Sustaining Domain Repositories for Digital Data: A White Paper

December 11, 2013

Prepared by Carol Ember (HRAF, Yale University)
and Robert Hanisch (VAO, Space Telescope Science Institute)

With contributions by George Alter (ICPSR, University of Michigan), Helen Berman (PDB, Rutgers University), Margaret Hedstrom (SEAD, University of Michigan), and Mary Vardigan (ICPSR, University of Michigan)

This document is the outcome of a workshop on "Sustaining Domain Repositories for Digital Data," June 24-25, 2013, which was convened by ICPSR at the University of Michigan and supported by a grant from the Alfred P. Sloan Foundation.

Executive Summary

The last few years have seen a growing international movement to enhance research transparency, open access to data, and data sharing across the social and natural sciences. Meanwhile, new technologies and scientific innovations are vastly increasing the amount of data produced and the resultant potential for advancing knowledge. Domain repositories — data archives with ties to specific scientific communities — have an indispensable role to play in this changing data ecosystem. With both content-area and digital curation expertise, domain repositories are uniquely capable of ensuring that data and other research products are adequately preserved, enhanced, and made available for replication, collaboration, and cumulative knowledge building. However, the systems currently in place for funding repositories in the US are inadequate for these tasks. Effective and innovative funding models are needed to ensure that research data, so vital to the scientific enterprise, will be available for the future. Funding models also need to assure equal access to data preservation and curation services regardless of the researcher's institutional affiliation. Creating sustainable funding streams requires coordination amongst multiple stakeholders in the scientific, archival, academic, funding, and policy communities.

Background

Not only has there been a vast increase in the amount of digital data, but there has also been global increase in activity related to research transparency, open access data, and data sharing. In February 2013, the U.S. Government's Office of Science and Technology

Policy (OSTP) issued a memorandum calling for all federal agencies with an annual R&D budget over $100 million to create plans for public access to research projects.[1] Recognizing these challenges, on June 24–25, 2013, representatives from 22 data repositories spanning the social and natural sciences met in Ann Arbor, MI. The meeting, organized by the Interuniversity Consortium for Political and Social Research (ICPSR) and supported by the Alfred P. Sloan Foundation, created a space to discuss the challenges facing repositories across domains, and to strategize around issues of sustainability.

Value and Role of Domain Repositories

Domain repositories in the social and natural sciences each serve a scientific community, which may be a traditional academic discipline, a subdiscipline, or an interdisciplinary network of scientists, united by a common focus. This in-depth knowledge enables domain repositories to enhance the data ecosystem far beyond data preservation and access. By combining domain-specific scientific knowledge, expertise in data stewardship, and close relationships with scientific communities, domain repositories accelerate intellectual discovery by facilitating reuse and reproducibility, ultimately building an enduring record that represents the richness, diversity, and complexity of the scientific enterprise.

Far from simply storing digital data, domain repositories can use these relationships to:

- **Manage** data in a way that maintains its understandability and usability for the scientific community.
- **Facilitate** data discovery and reuse through the development and standardization of metadata.
- **Provide Access** while ensuring necessary protections related to confidentiality and intellectual property.
- **Create** systems that facilitate future archiving (active data curation) while research is undertaken.
- **Respond** to the unique and evolving needs of scientific communities and othe**r** stakeholders.
- **Partner** with each community to create guidelines for data stewardship throughout the data life cycle.
- **Advocate** for transparency, data access, and data sharing.
- **Innovate** in the realm of data curation to address new and evolving forms of data.
- **Add Value** through the creation of data products that align with best practices and new technologies.
- **Collaborate** with related disciplines to achieve interoperability across scientific communities.
- **Mediate** among scientific communities and digital libraries and archives to implement the latest developments in information science.

[1] John P. Holdren, "Increasing Access to the Results of Federally Funded Scientific Research," Memorandum for the Heads of Executive Departments and Agencies, February 22, 2013, http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

The Challenge

Despite the growing demand for data sharing and access, domain repositories face an uncertain financial future in the United States. The need for data archives is rising due to open access mandates, research innovations, and the growing volume of scientific data that needs to be curated, preserved, and disseminated. Yet funding for domain repositories remains unpredictable and inadequate for the task at hand. Of particular concern is the mismatch between the long-term commitment to preservation inherent in the work of archiving, and the short-term and episodic funding upon which this work is based. Many archives rely primarily on project-based grants, even though the expectation of stakeholders is that data will be available and usable indefinitely.[2]

Another concern is that the push towards open access, while creating more equity of access for the community of users, creates a burden for domain repositories because it narrows their funding possibilities. Without care, this shift may create a different kind of inequity—less well-funded scholars or institutions will be less likely to have their products of research preserved for the future.

