

—Washington, DC: http://www.digitalpreservation.gov/ndsr/residency.html

—Boston: http://projects.iq.harvard.edu/ndsr_boston

—New York: http://ndsr.nycdigital.org/

—American Archives of Public Broadcasting: http://ndsr.americanarchive.org/.

Project Open Data: https://project-open-data.cio.gov/

U.S. Open Data: https://usopendata.org/

Related Reading

CLIR (Council on Library and Information Resources). 2013. *Research Data Management: Principles, Practices, and Prospects*. Washington, DC: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/NIST public aub160/pub160.pdf.

DataOne Education Modules. Available at https://www.dataone.org/education-modules

Digital Curation Centre (DCC) Training Materials. Available at http://www.dcc.ac.uk/training/train-trainer/disciplinary-rdm-training/curate-and-preserve

DLF E-Research Network. Available at https://www.diglib.org/groups/e-research-network/

Dooley, Jackie. 2015. *The Archival Advantage: Integrating Archival Expertise into Management of Born-digital Library Materials.* Dublin, OH: OCLC Research. Available at http://www.oclc.org/research/publications/2015/oclcresearch-archival-advantage-2015.html

Lyle, Jared, Jacob Carlson, and Ron Nakao. N.d. *An Applied Approach to Data Curation Training at the Inter-university Consortium for Political and Social Research (ICPSR)*. Available at http://ceur-ws.org/Vol-1016/paper14.pdf

Palmer, Carole L., Nicholas Weber, Trevor Muñoz, and Allen H. Renear. 2013. Foundations of Data Curation: The Pedagogy and Practice of "Purposeful Work" with Research Data. *Archive Journal 3* (Summer). Available at http://www.archivejournal.net/issue/3/archives-remixed/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data (2013)

Pearce-Moses, Richard, and Susan E. Davis, eds. 2008. *New Skills for a Digital Era.* Proceedings of a colloquium sponsored by National Archives and Records Administration; Society of American Archivists; and Arizona State Library, Archives and Public Records, May 31—June 2, 2006, Washington, DC. Chicago, IL: Society of American Archivists. Available at http://www.archivists.org/publications/proceedings/NewSkillsForADigitalEra.pdf.

Williford, Christa, and Charles Henry. 2012. *One Culture. Computationally Intensive Research in the Humanities and Social Sciences.* Washington, DC: Council on Library and Information Resources. Available at http://www.clir.org/pubs/reports/pub151.

APPENDIX 1

Analysis of Public Access Plans: Research Design and Methods

During the nearly two years when the plans for public access to data were unavailable, we worked on our approach to analysis by conducting a preliminary content analysis focused on the instructions from the Office of Science and Technology Policy (OSTP), the few rough drafts available from federal agencies, and a review of documents from governments in other countries that are addressing the same data management issues. When the U.S. federal plans were gradually released to the public, it became evident that agencies were not consistent in how they addressed the 2013 executive directive. Consequently, we redirected our analysis to examine the information as it was presented rather than applying a framework based on the executive directive. Twenty-one plans were analyzed (see Appendix 2), including the U.S. Agency for International Development (USAID) plan, although it is addressed separately because it was not from one of the agencies required to file a plan nor is it an operating division of one of these named agencies.

Content analysis is a powerful method for studying texts (Berelson 1952). Recognizing that the plans were developed to address a specific directive regarding the need for open data and public access, we used the plans as structured texts in our content analysis. The data were analyzed using the following process:

Task 1. Assignment of a level of analysis. We reviewed the documents to determine the appropriate level of analysis to be used. Because the documents were quite different in their specificity and scope, it became apparent that the analysis must be focused on broad concepts rather than specific characteristics.

Task 2. Creation of categories. Our first framework was developed using existing data policies and plans (beyond those in federal agencies) to identify key pre-defined categories to include in the analysis. These included specifics about cost, access, re-usability, data type, data analysis support, and data description. However, as the plans became available, it was clear that we needed to broaden the analysis significantly to focus instead on the key concepts of definitions of research data, open access to data, location for data storage, restrictions to data access, requirements for researchers' data management plans, and handling the output of research data analysis (e.g., articles). We also looked at how the plans addressed two important concepts in the executive directive: (1) meeting the goal of accelerating scientific discovery and fueling innovation, and (2) making the results of taxpayer-funded research available. We were particularly interested in creating a category for information about collaboration, but this was problematic. Although most agencies mentioned working with other agencies, they did not identify specific agencies or the mechanism they would use to work together.

Task 3. Analysis and results. The plans were assessed for existing similarities and differences, which led to the identification of important themes for understanding the emerging government data management environment.

Limitations and Future Research

Our analysis has several limitations. At the time of the analysis, not all plans were available from the agencies required to submit one. Furthermore, among those plans that were available, some were not yet finalized, and others did not specify if they were final. Although agencies use a common definition of data and address common elements related to data management, they often approach these elements in very different ways—which made some comparisons difficult.

Most important, the data landscape is a very dynamic environment that is difficult to predict.

APPENDIX 2

Links to Federal Department and Agency Public Access Plans Used for This Report

The public access plans used in this analysis were freely available on the Internet and were discovered via Internet searches for these documents at agencies that were required to comply with the February 22, 2013, OSTP policy memo, "Increasing Access to the Results of Federally Funded Scientific Research." Some of the plans in this analysis may still be awaiting OSTP review as of May 2016. The researcher has retained PDF versions of the plans used for analysis.

