Sustaining Domain Repositories for Digital Data:
A Call for Change from an Interdisciplinary Working Group of Domain Repositories

June 24–25, 2013
Interuniversity Consortium for Political and Social Research (ICPSR)
With support from the Alfred P. Sloan Foundation
Compiled by Cambridge Concord Associates

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 funding data collection to create plans for public access to research projects. 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, whether it 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 other 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** between scientific communities and digital libraries and archives to implement the latest developments in information science

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 commitments 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.

Another concern is that the push towards open access, while creating more equity of access for the community of users, creates more of a burden for domain repositories because it narrows their funding possibilities. Without care, this 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.*

This statement is 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

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