

# Collaborative Data Sharing Networks



Research and even business is becoming a collaborative enterprise that brings together multiple institutions, sectors and, increasingly, different countries. No where is this more apparent than in the natural sciences where the phenomenon being examined and questions being asked are not contained in the borders of one discipline, institution, country, or continent. Both a reason for and often the purpose of collaboration in the sciences is the need to amass, maintain, and share large and diverse structured data resources that no one research team or institution has the resources or expertise to collect, make available, and maintain.

Such data-centric collaborations among researchers are providing profound and valuable benefits to the scientific enterprise and the general public, including:

- **Enriching scientific knowledge** and accelerating scientific progress by encouraging researchers to generate new knowledge through using archival datasets in new ways and improving the quality and usefulness of existing datasets.
- **Fostering collaborative works** among researchers through the sharing of research datasets as well as materials, skills, and knowledge, and thereby increasing the quality of research.
- **Improving accountability** by encouraging a new ethos of open science and peer review that can increase accountability and reduce fraud related to data falsification and fabrication.
- **Increasing efficiency of research** effort through reducing the cost and time spent in collecting data and avoiding redundant data collection.
- **Expanding reputation and scientific merit** through increasing researchers' recognition and visibility by journals and peer committees, and, organizationally, improving data quality and efficiency and fostering trusted relationships among participating institutions.
- **Encourage long term data preservation and integrity** by reducing the redundancy and duplication in data processing, maintenance, and protection thereby reducing the cost and increasing the likelihood of long term data preservation.

Given the benefits of data-centric collaboration and sharing in the sciences, it is not surprising that organizational structures to facilitate this activity through the use of information technology are emerging. One such structure, called a collaborative data sharing network (CDSN), is being used to facilitate collaborations among dataset producers and users resulting in successful sharing of data and knowledge across traditional disciplinary, organizational, geographical, and political boundaries.

### FIVE CHARACTERISTICS OF A COLLABORATIVE DATA SHARING NETWORK (CDSN)

1. Collaboration of heterogeneous, autonomous, geographically dispersed, and inter-organizational social actors.
2. Members share common and compatible goals, including similar or different data and information.
3. Information may flow one-way, or the flow may be bi-directional.
4. Collaboration is mediated and dynamic within a trusted network.
5. Collaboration is supported with an interoperable infrastructure.

**AN EXAMPLE OF A CDSN** A prime example of a CDSN is DataONE, a collaborative earth observational data sharing networks initiative supported by the National Science Foundation. DataONE is taking advantage of

information and communication technologies to share data in a broader fashion than has been attempted in the past. It aims to ensure the preservation of, and access to, multi-scale, multi-discipline, and multi-national science data. DataONE is designed to transcend boundaries not only related to the field domains (e.g. biological and environmental), but also across organizational boundaries and, in the future, across national boundaries.

**DataONE**

Search this site

[f](#) [t](#) [in](#) [RSS](#)

**Wood Thrush**  
*Hylocichla mustelina*

Contributing to Understanding of Bird Species Distribution

**HELP WANTED**

Apply now for a  
DataONE Summer  
Internship

Latest News | DataONE co-sponsored Workshop on Software Tools for Sensor Networks Posted: 03/7/2012 DataONE 2012 Summer Internship Program

**About**  
What is DataONE?  
DataONE Organization  
Working Groups  
Partners  
Communication  
Videos  
Contact Us

**Participate**  
DataONE Users Group  
Member Nodes  
Internships  
Developer's Resources  
Open Positions  
Events Calendar

**Products**  
Investigator Toolkit  
Data Management Planning  
Best Practices  
Software Tools Catalog  
Publications

**Education**  
Training Activities  
Education Modules  
Graduate Courses

**ONE Mercury**  
In Testing

DataONE is a collaboration among many partner organizations, and is funded by the US National Science Foundation (NSF) under a Cooperative Agreement.  
1312 Basehart Dr SE 1 University of New Mexico Albuquerque, NM 87106 [Contact Us](#) | [Site Map](#)

A collaborative network such as DataONE consists of various members with various capabilities and resources. Its proposed participants range from individual field research stations to governmental organization (e.g., USGS, NASA, EPA). DataONE classifies these participants into users and nodes based on the level of services and fees for participating. Users are participants who will have capability to access and store datasets with no fees and nodes are the institutional-based participants who, upon joining DataONE, will have the ability to store, distribute, and coordinate datasets. DataONE itself will act as coordinating nodes that will mediate and direct the information flows and manage the connection between different member nodes. These diverse participants have different capabilities in terms of knowledge, experience, and resources. DataONE aims to connect multiple data repositories, collected and preserved by various organizations without regard to size and location.