Department of Agriculture (USDA)
http://www.usda.gov/documents/USDA-Public-Access-Implementation-Plan.pdf

Department of Commerce
http://open.commerce.gov/sites/default/files/Commerce%20Open%20Government20Plan%20Version%203_5%20%289-28-15%29%20Final.pdf

National Institute of Standards and Technology (NIST)
http://www.nist.gov/open/upload/NIST-Plan-for-Public-Access.pdf

National Oceanic and Atmospheric Administration (NOAA)
http://docs.lib.noaa.gov/noaa_documents/NOAA_Research_Council/NOAA_PARR_Plan_v5.04.pdf

Department of Defense (DOD)
http://www.dtic.mil/dtic/pdf/dod_public_access_plan_feb2015.pdf

Department of Energy (DOE)
http://energy.gov/downloads/doe-public-access-plan

Department of Health and Human Services (HHS)

Agency for Healthcare Research and Quality (AHRQ)
http://www.ahrq.gov/funding/policies/publicaccess/index.html

Office of the Assistant Secretary for Preparedness and Response (ASPR)
http://www.phe.gov/Preparedness/planning/science/Pages/AccessPlan.aspx

Centers for Disease Control and Prevention (CDC)
http://www.cdc.gov/od/science/index.htm

Food and Drug Administration (FDA)
http://www.fda.gov/ScienceResearch/AboutScienceResearchatFDA/ucm433459.htm

National Institutes of Health (NIH)
http://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf

Department of Labor (DOL)
http://www.dol.gov/digital-strategy/publicationprocess.htm
http://www.dol.gov/digital-strategy/inventoryschedule.htm

The two pages are not linked so must be accessed separately.

Department of Transportation (DOT)
https://www.transportation.gov/open/plan-chapter3

Department of Veterans Affairs (VA)
http://www.research.va.gov/resources/policies/public_access.cfm

Environmental Protection Agency (EPA)
http://www2.epa.gov/sites/production/files/2015-05/documents/opendatapolicyimplementation-plan_030415_finalb.pdf

Institute of Museum and Library Services (IMLS)
https://www.imls.gov/about-us/open-government/commitment-open-data

National Aeronautics and Space Agency (NASA)
http://science.nasa.gov/media/medialibrary/2014/12/05/NASA_Plan_for_increasing_access_to_results_of_federally_funded_research.pdf

National Science Foundation (NSF)
http://www.nsf.gov/news/special_reports/public_access/

Smithsonian Institution
https://www.si.edu/content/pdf/about/SmithsonianPublicAccessPlan.pdf

U.S. Geological Survey (USGS)
http://www.usgs.gov/usgs-manual/im/IM-OSQI-2015-01.html

Data management guide is available at
http://www.usgs.gov/datamanagement/index.php

Data are made available at
http://data.usgs.gov/datacatalog/#fq=dataType%3A(collection%20OR%20non-collection%20OR%20supercollection)&q=*%3A*

Not required to submit plan:

U.S. Agency for International Development (USAID)
https://www.usaid.gov/sites/default/files/documents/1868/579.pdf

APPENDIX 3

Projects on Digital Curation Curriculum and Skills Development Funded by the Institute of Museum and Library Services, 2004–2015

Text in brackets indicates a descriptive, rather than actual, title.

Project	Category	Award Year	Amount
Digital Library Education Program (DLEP). Indiana University. http://www.lis.illinois.edu/academics/programs/cas-dl	Curriculum, skills development	2004	$939,618
Preserving Access to Our Digital Future: Building an International Digital Curation Curriculum (DigCCurr I). Key staff: Helen Tibbo, Cal Lee. University of North Carolina at Chapel Hill. http://www.ils.unc.edu/digccurr/digccurr_I_final_report_031810.pdf	Curriculum	2006	$562,041
Extending Data Curation to the Humanities. University of Illinois at Urbana–Champaign.	Curriculum	2008	$892,028
[Digital Curation/Digital Preservation Internships]. Key staff: Elizabeth Yakel, Paul Conway. University of Michigan.	Training, skills development	2008	$631,816
Curriculum, Cooperation, Convergence, Capacity— Four C's for the Development of Cultural Heritage Institutions. Simmons College, Boston, MA.	Curriculum, skills development, training	2008	$455,639
DigCurr II. Key staff: Helen Tibbo, Cal Lee. University of North Carolina at Chapel Hill.	Curriculum	2008	$878,634
[Digital curation curriculum on the management and preservation of science-related information plus scholarships to students with a background in the sciences]. Syracuse University, Syracuse, NY.	Curriculum	2009	$706,200
[ESOPI]. University of North Carolina at Chapel Hill.	Curriculum	2009	$803,258
Closing the Digital Curation Gap [funded by National Leadership Grants for Libraries]. University of North Carolina at Chapel Hill.	Continuing education, capabilities	2009	$249,623
Preparing Information Professionals as Digital Managers. Pratt Institute, New York, NY.	Curriculum, skills development	2010	$971,407
Understanding Curation Through the Use of Data Curation Profiles. Purdue University, West Lafayette, IN.	Continuing education	2010	$187,242
Data Curation Education in Research Centers (DCERC). Key staff: Carole Palmer.	Curriculum, training	2010	$988,543
[Scholarships to develop faculty to train e-science or data curation librarians]. Syracuse University, Syracuse, NY.	Curriculum, capabilities	2011	$741,936
Information: Curate, Archive, Manage, Preserve (iCAMP). Key staff: William Moen. University of North Texas, Denton, TX. http://icamp.unt.edu/icamp/content/icamp-project	Curriculum	2011	$624,663
SciData project. University of Tennessee, Knoxville, TN.	Curriculum	2011	$546,472
Graduate specialization in Sociotechnical Data Analytics (SODA). Key staff: Catherine Blake. University of Illinois at Urbana–Champaign.	Curriculum	2012	$498,777
Closing the Gap: Identifying Needs in Continuing Education for Managing Cultural Heritage Data. Key staff: Charles Henry. Council on Library and Information Resources (CLIR), Washington, DC.	Continuing education	2013	$164,243
Testing the National Digital Stewardship Residency (NDSR) Model in the Boston Area. Key staff: Andrea Goethals, Nancy McGovern. Harvard Library, Cambridge, MA.	Continuing education, skills development	2013	$498,385
National Digital Stewardship Residency in New York. New York Metropolitan Reference and Resource Library Agency, New York, NY.	Continuing education, skills development	2013	$498,135
Curate Cloud: Building Digital Curation Excellence through Professional Education, Cloud Computing and Community Outreach. University of Maryland, College Park, MD.	Continuing education, skills development	2013	$299,999