A Call for Change

Domain repositories must be funded as the essential piece of the U.S. research infrastructure that they are. This means:

- Ensuring funding streams that are long-term, uninterrupted, and flexible.
- Creating systems that promote good scientific practice.
- Assuring equity in participation and access.

There may not be one solution to the problem — repositories may very well need different funding models across domain and repository type. But in every case, *creating sustainable funding streams will require the coordinated response of multiple stakeholders in the scientific, archival, academic, funding, and policy communities.*

1. Introduction

The available evidence strongly suggests that sharing research data results in the production of more science, and that strategic investments in data curation and preservation are highly efficient ways to spend scarce public funds. By bridging the gap between scientific communities and the rapid changes in technology and information science, domain repositories play a key role in assuring that valuable data remain accessible, discoverable, and meaningful. However, most domain repositories lack sustained sources of funding,

[2] Lyle, Jared, Alter, George, and Vardigan, Mary. "'The Price of Keeping Knowledge' Workshop: ICPSR Position Paper." (2013) http://www.knowledge-ex-change.info/Admin/Public/DWSDownload.aspx?File=%2FFiles%2FFiler%2Fdownloads%2FPrimary+Research+Data%2FWorkshop+Price+of+Keeping+Knowledge%2FJared+Lyle+ICPSR_Position+Paper_Price+workshop_public.pdf

and there is a mismatch between their role as long-term guardians of valuable scientific resources and the short-term perspective of current funding mechanisms.

Domain repositories are archives of digital and/or digitized information related to an area of research. They can be broad in scope, such as the Inter-university Consortium for Political and Social Research (ICPSR[3]) at the University of Michigan, which supports research in the social and behavioral sciences, or more tightly focused, such as the Mikulski (Multi-mission) Archive at Space Telescope (MAST[4]) at the Space Telescope Science Institute, which contains data from NASA space astronomy missions in the optical, ultraviolet, and near-infrared parts of the electromagnetic spectrum. Domain repositories represent a significant body of knowledge that spans multiple experiments, facilities, studies, etc., and whose data collections support ongoing scientific research through integrated discovery and access to heterogeneous data.

Domain repositories are playing an increasingly important role in facilitating research and promoting the re-use and re-purposing of data, thereby enhancing the return on investment from public research funding. Figure 1 shows the use of archival data in MAST, for the Hubble Space Telescope (HST), as a function of time. In 2012, of the nearly 850 peer-reviewed publications based on HST data, nearly 500 (60%) were based

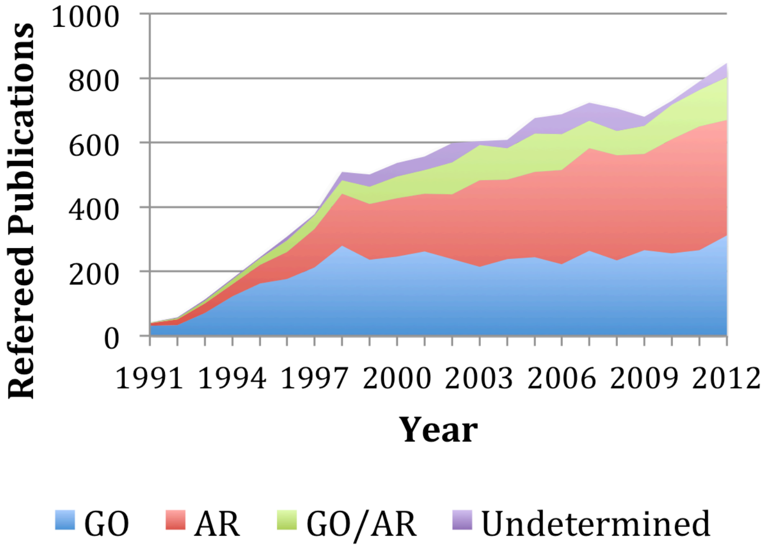

Figure 1. Number of refereed publications based on Hubble Space Telescope data in the Multi-mission Archive at the Space Telescope Science Institute. GO = guest observer programs (papers published by the principal investigator and immediate collaborators), AR = archival research (papers published by researchers not affiliated with the principal investigator), GO+AR = papers that include both GO and AR data, and Undetermined = papers for which the origin of the data is unclear.