**CHALLENGES** Notwithstanding the many benefits of data sharing, CDSNs such as DataONE face the same challenges of most data sharing initiatives. These challenges are embedded in social, legal, economic, and political factors and fall into four broad categories: technological, organizational, legal and policy barriers, and local context.

**Technological barriers** to data sharing exist when data sharing entities do not have compatible data architectures and technological infrastructures or consistent data definitions and standards. Data with different formats, definitions, content, and from multiple sources are difficult and costly to integrate into a single useable data repository or to improve so they are suitable for sharing.

**Social, organizational, and economic barriers** such as structural conflicts, managerial practices, lack of funding, institutionalized disincentives, and professional cultures can discourage data sharing. The intense

competition in scientific fields may, for example, contribute to resistance to sharing data. Research about scientific data sharing has shown that fear for reputational damage if data is found to be faulty or lacking in some way is a deterrent to data sharing. Another deterrent is the lack of relevant resources to prepare data for sharing and to sustain sharing mechanisms. Scientists and institutions are not often recognized or rewarded for making datasets openly available and usually can't spare the time or resources to prepare the labor-intensive documentation necessary to share data. Arranging for outside access and storage may involve lengthy and onerous negotiations or drawn-out administrative processes.

**Legal and policy frameworks** created by government, funding agencies, or other regulatory bodies often complicate the process of data sharing. Legal and policy mechanisms can create a paradoxical situation in relation to data sharing and may be the greatest obstacle in building a knowledge network. On the one hand, such frameworks can enhance data sharing by ensuring proper and accountable use of data and information as well as mandating the sharing of data. On the other hand, rigidity of policies and regulations, such as those designed to address privacy concerns, served to protect, can often inhibit data sharing. Unresolved legal issues have been found to deter or restrain collaboration, even if the scientists or institutions are prepared to proceed.

**Local context**, in the case of DataONE and the natural sciences, can create unique challenges to data sharing. Datasets in ecological research are complex, heterogeneous, and highly context dependent. Natural scientists usually pursue a specific question about a specific phenomenon at a specific site. Each subject might have different characteristics and require a different methodology. Data quality is highly correlated with the context underlying production, storing, and initially intended use. Using a diversity of data from multiple sources and contexts may lead scientists to question the data's reliability and its research value or usability.

**ORGANIZATIONAL SUPPORT & COMMITMENT FOR CDSN** Organizational support plays a major role in sharing research datasets, particularly considering the heterogeneity of collaborators and complexity of the data sharing process (Sayogo & Pardo, 2011; 2012). Analysis using a logistic regression and structural equation modeling technique of survey responses from 587 researchers found that organizational involvement is crucial for two reasons:

- Providing support for data management.
- Reducing the burden of complex data sharing process for the researcher.

The study also found that organizational support significantly influences the intention of researchers to publish their datasets.

**CRITICAL CAPABILITIES** The success of CDSNs such as DataONE depends on the ability of many, if not most, of the participating entities to overcome the challenges described above. Success of a CSDN then requires new understanding of data sharing and calls attention to the following questions:

- What kinds of capabilities are needed to effectively participate in cross-boundary scientific data sharing?
- Given the variations in the capabilities of scientific data stakeholders, what factors are critical to the success of data integration and reuse in a scientific CSDN?

It is precisely these types of questions that CTG and others are trying to answer for DataONE and similar types of CDSNs. Through previous research, CTG modeled the complexity of data sharing initiatives including the interdependencies of technical and organizational capabilities and the relationship between those capabilities and successful data sharing. Building on this past research and new data on DataONE, four categories of capabilities continue to stand apart as critical to the success of a data sharing initiative.