Project	Category	Award Year	Amount
Curating Research Assets and Data Using Lifecycle Education (CRADLE): Data Management Tools for Librarians, Archivists, & Content Creators. University of North Carolina at Chapel Hill. http://cradle.web.unc.edu/	Continuing education	2013	$499,002
[Capacity-building project to educate six master's students in the area of scientific data curation]. University of Tennessee, Knoxville, TN.	Curriculum, skills development	2014	$438,991
American Archive of Public Broadcasting (AAPB) National Digital Stewardship Residency (NDSR) Project. Key staff: Karen Cariani. WGBH, Boston, MA.	Continuing education, skills development	2015	$450,126
Learning, Evaluation and Analysis Project II (LEAP-II). Key staff: Yvonne Chandler. University of North Texas, Denton, TX.	Curriculum, skills development	2015	$499,991

Examples of Non-U.S. Projects and International Collaborations

Project	Institution	Year(s)	Funders
Closing the Digital Curation Gap (CDCG) http://www.dcc.ac.uk/projects/closing-digital-curation-gap	Digital Curation Centre	2009–2011	IMLS; JISC
Research Data Management Skills Support Initiative (DaMSSI-ABC) http://www.dcc.ac.uk/training/damssi	Digital Curation Centre	2010–2013	JISC; Research Information Network (RIN)
DigCurV http://www.dcc.ac.uk/projects/digcurv	Fondazione Rinascimento Digitale; Goettingen State and University Library; Humanities Advanced Technology Institute, IMLS, & more	2011–2013	European Commission
Timeline Business Processes and Services (TIMBUS)	Alliance Permanent Access to the Records of Science in European Network (APARSEN)	2013	European Commission

APPENDIX 4

Competency Categories and Skills Defined for the Four Curriculum Models

Framework	Competency Category	Skill/Topic [Number of Subtopics]
DigCCurr Matrix	Contexts	Cultural Context
DigCCurr Matrix	Contexts	Disciplinary Context
DigCCurr Matrix	Contexts	Institutional or Organizational Context
DigCCurr Matrix	Contexts	Professional Context/History of Professional Activities
DigCCurr Matrix	Contexts	Professional Context/ Professional Development
DigCCurr Matrix	Functions and Skills	Access [9]
DigCCurr Matrix	Functions and Skills	Administration [24]
DigCCurr Matrix	Functions and Skills	Advocacy & Outreach [5]
DigCCurr Matrix	Functions and Skills	Analysis & Characterization of Digital Objects [2]
DigCCurr Matrix	Functions and Skills	Analysis & Evaluation of Producer Information Environment [4]
DigCCurr Matrix	Functions and Skills	Archival Storage [8]
DigCCurr Matrix	Functions and Skills	Collaboration, Coordination, & Contracting [6]
DigCCurr Matrix	Functions and Skills	Common Services [3]
DigCCurr Matrix	Functions and Skills	Data Management [5]
DigCCurr Matrix	Functions and Skills	Description, Organization, & Intellectual Control [12]
DigCCurr Matrix	Functions and Skills	Destruction & Removal [0]
DigCCurr Matrix	Functions and Skills	Identifying, Locating, & Harvesting [5]
DigCCurr Matrix	Functions and Skills	Ingest [8]
DigCCurr Matrix	Functions and Skills	Management [5]
DigCCurr Matrix	Functions and Skills	Preservation Planning & Implementation [6]
DigCCurr Matrix	Functions and Skills	Production [4]
DigCCurr Matrix	Functions and Skills	Purchasing & Managing Licenses to Resources [3]
DigCCurr Matrix	Functions and Skills	Reference & User Support Services [4]
DigCCurr Matrix	Functions and Skills	Selection, Appraisal, & Disposition [7]
DigCCurr Matrix	Functions and Skills	Systems Engineering & Development [9]
DigCCurr Matrix	Functions and Skills	Transfer/First-Level Subfunctions [3]
DigCCurr Matrix	Functions and Skills	Transformation of Digital Objects [0]
DigCCurr Matrix	Functions and Skills	Use, Re-use, & Adding Value to Accessed Information [0]
DigCCurr Matrix	Functions and Skills	Validation & Quality Control of Digital Objects-Packages [5]
DigCCurr Matrix	Functions and Skills/Meta-level Functions	Analysis & Documentation
DigCCurr Matrix	Functions and Skills/Meta-level Functions	Audit of Curation Functions
DigCCurr Matrix	Functions and Skills/Meta-level Functions	Certification
DigCCurr Matrix	Functions and Skills/Meta-level Functions	Education and Sharing
DigCCurr Matrix	Functions and Skills/Meta-level Functions	Evaluation & Audit
DigCCurr Matrix	Functions and Skills/Meta-level Functions	Monitoring and Logging
DigCCurr Matrix	Functions and Skills/Meta-level Functions	Process Mapping
DigCCurr Matrix	Functions and Skills/Meta-level Functions	Research & Development
DigCCurr Matrix	Mandates, Values, and Principles	Abstraction