[3] http://www.icpsr.umich.edu/icpsrweb/landing.jsp

[4] http://archive.stsci.edu/

wholly or in part on archival data. The archival research category (AR) represents papers published by researchers not connected in any way with the original guest observer (GO). An example from the social sciences is the expansion of cross-cultural research after the formation of the nonprofit membership consortium called the Human Relations Area Files (HRAF) in 1949. The annually-growing collection of ethnographic information subject-indexed at the paragraph level was designed to facilitate finding information quickly so that researchers could test hypotheses on worldwide information. Prior to the formation of the "HRAF files" (as they were known for short), there were only ten published cross-cultural studies. Only available to a few founding institutions, there were only 18 studies in the next decade, but as the HRAF Collection became more widely distributed in the late 1950s to over 140 institutions, there were 100 published cross-cultural studies in the decade from 1958-1967. Not all of these studies used HRAF, but it is estimated that about 42% of the published studies did so. Other cross-cultural datasets were created in the late 1960s and although there are as yet no complete available counts today there were over 400 cross-cultural studies by 1981 and over 700 by 1989.[5] Studies of other domains in both the social and natural sciences have shown marked improvement in research returns from archived data.[6]

Domain repositories play a different role from institutional repositories. The latter focus on research products across an organization—a university or association of universities, for example—and thus deal with a tremendous diversity of data types that are not necessarily comparable to each other. Domain repositories focus on data that benefit from being used in relation to, and in comparison with, other data in the collection. Thus, domain repositories call for a particularly active *curation* function in order to assure the greatest possible interoperability amongst the datasets in the collection.

In some fields domain repositories are both well supported and heavily utilized. In others there is a lack of funding for even the basic infrastructure of storage and preservation, meaning that the products of publicly funded research are in danger of being permanently lost, and the opportunities for creative re-use of such information are also lost. Where funding does exist, it is often associated with a particular project or program and lacks any long-term commitment. This can be disruptive to the point of also leading to data

[5] A more complete list of studies is under development at the Human Relations Area Files. For preliminary estimates see David Levinson, "Holocultural studies based on the Human Relations Area Files," *Cross-Cultural Research* 4 (1978): 296; David Levinson, “Bibliography of substantive worldwide cross-cultural studies,” *Cross-Cultural Research* 24 (1990): 105; Carol R. Ember, “Human Relations Area Files” in William Sims Bainbridge, ed. *Leadership in Science and Technology: A Reference Handbook*, Vol. 2. Los Angeles: Sage, 2012, p. 622. These estimates concentrate on hypothesis-testing studies of ten or more societies and do not include other uses of the databases.

[6] Pienta, Amy M.; Alter, George C.; Lyle, Jared A. "The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data." (2010) http://hdl.handle.net/2027.42/78307; Piwowar HA, Day RS, Fridsma DB "Sharing Detailed Research Data Is Associated with Increased Citation Rate." PLoS ONE (2007): 2(3): e308. doi:10.1371/journal.pone.0000308; Heather A Piwowar, Todd J Vision, Michael C Whitlock. "Data archiving is a good investment." Nature 473, 285 (19 May 2011) doi:10.1038/473285a

loss. The broader community of data creators and users does not fully appreciate what it takes to preserve data for future use. This leads to assumptions that online storage using systems like Dropbox are adequate, ignoring the needs of curation, preservation, interoperability, and metadata. The February 2013 directive from the Office of Science and Technology Policy[7] to federal agencies to increase access to the results of federally funded scientific research establishes the government's awareness of the problem, but provides no new resources toward its solution.

2. Domain Repository Functions

Domain repositories have a variety of functions in their role as enablers of research.