1. **Collaborative management capabilities** include strategic planning, organizational compatibility, and resource management. These capabilities are necessary for mobilizing the resources and building the organizational structures necessary to participate in a CSDN. Assessments of this set of capabilities prior to entering could be used to decide the level of participation appropriate for each member node and services users could reasonably expect from that node.
2. **Data governance and policy capabilities** include data assets requirements, governance, information policies, and secure environment. These refer to the ability of an entity, in this case an institution considering becoming a member of a scientific CSDN, to provide and encourage sharing through wide-ranging, clear, and precise information policies and management practices including policies on data stewardship, use, and security. This requires also the governance of data collection, description, usage, sharing, reuse and long term preservation. These capabilities are critical to supporting open sharing of research datasets, particularly to mitigate the fear of data misuse and misinterpretation. Intellectual property has been found to be a major concern in sharing ecological research datasets. Scientists are wary of the issue of recognition

for data ownership.

3. **Collaborative space and operational agreements** that address all the elements necessary to collaborate are critical for a collaborative network. These elements include not only the infrastructure but also other elements essential for fostering collaboration and managing interdependencies among stakeholders such as effective communication procedures, working principles, and operational protocols. This capability is essential to sustaining the collaboration. Collaboration-ready entities are entities with successful collaboration experience who actively seek new opportunities for partnering. They are entities with the negotiation skills and experience necessary to achieve agreement, compromise, and mutual understandings on the distribution of authority and responsibilities within a cooperative network.
4. **Technology capability** includes technology acceptance, technology knowledge, and technology compatibility. Technology acceptance refers to the attitudes of entities toward technological change and their degree of comfort in accepting the new technology. Previous experience with technology often results in a more receptive attitude toward technology-based data sharing initiatives. Technology compatibility includes the presence of agreed-upon standards, interconnectivity among entities, and a staff experienced in sharing activities.

**FACTORS FOR SUCCESS** CTG's extensive work in cross-boundary information sharing and collaboration has consistently identified three factors as critical to the success of cross-boundary data sharing initiatives. Preliminary insights from scientific data sharing initiatives support these findings:

- **High-level of Trust.** Collaboration requires peer relationships between actors where trustworthiness is the most prominent ingredient. Hierarchical mechanisms do not exist in the governance of collaborative networks. The participating entities are autonomous and heterogeneous. Therefore, it is necessary to have common working principles, value systems, policies, and set of base trustworthiness criteria. Trust and trustworthiness are important determinants in ensuring a successful data sharing initiative.
- **Common Working Principles, Values, Policies, and Organizational Commitment.** Incentivizing participants to continuously participate in a CSDN is essential to its success. Research shows that previous efforts in biology networks and collaborative databases have failed, in large part because of minimum contribution by members. For example, in one collaborative network in biology, 70% of the data contributions came from the founding members while only 30% from other contributors. A successful data sharing initiative also depends on well-documented procedures for collaboration and sharing of different resources, assessment of collaboration readiness, and measurement of the alignment between the value systems, principles, and policies.
- **Harmonization of Multiple Contexts.** Lastly, a data sharing initiative that transcends geographical boundaries must also deal with harmonizing different external contextual factors. Participants from different geographical regions may speak different languages, uses different scientific notations and laws/regulation, and be concerned with different cultural issues. Acknowledging these differences and developing mechanisms to work with and within them is a prerequisite for a successful data sharing initiative.

**CONCLUSION** The sharing of research datasets is recognized as central to global efforts to advance science. Ensuring the success of sharing however, is a difficult and challenging endeavor that goes beyond a single knowledge domain, organization, or nation. Encouraging the sharing of datasets in a collaborative network in the interest of advancing science requires balancing expected benefits with identified challenges. If data sharing is conducted within a collaborative network such as DataONE, where the actors are autonomous, heterogeneous, and geographically dispersed, sharing is not purely based on personal decision but also affected by social and institutional arrangements. Looking at scientific data sharing through the lenses of CTG's work in information sharing and collaboration provides new insights into how capability for data sharing in the scientific community can be created and advances in science enabled.

Djoko Sigit Sayogo, Graduate Assistant Theresa Pardo, Director Alan Kowlowitz, Senior Fellow