Framework	Competency Category	Skill/Topic [Number of Subtopics]
DigCCurr Matrix	Mandates, Values, and Principles	Accountability
DigCCurr Matrix	Mandates, Values, and Principles	Adaptability
DigCCurr Matrix	Mandates, Values, and Principles	Authenticity
DigCCurr Matrix	Mandates, Values, and Principles	Automation
DigCCurr Matrix	Mandates, Values, and Principles	Chain of Custody
DigCCurr Matrix	Mandates, Values, and Principles	Collection
DigCCurr Matrix	Mandates, Values, and Principles	Context
DigCCurr Matrix	Mandates, Values, and Principles	Continuum Approach
DigCCurr Matrix	Mandates, Values, and Principles	Critical Inquiry
DigCCurr Matrix	Mandates, Values, and Principles	Diversity
DigCCurr Matrix	Mandates, Values, and Principles	Encapsulation
DigCCurr Matrix	Mandates, Values, and Principles	Evidence
DigCCurr Matrix	Mandates, Values, and Principles	Informating
DigCCurr Matrix	Mandates, Values, and Principles	Interoperability
DigCCurr Matrix	Mandates, Values, and Principles	Long Term
DigCCurr Matrix	Mandates, Values, and Principles	Modularity
DigCCurr Matrix	Mandates, Values, and Principles	Open Architecture
DigCCurr Matrix	Mandates, Values, and Principles	Organizational Learning
DigCCurr Matrix	Mandates, Values, and Principles	Provenance
DigCCurr Matrix	Mandates, Values, and Principles	Robustness
DigCCurr Matrix	Mandates, Values, and Principles	Scale and Scalability
DigCCurr Matrix	Mandates, Values, and Principles	Significant Properties
DigCCurr Matrix	Mandates, Values, and Principles	Stakeholders
DigCCurr Matrix	Mandates, Values, and Principles	Standardization
DigCCurr Matrix	Mandates, Values, and Principles	Sustainability
DigCCurr Matrix	Mandates, Values, and Principles	Trust
DigCCurr Matrix	Prerequisite Knowledge	Definitions of Technology
DigCCurr Matrix	Prerequisite Knowledge	Essential Characteristics of... ICT [information and communication technology] Landscape
DigCCurr Matrix	Prerequisite Knowledge	History and Evolution of ICTs
DigCCurr Matrix	Prerequisite Knowledge	Terminology
DigCCurr Matrix	Transition Point in Information Continuum	Transition Points of Digital Objects
DigCCurr Matrix	Type of Resource	Format
DigCCurr Matrix	Type of Resource	Genre
DigCCurr Matrix	Type of Resource	Level of Abstraction
DigCCurr Matrix	Type of Resource	Level of Aggregation
DigCCurr Matrix	Type of Resource	Medium
DigCurV	Knowledge and Intellectual Abilities (KIA)/Data Skills [KIA5]	Data structures and types [KIA5.1]
DigCurV	Knowledge and Intellectual Abilities (KIA)/Data Skills [KIA5]	Database types and structures [KIA5.3]
DigCurV	Knowledge and Intellectual Abilities (KIA)/Data Skills [KIA5]	Execute analysis of and forensic procedures in digital curation [KIA5.4]
DigCurV	Knowledge and Intellectual Abilities (KIA)/Data Skills [KIA5]	File types, applications, and systems [KIA5.2]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Evaluation Studies [KIA3]	Conduct usability evaluation [KIA3.6]

Framework	Competency Category	Skill/Topic [Number of Subtopics]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Evaluation Studies [KIA3]	Conduct user needs analysis [KIA3.3]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Evaluation Studies [KIA3]	Continuous monitor and evaluate digital curation technologies [KIA3.4]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Evaluation Studies [KIA3]	Monitor and assess needs of designated community [KIA3.5]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Evaluation Studies [KIA3]	Prioritize curation activities based on value of digital objects and the risks facing them [KIA3.7]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Evaluation Studies [KIA3]	Prioritize funding for curation activities based on the value of digital objects and the risks facing objects [KIA3.1]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Evaluation Studies [KIA3]	Respond to findings from user studies constructively in future decision-making [KIA3.2]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Information Skills [KIA4]	Apply metadata standards [KIA4.6]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Information Skills [KIA4]	Deploy appropriate information-seeking strategies [KIA4.3]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Information Skills [KIA4]	Information-seeking strategies, access technologies, and user sharing behaviors [KIA4.1]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Information Skills [KIA4]	Key metadata standards for sector/subject [KIA4.4]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Information Skills [KIA4]	Relationship between appropriate controlled vocabularies and metadata standards [KIA4.7]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Information Skills [KIA4]	Select metadata standards [KIA4.5]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Information Skills [KIA4]	Support information access and sharing [KIA4.2]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Selection/Appraisal [KIA2]	Articulate information and records management principles [KIA2.2]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Selection/Appraisal [KIA2]	Articulate the benefits and long-term value of collections [KIA2.3]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Selection/Appraisal [KIA2]	Contribute to institutional policies, including criteria for selection/appraisal [KIA2.4]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Selection/Appraisal [KIA2]	Information- and records-management principles [KIA2.5]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Selection/Appraisal [KIA2]	Institutional policies, including criteria for selection/ appraisal [KIA2.6]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Selection/Appraisal [KIA2]	Maximize benefits and long-term value of collections [KIA2.1]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Selection/Appraisal [KIA2]	Plan application of selection/appraisal criteria to collections [KIA2.7]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Apply appropriate technological solutions [KIA1.9]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Current and emerging subject landscape (trends, people, institutions) [KIA1.3]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Designated community [KIA1.7]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Develop a professional network for support [KIA1.10]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Digital curation and preservation terminology [KIA1.13]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Digital curation tools (at high level) [KIA1.11]

Framework	Competency Category	Skill/Topic [Number of Subtopics]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Digital preservation standards [KIA1.12]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Fundamental digital curation principles, including life cycles [KIA1.6]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Information technology definitions and skills [KIA1.15]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Relevance of, and need for, digital curation activity within subject context [KIA1.2]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Respective responsibilities for digital curation across institution [KIA1.4]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Scope of own role within institutional context [KIA1.17]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Scope of team responsibilities within institution [KIA1.14]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Scope the boundaries for digital curation at institution [KIA1.5]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Select and apply digital curation and preservation techniques [KIA1.16]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Select appropriate technological solutions [KIA1.8]
DigCurV	Knowledge and Intellectual Abilities (KIA)/ Subject Knowledge (KIA1)	Subject-specific knowledge and definitions [KIA1.1]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Audit and certification standards [MQA2.1]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Audit of curation functions [MQA2.8]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Benefits of audit process, and relevance of audit results [MQA2.2]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Certification of repositories or programs [MQA2.9]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Institutional liabilities in audit process [MQA2.3]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Lead repository through certification process [MQA2.5]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Level of audit appropriate to institution [MQA2.4]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Maintain documentation in preparation for audit process [MQA2.10]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Prepare effectively for an audit of curation functions [MQA2.7]
DigCurV	Management and Quality Assurance (MQA)/ Audit and Certification [MQA2]	Respond to audit report and build new service plan where required [MQA2.6]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Create a team environment [MQA3.10]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Creation, management, and monitoring of project plans [MQA3.13]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Data management requirements [MQA3.15]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Deal with data curation challenges through structured planning [MQA3.17]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Make sound decisions based on information produced by project team [MQA3.8]