- *Preservation:* Long-term preservation is a core necessity for data re-use. Preservation includes multiple services, including back-up/safe-store copies, security, format migration, and media migration. While most repositories have backup storage and migration strategies for the short-term, development of a comprehensive preservation policy is difficult for smaller repositories. Few repositories have the staff resources or expertise to implement audit and certification standards. Access to shared infrastructure and/or shared access to technical expertise in long-term preservation are urgently needed.
- *Curation:* The traditional meaning of "curation" is to select and document the elements of a collection including provenance: the history of how data were created and transformed. For digital collections curation also includes defining and populating metadata that are used in supporting search and discovery capabilities, and in making the data understandable for re-use. To maximize efficiency, domain repositories are ideally placed to design systems that allow researchers to prepare data for archiving as it is collected and analyzed.[8]
- *Interoperability:* New discoveries can come from re-examination of a single data set, but more frequently are the result of comparison and integration across data sets. Such comparisons require interoperability, i.e., consistency of metadata definitions, use of standard data formats, and/or the provision of translators between formats. Domain repositories are ideally positioned to create interoperability and to build bridges across related disciplines.
- *Supporting reproducibility and integrity of the research process:* A fundamental tenet of scientific research is that it should be reproducible. Access to the data underlying research publications is therefore a fundamental requirement. While cases of deliberate falsification of research results are rare, access to data assures the integrity of the research process and upholds the legitimacy of the scientific enterprise.

[7] http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

[8] See, for example, a vision statement in a report from an NSF/Wenner-Gren supported workshop, May 18-20, 2009. http://anthrodatadpa.org/addpa/home/vision-statement

- *Citation and linking:* Data citation demonstrates the value and impact of the research funding, and can serve as a metric for funding agencies to understand how data are being used. Linking data to and from publications facilitates research and supports the transparency, reproducibility, and integrity.
- *Controlling access to proprietary/confidential data:* Ideally all publicly funded research data become openly available, perhaps after a reasonable period during which time the principal investigator has exclusive access to accomplish primary research objectives. In some fields, however, access controls are needed to protect the confidentiality of subjects (e.g., in medical research) or intellectual property of commercial entities.
- *Community engagement (data providers, data consumers):* Engaging a community is the best way to maximize the value of a repository. Repositories serve multiple communities starting with domain scientists and expanding to students, policy makers, and sometimes the general public. Since new uses are often unanticipated—and of high impact—data must be made widely available.
- *Big data:* As new forms of "Big Data" emerge, data sets are becoming too large to to download, store, or analyze on a desktop computer. Analysis requires moving the algorithm to the data. Domain repositories will increasingly be called upon to manage peta-scale collections, implying infrastructure costs for storage and computational services and management of access for remote users.

3. Infrastructure Issues

- *Preservation, storage, and migration:* Most domain repositories have back-up and storage solutions that include redundant off-site storage, but storage and backup systems in themselves do not constitute long-term preservation of digital data. Digital repositories must have persistent ways of describing, discovering, accessing, and assuring the integrity of the objects in their charge. Metadata schemas such as PREMIS (PREservation Metadata: Implementation Strategies) are necessary to discover, search, and access databases as hardware, software, and community standards change.[9] Repositories must have the administrative and technical capacity to manage digital preservation and to migrate to new hardware and software standards as they emerge. In the present climate, domain repositories have difficulty supporting technical systems and staff expertise required for long-term preservation.
- *Certification:* Several standards for repository certification are now available, including the World Data System (WDS) and Data Seal of Approval (DSA). Certification could serve as a way to communicate the added value of domain repositories. It could also foster trust among stakeholders (e.g. funders, journals) that repositories are institutionally and financially sustainable, and encourage data producers to deposit

[9] David Gewirtz (Chair), Laura Welcher, Dean Snow, Michael Fischer, David R. Hunt, and Mark Mahoney, "Storage/Backup and Long-Term Preservation Breakout Group Report" of an NSF/Wenner-Gren supported workshop, May 18-20, 2009-- http://anthrodatadpa.org/addpa/chair-reports/storagebackup-issues; Angela A. Dappert and Markus Enders, excerpts from "Digital preservation metadata standards" Information Standards Quarterly 22 (2010). http://www.loc.gov/standards/premis/FE_Dappert_Enders_MetadataStds_isqv22no2.pdf

data in high-quality archives. Going through the certification process can help a repository to improve its practices and learn about itself. On the other hand, certification can be very labor intensive for archives already stretched thin. The value of certification is not always clear, and such a process is not a substitute for developing trust among data producers through community engagement. Moreover, particular domains may have valuable systems set up that clash with the expectations of certification—for instance, in oceanography researchers are required to put their data in certain repositories and certification would require the repositories to reject more of this data.