Framework	Competency Category	Skill/Topic [Number of Subtopics]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Make sustainable storage decisions in institutional context [MQA3.12]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Plan and implement sound staff training and development [MQA3.11]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Produce relevant information to support decision-making [MQA3.16]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Project management concepts and techniques [MQA3.18]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Recruit and motivate staff [MQA3.9]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Reputation management [MQA3.4]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Resources required for digital curation activity, including energy consumption [MQA3.3]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Respond to staff recruitment, training, and development needs [MQA3.5]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Undertake business continuity management, including disaster planning [MQA3.2]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Undertake business planning in line with corporate/ institutional goals [MQA3.7]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Undertake financial planning, cost analysis, and economic sustainability [MQA3.6]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Undertake project management activities and innovative practices [MQA3.14]
DigCurV	Management and Quality Assurance (MQA)/ Resource Management [MQA3]	Undertake strategic planning [MQA3.1]
DigCurV	Management and Quality Assurance (MQA)/ Risk Management [MQA1]	Apply risk management practice, techniques, and standards to digital curation activities within institutional risk management context [MQA1.3]
DigCurV	Management and Quality Assurance (MQA)/ Risk Management [MQA1]	Assess, analyze, monitor, and communicate risks [MQA1.4]
DigCurV	Management and Quality Assurance (MQA)/ Risk Management [MQA1]	Risk management theory and standards [MQA1.2
DigCurV	Management and Quality Assurance (MQA)/ Risk Management [MQA1]	Undertake succession planning [MQA1.1]
DigCurV	Personal Qualities (PQ)/Communication and Advocacy Skills [PQ2]	Articulate importance of digital curation to peers, other staff, and public [PQ2.2]
DigCurV	Personal Qualities (PQ)/Communication and Advocacy Skills [PQ2]	Articulate value of collections to peers, other staff, and public [PQ2.3]
DigCurV	Personal Qualities (PQ)/Communication and Advocacy Skills [PQ2]	Communicate across domains, staff groups, and with other relevant communities [PQ2.1]
DigCurV	Personal Qualities (PQ)/Communication and Advocacy Skills [PQ2]	Communication protocols for designated community [PQ2.9]
DigCurV	Personal Qualities (PQ)/Communication and Advocacy Skills [PQ2]	Engage with wider digital curation community [PQ2.8]
DigCurV	Personal Qualities (PQ)/Communication and Advocacy Skills [PQ2]	Make case for funding of digital curation activity [PQ2.4]
DigCurV	Personal Qualities (PQ)/Communication and Advocacy Skills [PQ2]	Make case for staff training and development [PQ2.7]
DigCurV	Personal Qualities (PQ)/Communication and Advocacy Skills [PQ2]	Manage and foster stakeholder relationships [PQ2.5]
DigCurV	Personal Qualities (PQ)/Communication and Advocacy Skills [PQ2]	Plan and deliver dissemination activities [PQ2.6]

Framework	Competency Category	Skill/Topic [Number of Subtopics]
DigCurV	Personal Qualities (PQ)/Integrity [PQ1]	Demonstrate leadership in high-quality standards of work [PQ1.4]
DigCurV	Personal Qualities (PQ)/Integrity [PQ1]	Identify malpractice [PQ1.5]
DigCurV	Personal Qualities (PQ)/Integrity [PQ1]	Make transparent decisions [PQ1.3]
DigCurV	Personal Qualities (PQ)/Integrity [PQ1]	Responsibility, accountability, and good practice in digital curation [PQ1.1]
DigCurV	Personal Qualities (PQ)/Integrity [PQ1]	Value of policy formulation to deal with malpractice [PQ1.2
DigCurV	Personal Qualities (PQ)/Responsiveness to Change [PQ3]	Assess, extend and generate digital curation models for cultural heritage domain [PQ3.7]
DigCurV	Personal Qualities (PQ)/Responsiveness to Change [PQ3]	Cultivate and maintain relationships with other relevant sources of information in digital curation (individuals/services/institutions) [PQ3.4]
DigCurV	Personal Qualities (PQ)/Responsiveness to Change [PQ3]	Emerging developments in discipline and their applicability to digital curation activity in the institution [PQ3.3]
DigCurV	Personal Qualities (PQ)/Responsiveness to Change [PQ3]	Maintain continuous awareness of emerging developments in digital curation [PQ3.8]
DigCurV	Personal Qualities (PQ)/Responsiveness to Change [PQ3]	Potential developments in business models, strategic planning, and management models in digital curation [PQ3.1]
DigCurV	Personal Qualities (PQ)/Responsiveness to Change [PQ3]	Potential of developments in digital curation to influence new services and tools [PQ3.2]
DigCurV	Personal Qualities (PQ)/Responsiveness to Change [PQ3]	Translate current digital curation knowledge into new services and tools [PQ3.9]
DigCurV	Personal Qualities (PQ)/Responsiveness to Change [PQ3]	Translate knowledge of technology and processes into services and tools for needs of designated community [PQ3.6]
DigCurV	Personal Qualities (PQ)/Responsiveness to Change [PQ3]	Value of new and emerging digital curation technologies and processes [PQ3.5]
DigCurV	Professional Conduct (PC)/Ethics, Principles and Sustainability [PC3]	Adhere to principles of ethical conduct [PC3.4]
DigCurV	Professional Conduct (PC)/Ethics, Principles and Sustainability [PC3]	Embed principles of ethical conduct throughout institutional policies (including those affecting curation activity) [PC3.3]
DigCurV	Professional Conduct (PC)/Ethics, Principles and Sustainability [PC3]	Energy consumption and carbon footprint of digital curation activity [PC3.2]
DigCurV	Professional Conduct (PC)/Ethics, Principles and Sustainability [PC3]	Evaluate and treat employees fairly [PC3.5]
DigCurV	Professional Conduct (PC)/Ethics, Principles and Sustainability [PC3]	Social and ethical responsibility in digital curation [PC3.1]
DigCurV	Professional Conduct (PC)/Regulatory Compliance [PC2]	Apply appropriate actions to curation workflow to ensure compliance with legal and policy frameworks and relevant standards [PC2.4]
DigCurV	Professional Conduct (PC)/Regulatory Compliance [PC2]	Contribute to institutional regulatory framework in which digital repositories operate [PC2.3]
DigCurV	Professional Conduct (PC)/Regulatory Compliance [PC2]	Incorporate legal requirements into institutional policies [PC2.2]
DigCurV	Professional Conduct (PC)/Regulatory Compliance [PC2]	Institution's legal culpabilities in digital curation activity [PC2.1]
DigCurV	Professional Conduct (PC)/Regulatory Compliance [PC2]	Select and apply validation techniques to detect policy infringement [PC2.5]