- *Workforce:* There is a shortage of qualified people to work at repositories, particularly in developing data management systems, defining data models, and programming. Repositories need a balance between people who understand curation, the technology, and the science. Moreover, there is a lot of workforce turnover, so substantial effort is going into training and retraining; once people are trained they become desirable in the market and there is competition for them. This leads to a lack of continuity and a loss of institutional knowledge. In terms of funding, there is a mismatch between budget levels and the salaries required to retain good people. There is also a mismatch between job classifications like archivist and librarian, and the actual demands of the data world. The implication is that repositories cannot operate effectively on soft money; there needs to be an underlying sustainable funding base.
- *Institutional repositories:* Institutional repositories (IRs) can be beneficial in that they are permanent; libraries have invested in them and they are established institutions, so the data they hold can be expected to be supported long-term. They can also get data that would otherwise be lost. IRs sometimes collaborate with research teams and are a source for gathering information during active creation. The tremendous diversity of data held in IRs means that metadata is not as rich as in a domain repository, thus compromising discoverability and interoperability, and most IRs lack the capacity to migrate data to new formats as software changes. Data curation experience, expertise, and capacity at IRs are also limited.
- *Commercial providers, Cloud:* It would seem to be attractive to utilize commercial cloud-based storage as the underlying infrastructure for domain repositories. Several cost studies suggest that it is rarely cost-effective, the exception being static collections with modest I/O requirements.[10]

4. Funding Models

Domain repositories operate with a number of different funding models, including hybrid approaches that help to increase the dependability and stability of their income stream. However, domain repository leaders agree that the current solutions are inadequate for ensuring that research data is preserved and accessible for the long-term; new, stable, long-term solutions are needed.

[10] Berriman, Deelman, Juve, et al., 2010. http://arxiv.org/abs/1010.4813

After discussing ideal funding model principles, we describe some of the current models used by domain repositories. We then compare how well the present models fit many of the ideal principles.

Principles

- Research data are a public good,[11] therefore it is necessary to facilitate equal access and equal opportunity to data services.
- Science requires a permanent and durable record representing the richness, diversity, and complexity of the scientific enterprise.
- Domain repositories provide essential domain expertise and responsiveness to a scientific community. Domain repositories also help to assure efficient allocation of expertise to curation and can foster interoperability across related domains.
- Preparing data for archiving during the research process (active curation) is much more efficient than preparing legacy data for archiving.
- Sufficient, flexible, long-term, and uninterrupted support must be ensured.
- Effective partnerships with the international community and the private sector should be sought and encouraged.
- Sources of revenue and the nature of expenditures should be transparent and fully accountable.

Current Models for Funding Repositories

Membership: A repository is supported by member institutions, such as universities and libraries, who pay dues and fees to support repository activities. Usually some rights, particularly access, are restricted to members.

Submission fees: The author or a sponsor pays the repository for preparing and archiving deposited data. Submission fees may be coordinated with publication of results in a journal.

Institutional support: Some universities support repositories for specific disciplines with institutional funds. Many universities have created institutional repositories, which are intended to service all scientific domains.

Federal Funding for Special Projects: Most domain repositories receive a large part of their funding in the form of grants from federal agencies or private foundations. These grants have limited duration, but some repositories have renewed grants for decades. Examples include NASA's space physics, astrophysics, and planetary data archives, which have been supported at reasonably steady levels for 25 years or more. ICPSR hosts archives for a number of federal agencies and foundations, including sponsors in NIH and the Department of Justice. The Protein Data Bank receives grant funds from seven federal sponsors.

[11] "Public good" in the economics sense, i.e., "A product that one individual can consume without reducing its availability to another individual and from which no one is excluded." http://www.investopedia.com/terms/p/public-good.asp (August 2013).

Possible New Models for Funding Repositories

Commercial services: Repositories may build services to make money from archived data. This could have limited use for academic users, or only be available to corporate consumers.

User fees: Some kinds of data are costly to distribute, and end users may be charged for costs associated with data access. For example, when datasets are extremely large, it is more efficient and economical to provide computing on central resources. Similarly, there are costs to providing access to data in which confidential information from research subjects must be protected.[12] Repositories may charge fees for services provided directly to data users.

Overhead: Universities allocate a percentage of all research grants (possibly from indirect costs) towards a fund for data archiving. Universities then make decisions about how data are selected for archiving and which repositories are used.

Infrastructure: Funding agencies pay for archives directly as a necessary aspect of research infrastructure. The funding model is structured for long-term investment, rather than being tied to three-year grant cycles. While this may appear similar to federal funding for special projects, *a percentage of federal research funding would be set aside for digital data archiving and preservation in all disciplines*.