Framework	Competency Category	Skill/Topic [Number of Subtopics]
DigCurV	Professional Conduct (PC)/Regulatory Requirements [PC1]	Contribute to national/international regulatory frameworks in which digital repositories operate [PC1.3]
DigCurV	Professional Conduct (PC)/Regulatory Requirements [PC1]	Domain policies and standards for management and preservation of digital objects [PC1.2]
DigCurV	Professional Conduct (PC)/Regulatory Requirements [PC1]	Legal frameworks in which digital curation is taking place [PC1.1]
NRC Report	Functions and Skills/Archiving and preservation	Approaches (e.g., emulation, migration, canonicalization)
NRC Report	Archiving and Preservation	Assess trustworthiness of repositories (e.g., TRAC)
NRC Report	Archiving and Preservation	Authenticating users
NRC Report	Archiving and Preservation	Forensic role of digital repositories
NRC Report	Archiving and Preservation	Integrity and security
NRC Report	Archiving and Preservation	Preservation models (e.g., OAIS, LOCKSS, PLANETS)
NRC Report	Archiving and Preservation	Resources, methods, and data practices of disciplines
NRC Report	Data Analytics	Algorithmic thinking and programming
NRC Report	Data Analytics	Data mining
NRC Report	Data Analytics	Hypothesis development and testing
NRC Report	Data Analytics	Information extraction
NRC Report	Data Analytics	Performance evaluation and risk analysis
NRC Report	Data Analytics	Research design
NRC Report	Data Analytics	Sampling techniques
NRC Report	Data Analytics	Statistics
NRC Report	Data Collection and Management	Data acquisition or harvesting
NRC Report	Data Collection and Management	Deselecting and destroying data
NRC Report	Data Collection and Management	Gathering and analyzing requirements
NRC Report	Data Collection and Management	Identifying and selecting data
NRC Report	Data Collection and Management	Ingestion or deposit (e.g., identifiers, citations, versions)
NRC Report	Data Collection and Management	Linkages to literature and data
NRC Report	Data Collection and Management	Prepare for use (e.g., cleaning, reformatting, anonymizing)
NRC Report	Data Collection and Management	Provenance and context for preservation
NRC Report	Data Collection and Management	Support data annotation and publication
NRC Report	Data Practices	Data processing, transformation, documentation processes
NRC Report	Data Practices	Disciplinary, professional, and institutional practices
NRC Report	Data Practices	Quantitative and qualitative data types and formats
NRC Report	Data Practices	Research methods, instruments, tools, and protocols
NRC Report	Data Practices	Standards (e.g, data, schemas, ontologies, technologies)
NRC Report	Data Practices	Standards of evidence, quality, and uncertainty
NRC Report	General Background and Abilities	Communicate, collaborate, innovate, prioritize
NRC Report	General Background and Abilities	Heterogeneity, complexity, and volume
NRC Report	General Background and Abilities	Math and science; domain specialization
NRC Report	Management and Administration	Cost-benefit analysis
NRC Report	Management and Administration	Cross-institutional coordination
NRC Report	Management and Administration	Expectation management, complaint handling

Framework	Competency Category	Skill/Topic [Number of Subtopics]
NRC Report	Management and Administration	Grant and report writing
NRC Report	Management and Administration	Project management and planning
NRC Report	Management and Administration	Staff development
NRC Report	Management and Administration	Strategic planning
NRC Report	Management and Administration	Supervision
NRC Report	Management and Administration	Training
NRC Report	Policy and Planning	Archival principles (e.g., collection, retention, preservation, rescue)
NRC Report	Policy and Planning	Conformance with legal mandates, best practices, expectations
NRC Report	Policy and Planning	Institutional, national, international policies
NRC Report	Policy and Planning	Intellectual property, rights management, licensing, agreements
NRC Report	Policy and Planning	Risk assessment, disaster planning, sustainability
NRC Report	Presentation and Visualization	Evaluation of products, algorithms, specific programs
NRC Report	Presentation and Visualization	Information design and contextualization
NRC Report	Services and Support	Current awareness services (push and pull)
NRC Report	Services and Support	Enhancement, including metadata, annotation, linking
NRC Report	Services and Support	Information resource development
NRC Report	Services and Support	Instruction and training
NRC Report	Services and Support	Liaison and consulting
NRC Report	Services and Support	Outreach, advocacy, and promotion
NRC Report	Services and Support	Support for virtual communities
NRC Report	Technologies, Tools, and Infrastructure	Access systems
NRC Report	Technologies, Tools, and Infrastructure	Data acquisition (e.g., instrumentation, sensors, laboratory notebooks)
NRC Report	Technologies, Tools, and Infrastructure	Data modeling
NRC Report	Technologies, Tools, and Infrastructure	Database design, construction, and management
NRC Report	Technologies, Tools, and Infrastructure	Interoperability
NRC Report	Technologies, Tools, and Infrastructure	Markup languages
NRC Report	Technologies, Tools, and Infrastructure	Network architecture
NRC Report	Technologies, Tools, and Infrastructure	Preservation systems
NRC Report	Technologies, Tools, and Infrastructure	Repository infrastructure
NRC Report	Technologies, Tools, and Infrastructure	Software development environments
NRC Report	Technologies, Tools, and Infrastructure	System administration
NRC Report	Technologies, Tools, and Infrastructure	Technology assessment
NRC Report	Technologies, Tools, and Infrastructure	Usability testing
NRC Report	Technologies, Tools, and Infrastructure	Web services
NRC Report	Values and Principles	Analyze ethical dilemmas to recommend resolutions
NRC Report	Values and Principles	Conformance to relevant privacy provisions, legitimate expectations of users
NRC Report	Values and Principles	Ethical, legal, cultural, and economic considerations for evidentiary record
NRC Report	Values and Principles	Principled activities grounded in fundamental values
NRC Report	Values and Principles	Regulations, policies, norms, values of access, privacy, retention, repurposing