Advantages and Disadvantages of Different Models

The chart below reviews most of the models discussed above in terms of some of the most important principles of economic dependability needed for long-term sustainability and equity. The infrastructure model does not yet exist, but we include it because in the view of this group it is the only one that maximizes equity and economic dependability.

Funding Models	Potential for Economic Stability Needed for Long-Term Sustainability	Potential for Open Access to Research Data	Potential for Equity for Deposits by Individual Researchers	Potential for Equity for Universities/ Institutions
Membership Dues	Moderate; subject to institutional budgets and priorities	Low	Moderate	Low
Submission Fees	Low to Moderate; subject to policies of funding agencies and publications;	High	Low; costs transferred from end users to data producers	Low
Institutional support	Moderate; subject to institutional budgets and priorities	High	Low	Low
Federally-sponsored Special Projects	Low; subject to changes in national research priori-	High	Limited to designated research	High

[12] NSF policy allows investigators to recover the "incremental costs" of sharing data.

	ties			
Commercial services	Low; at risk of commercial and sponsored competition	Moderate	Low	Low
User fees	Low; unlikely to cover costs of data curation and preservation	Low	High; costs transferred from data producers to end users	Low
Overhead	Moderate; subject to changes in national research priorities and institutional commitment	High	Moderate	Low; favors research at well-funded universities
Infrastructure	Moderate to High; subject to political commitment	High	High	High

5. Recommendations

We have come together from a wide range of disciplines to assert the importance of data archiving and the essential role that domain repositories play in curating data for future re-use. We recognize that materials and methods of science vary by domain, as do the sources and modes of research funding. A single funding model may not fit all disciplines, but new approaches are urgently needed.

Data repositories have been heavily dependent on two models, which have clear disadvantages: grants and memberships. Research grant competitions subject data repositories to review criteria and time horizons that are inconsistent with their core function as long-lived memory institutions. Membership models do not provide public access to data, and they favor researchers at institutions with more resources.

We propose the following principles to encourage data stewardship and support sustainable data repositories

1. Commit to sustaining institutions that assure the long-term preservation and viability of research data.

Agencies supporting research must back up the new open access requirements with funding to ensure their success. Overall, the funding model that provides the highest level of stability, best access for both ingest and retrieval, and greatest equity amongst organizations, is the infrastructure model. The percentage of the total research budget needed to support this approach is likely to be domain specific. We estimate that successful domain repositories can be operated at funding levels of less than 5% of the total research budget (Some fields might be as low as 1%; the cost might rise to 10% in fields with high data rates or particularly diverse and complex metadata.) These are modest costs to assure a strong return on public investments in the research and to enable uses of data unanticipated by the original investigators.

2. Promote cooperation among funding agencies, universities, domain repositories, journals and other stakeholders.

Archiving and preserving scientific data for re-use will require contributions from all participants in the research enterprise. Funding agencies should re-direct resources to support data curation and archiving. Universities should provide facilities for researchers and assure compliance with archiving requirements. Domain repositories should establish partnerships with universities and institutional repositories to provide facilities and expertise across all disciplines. Professional associations should revise their codes of ethics to affirm the value of research transparency and data access. Journals should require that authors provide access to the data used in their publications. In many disciplines this will involve a cultural shift and innovations in scientific workflows designed to capture and document both data and research methods.

3. Support the human and organizational infrastructure for data stewardship as well as the hardware.

Funding for data stewardship must include trained professionals, organizations with the capacity to persist over time, and community standards for metadata and preservation.

4. Establish review criteria appropriate for data repositories.

Data repositories should be evaluated by criteria that are consistent with their mission as institutions entrusted with the long-term preservation of scientific memory.

Relevant criteria are:

- service to the community
- adherence to and development of standards for metadata and data curation
- compliance with standards for trusted digital repositories

5. Incentivize PIs to archive data.

Funding policies should reward PIs for good data management practices and data sharing. PIs should not be faced with a tradeoff between accomplishing their scientific objectives and sharing their data. For example, data curation for archiving should be funded by mechanisms that are outside the funding allocated for the principle aims of a research project.

Acknowledgments

We are grateful to the Alfred P. Sloan Foundation for its support of the workshop on "Sustaining Domain Repositories for Digital Data." The Inter-university Consortium for Political and Social Research (ICPSR) convened the workshop and continues to support this initiative. Consultants from Cambridge Concord Associates provided invaluable advice and assistance before, during, and after the workshop.