Framework	Competency Category	Skill/Topic [Number of Subtopics]
NRC Report	Values and Principles	Values and principles of respective discipline, service organization
NDSA Survey	Functions and Skills/Digital Preservation	Content replication
NDSA Survey	Functions and Skills/Digital Preservation	Creation of access copies
NDSA Survey	Functions and Skills/Digital Preservation	Descriptive cataloging
NDSA Survey	Functions and Skills/Digital Preservation	Development and maintenance of tools
NDSA Survey	Functions and Skills/Digital Preservation	Development of guidelines for content creators
NDSA Survey	Functions and Skills/digital preservation	Development of preservation policies and strategy
NDSA Survey	Functions and Skills/Digital Preservation	Digitization
NDSA Survey	Functions and Skills/Digital Preservation	Emulation
NDSA Survey	Functions and Skills/Digital Preservation	File format identification
NDSA Survey	Functions and Skills/Digital Preservation	File format validation
NDSA Survey	Functions and Skills/Digital Preservation	Fixity checks
NDSA Survey	Functions and Skills/Digital Preservation	Metadata creation/extraction
NDSA Survey	Functions and Skills/Digital Preservation	Normalization of files
NDSA Survey	Functions and Skills/digital preservation	Preservation education, training and outreach
NDSA Survey	Functions and Skills/Digital Preservation	Preservation planning
NDSA Survey	Functions and Skills/Digital Preservation	Research
NDSA Survey	Functions and Skills/Digital Preservation	Secure storage management
NDSA Survey	Functions and Skills/Digital Preservation	Selection for preservation
NDSA Survey	Functions and Skills/Digital Preservation	Technology watch
NDSA Survey	Functions and Skills/Digital Preservation	Transformation / migration of formats

APPENDIX 5

Job Postings to the American Library Association (ALA) DigiPres Electronic Mailing List, 2013–2015

Explanations of the columns in the table:

Date: Month and year of the job posting on the ALA digipres listserv.

Title: Job title in the posting.

Type of Institution: General term for the type of institution that posted the job.

Desc?: Indication if an extended description of the job is contained in the e-mail [E], included as an attachment [A] or at URL [U], or not included at all (i.e., not described and pointing to an inactive link [N]). Lower case 'e' indicates there is a very brief description in e-mail with no attachment or active link.

U.S.: Indication if institution that posted the job is in the United States (Y or N).

Date	Job Posting Title	Type of Institution	Desc?	U.S.?
2013/01	Project Manager, Preservation & Conservation	Academic Library	E	Y
2013/01	Digitization Program Librarian	National Collection	E	Y
2013/01	Museum Archives Manager [hybrid]	Art Museum	E	Y
2013/01	Digital Archivist	Academic Library	E	Y
2013/02	Metadata Cataloger	Library Services Provider	E	Y
2013/02	Digital Preservation Manager	Academic Library	E	Y
2013/02	Director, Preservation Services	Preservation Services	e	Y
2013/02	Metadata Specialist	Library Services Provider	E	Y
2013/02	Digital Library Software Engineer	Curation Services Provider	e	Y
2013/02	Project Developer, Data Management Plan Tool	Curation Services Provider	E	Y
2013/02	Production Systems Architect and Administrator	Curation Services Provider	E	Y
2013/03	Project Archivists [multiple jobs, physical collections]	Academic Library	E	Y
2013/03	Digital Archivist	City Archives	E	N
2013/04	Library Technicians [multiple jobs]	Library Services Provider	E	Y
2013/05	Metadata Specialists [multiple jobs]	Library Services Provider	E	Y
2013/05	Assistant Director, Outreach Librarian	Library Services Provider	N	Y
2013/05	Digital Metadata Librarian	Library Services Provider	N	Y
2013/05	Ingest Operator–Media Coordinator	Library Services Provider	E	Y
2013/05	Library Clerks [multiple jobs, physical collections]	Library Services Provider	E	Y
2013/06	Data Management Services Librarian	Academic Library	E	Y
2013/06	Special Collections Librarian [hybrid]	Academic Library	E	Y
2013/07	Metadata Librarian	Academic Library	e	Y
2013/07	Tape Value / Media Archives Specialist	Library Services Provider	E	Y
2013/07	Digital Projects Librarian	Academic Library	A	Y
2013/07	Digital Continuity–Senior Adviser	National Archives	E	N
2013/07	Digital Preservation Manager	National Library	E	N
2013/08	University Archivist [hybrid]	Academic Library	E	Y
2013/08	Digital Services Librarian	Academic Library	e	Y
2013/08	Digital Initiatives Librarian	Academic Library	E	Y