Appendix

- Workshop agenda
- Endorsers of the "Call for Change"

Sustaining Domain Repositories for Digital Data

June 24-25, 2013
Kuenzel Room, Michigan Union
Convened by ICPSR at the University of Michigan
Supported by a Grant from the Alfred P. Sloan Foundation

Day 1: June 24

8:30 Breakfast

9:00 Welcome & Opening Comments George Alter, Director, ICPSR
University of Michigan

Introductions & Agenda Overview Elaine Kuttner, Principal
Cambridge Concord Associates

9:15 Sharing Repository Experiences

10:30 Break

11:00 Panel Discussion: Infrastructure at the National and International Levels

Robert Chen (CIESIN, Columbia University), CODATA and the International Council for Science (ICSU)

Paul Uhlir (National Academies), Research Data Alliance

Jared Lyle and Mary Vardigan (ICPSR, University of Michigan), National Digital Stewardship Alliance and Data Seal of Approval

Chair: Sayeed Choudhury (Data Conservancy, Johns Hopkins University)

12:15 Lunch

1:15 Small Group Discussions: Repository Operations

2:15 Small Group Discussions: Domain Repositories and the Academic Ecosystem

3:30 Break

3:45 Panel Discussion: Funding Models

George Alter (University of Michigan), ICPSR

Helen Berman (Rutgers University), Worldwide Protein Data Bank

Todd Vision (University of North Carolina), Dryad

5:00 Adjourn

Day 2: June 25

8:00 Breakfast

8:30 Working Groups: Drafting a Common Statement

11:00 Moving Forward

12:00 Adjourn

"Sustaining Domain Repositories for Digital Data: A Call for Change from an Interdisciplinary Working Group of Domain Repositories" has been endorsed by:

Karen Adolph*, Databrary Project, New York University

George Alter*, Inter-university Consortium for Political and Social Research, University of Michigan

Helen Berman*, Research Collaboratory for Structural Bioinformatics Protein Data Bank, Rutgers University

Bobray Bordelon*, Cultural Policy & the Arts National Data Archive, Princeton University

Thomas M. Carsey, HW Odum Institute for Research in Social Science, University of North Carolina

Robert S. Chen*, Center for International Earth Science Information Network, Columbia University

Sayeed Choudhury*, Principal Investigator of the Data Conservancy

Christopher Cieri*, Linguistic Data Consortium, University of Pennsylvania

Jonathan Crabtree*, HW Odum Institute for Research in Social Science, University of North Carolina

Mercè Crosas, Dataverse, Director of Data Science at IQSS, Harvard University

Ruth E. Duerr*, National Snow and Ice Data Center, University of Colorado

Colin Elman, Qualitative Data Repository, Syracuse University

Carol R. Ember*, Human Relations Area Files, Yale University

Florence Fetterer, Manager, NOAA@NSIDC,National Snow and Ice Data Center

Roger Finke, Association of Religion Data Archives, Pennsylvania State University

Rick O. Gilmore*, Databrary Project, The Pennsylvania State University

Robert J. Hanisch*, Virtual Astronomical Observatory, Space Telescope Science Institute

Margaret Hedstrom*, SEAD DataNet and School of Information, University of Michigan

Paul Herrnson*, Roper Center, University of Connecticut

Diana Kapiszewski, Qualitative Data Repository, Georgetown University

Gary King, Albert J. Weatherhead III University Professor and Director for IQSS, Harvard University

Eugene Kolker, MOPED Database, Seattle Children's Research Institute & DELSA Global

Kerstin Lehnert, Integrated Earth Data Applications, Columbia University

Francis P. McManamon*, Executive Director, Center for Digital Antiquity, Arizona State University

William Michener, DataONE and Professor and Director of e-Science Program, University Libraries, University of New Mexico

Steven Ruggles*, TerraPopulus and Integrated Public Use Microdata Series, University of Minnesota

Mark C. Serreze, National Snow and Ice Data Center, University of Colorado

Libbie Stephenson*, UCLA Social Science Data Archive, University of California, Los Angeles

Victoria Stodden, RunMyCode, Columbia University

Alexander Szalay*, Virtual Astronomical Observatory, Johns Hopkins University

Todd Vision*, Dryad Digital Repository, National Evolutionary Synthesis Center

* Participant in the June 24-25 workshop.