Date	Job Posting Title	Type of Institution	Desc?	U.S.?
2013/08	Digital Preservation Librarian	Academic Library	E	Y
2013/09	Digital Preservation Manager	National Library	E	Y
2013/09	Senior Software Developer	Preservation Services	E	Y
2013/10	Systems Librarian	Library Services Provider	E	Y
2013/11	Head, Digital Services Unit [repost 2014/05&12]	Academic Library	E	Y
2013/11	Project Manager, Digital Stewardship Residencies	Academic Library	N	Y
2013/11	Lecturer–Preservation Librarian	Academic Library	E	Y
2013/11	Head, Preservation and Collection Development	National Collection	e	Y
2013/12	Strategic Program Specialist	Library Services Provider	N	Y
2013/12	Preservation Librarian	Academic Library	A	Y
2014/02	Preservation and Collection Management Librarian	Academic Library	E	Y
2014/02	Archivist / Metadata Specialist	Academic Library	E	Y
2014/02	Head, Preservation and Reformatting	Academic Library	E	Y
2014/03	Digital Preservation Technical Specialist	National Library	E	N
2014/03	Repository Technical Project Manager	Library Professional Association	E	Y
2014/03	Digital Asset Management Librarian	Library Services Provider	E	Y
2014/03	Conversion Support Services Manager [AV]	National Collection	N	Y
2014/03	Preservation Project Manager [AV]	Media Services Provider	E	Y
2014/05	Digital Archivist	City Archives	N	N
2014/05	Quality Control Specialist, Media Digitization	Academic Library	e	Y
2014/05	Head, Metadata and Digitization	Academic Library	E	Y
2014/06	[Announcement of 5 vacancies at e-research center]	University	N	N
2014/06	Head, Digital Research and Publishing	Academic Library	Y	Y
2014/07	Director, Digital Scholarship and Publishing	Academic Library	N	Y
2014/07	Digital Initiatives Librarian	Academic Library	N	Y
2014/07	Digital Preservation Analyst	Academic Library	e	Y
2014/07	Digital Preservation Process Administrator	National Archives	E	N
2014/08	Rights and Reproductions Coordinator	Art Museum	e	Y
2014/08	Digital Initiatives Librarian	Academic Library	E	Y
2014/08	Data Management Services Librarian	Academic Library	E	Y
2014/08	Director, Imaging Services	Library Services Provider	e	Y
2014/08	Executive Director	Library Professional Association	N	Y
2014/09	Media Services Program Manager	Library Services Provider	E	Y
2014/09	Curation Archivist	Academic Library	E	Y
2014/11	Digital Preservation Librarian	Academic Library	E	Y
2014/12	Data Librarian	Academic Library	E	Y
2015/01	Digital Initiatives Librarian	Academic Library	E	Y
2015/01	Digital Archivist	National Collection	N	Y
2015/02	Digital Archivist	Academic Library	E	Y
2015/02	Director of Library	Academic Library	E	Y
2015/02	Special Collections Librarian [hybrid]	Academic Library	e	Y
2015/02	Team Leader, Digital Learning and Scholarship	Academic Library	E	Y
2015/02	Electronic Resources Librarian	Academic Library	e	Y
2015/03	Expert Audiovisual Consultants	National Collection	e	Y

Date	Job Posting Title	Type of Institution	Desc?	U.S.?
2015/03	Digital Infrastructure Librarian	Academic Library	E	Y
2015/03	Project Manager/DP Specialist	Academic Library	E	Y
2015/03	Data Services Librarian/Specialist	Academic Library	E	Y
2015/03	Quality Control Specialist	University	e	Y
2015/03	Quality Control Specialist Audio	University	e	Y
2015/03	Head of Research and Development Department	Academic Library	N	N
2015/03	Head of Audio Moving Image Preservation	Public Library	U	Y
2015/04	AV Archivist	National Collection	N	Y
2015/04	Digital Archivist	Academic Library	E	Y
2015/04	Conservator	Academic Library	E	Y
2015/04	Archivist	National Collection	U	Y
2015/05	Digital Archivist	Historical Society	U	Y
2015/05	Digital Projects Librarian	Academic Library	E	Y
2015/05	Head of Digital Preservation	Public Library	E	Y
2015/05	Metadata Librarian	Academic Library	E	Y
2015/06	Applications Analyst	National Collection	N	Y
2015/06	Program Coordinator, Digital Stewardship Residencies	Library Services Provider	N	Y
2015/07	Associate Dean for Academic Services	Academic Library	A	Y
2015/07	Manager, Rare Books and Special Collections	Academic Library	E	N
2015/07	Library Director	Academic Library	E	N
2015/07	Web Archiving and Emerging Formats Librarian	Academic Library	E	Y
2015/07	Information Technology Analyst III [digital library program]	Academic Library	E	Y
2015/08	Preservation Librarian	Library Services Provider	N	Y
2015/08	Executive Director	DP Membership Organization	U	N
2015/09	Collection Design and Assessment Librarian	Academic Library	A	Y
2015/09	Audiovisual Archivist	Academic Library	E	Y
2015/09	Digital Preservation Analyst	Academic Library	E	Y
2015/09	Head, Special Collections and University Archives	Academic Library	E	Y
2015/09	Digital Archivist	Academic Library	U	N
2015/09	Manager, Library Operations–Digital Services	Academic Library	E	Y
2015/09	Digital Projects Librarian	Academic Library	E	Y
2015/09	Metadata Librarian	Academic Library	E	Y
2015/09	Software Developer	Academic Library	E	Y
2015/10	Newspaper Digitization Project Librarian	Academic Library	e	Y
2015/10	Systems Archivist	Preservation Services	U	N
2015/10	Library Information Associate (Digital)	Academic Library	E	Y
2015/11	Preservation Services Manager	Academic Library	E	Y
2015/11	Multimedia and Digital Collections Archivist	Academic Library	E	Y
2015/11	Digital Curation Coordinator	Library Services Network	A	Y