



THE CENTER FOR  
**OPEN DATA ENTERPRISE**

**Open Data Roundtable on Sharing and Applying Research Data:  
KEY TAKEAWAYS**

*In 2016, the White House Office of Science and Technology Policy and the Center for Open Data Enterprise co-hosted four Open Data Roundtables to identify case studies, lessons learned, and best practices in open data across the federal government. Open data from government is free, publicly-available data that anyone can use and republish. The Center has summarized key takeaways from these Roundtables, which brought together experts from inside and outside of government with technical, policy, and legal backgrounds. The Center will publish a full report of Roundtable findings in fall 2016.*

**BACKGROUND**

On May 25th, 2016, the White House and the Center for Open Data Enterprise co-hosted a Roundtable to address a key issue: **How to best share and apply government-funded research data**. Open science, an approach that gives greater access to publicly funded research, results in a wide range of benefits. However, researchers face many challenges in sharing data, including the need for data platforms, concerns about privacy, confidentiality, and intellectual property, and incentives that discourage data sharing.

This Roundtable brought together 90 participants from government, nonprofits, academia, and the private sector to address research data. Roundtable participants were not asked to develop consensus recommendations but to share their own observations and suggestions.

**ISSUES TO ADDRESS AND PROPOSED SOLUTIONS BY RESEARCH DOMAIN**

For discussions during breakout sessions, participants were grouped into tables by research domains: Arctic; Cancer & Biomedical; Climate; Health & MyData; Infectious Disease; Materials Science; Smart Cities, Transportation & Energy; and Soil & Agriculture. The following sections describe the challenges and proposed strategies for improving open data in each research domain.

**ARCTIC**

The table was comprised of participants with expertise in the Arctic region and an interest in using satellite, energy, and biodiversity data to advance community sustainability and resilience. See: [Arctic Executive Steering Committee](#).

Arctic researchers are studying a region experiencing rapid climate-related changes, making it critical to capture, use, and share data in real time. Their work involves collaborating with an unusually diverse range

of stakeholders. These range from local tribes, who play an important role in data gathering, to international governments whose cooperation is essential. The challenges for Arctic researchers include working with each group diplomatically and coordinating data and analyses from different sources.

Arctic experts at the Roundtable proposed the following solutions:

- Ensure that outreach to tribal leaders and research engagements are coordinated across agencies.
- Develop an initial international agreement that advances open data in Arctic science. This agreement would set new priorities to increase resources for data repositories, information management, and innovative tools. International data-sharing would help create the immense datasets needed for Arctic research.
- Share data across institutions and initiatives to validate data by comparing multiple datasets and reduce any redundancies in data collection. This is particularly important for Arctic data because it changes so rapidly.
- Fund cloud-based data management tools to match the pace of data collection. Develop an open engagement platform for the Arctic, with tools that anyone can use.

## **CANCER AND BIOMEDICAL RESEARCH**

The table was comprised of participants with expertise in cancer and biomedical research with an interest in using data to advance prevention, screening, diagnosis, and treatment of cancer. See: [Cancer Moonshot Initiative](#).

The Cancer Moonshot Initiative promises to boost the production, use, and sharing of cancer-related research data. While this ambitious program is well funded, it will face many challenges in developing effective ways to share the research data essential to its progress. Research groups use different formats, structures, processes, and standards in collecting and managing their data, making it difficult to synthesize their work.

Participants proposed developing a single data repository to integrate existing data platforms relevant to cancer research and the Cancer Moonshot. This “Cancer Data Constellation” would leverage the value of diverse datasets and accelerate research. It would use common standards and persistent identifiers to link research programs and researchers to the datasets they provide, which would both enable interoperability and make it possible to give credit to academic researchers who share their data.

## **CLIMATE**

The table was comprised of participants with expertise in climate data as it relates to weather, geography, oceanography, agriculture, etc. See: [U.S. Climate Resilience Toolkit](#).

Major challenges in managing and analyzing climate data include the size, scope, variety, and complexity of the data involved. The massive collections of satellite, weather, and other data used in climate science are simply too large for researchers to download them and analyze them on their own systems.

Climate data experts proposed establishing a paradigm of moving analytic services to the data through cloud hosting rather than downloading data for localized analysis. This strategy would help remove barriers to data use while also creating incentives for data sharing. The project would encourage researchers to move data to the cloud, begin to analyze data there, and develop new standards and approaches through that work.

## **HEALTH & MYDATA**

The table was comprised of participants with expertise in patient-centered initiatives and an interest in how research data affects patients. See: [MyData Initiative](#).

The personal nature of health data makes sharing it especially challenging. Concerns over privacy are not only ethical but legal. Researchers must comply with the Health Insurance Portability and Accountability Act (HIPAA) and the Family Educational Rights and Privacy Act (FERPA), which limit data sharing. Obtaining informed consent for data release can help solve privacy concerns in principle, but can be difficult to implement.

As one possible solution, Roundtable participants proposed ways to incentivize and enable patients to share their data voluntarily. Patients could share their health information through a data marketplace, assigning real value to individuals' data that all stakeholders recognize. At the same time, researchers could work with healthcare providers to communicate to patients about opportunities to share their data, the value of the data, and the impact it can have.

## **INFECTIOUS DISEASE**

The table was comprised of participants with expertise in combating infectious diseases and surveillance research data. See: [Zika Challenge](#).

Infectious disease researchers face a number of challenges due to the international nature of their work. High-income countries use information from low-income countries, which may be concerned about how data about disease will reflect on them. Some national governments have asserted "viral sovereignty," claiming that they have the right to control the research and data about pathogens within their borders, even if there is a risk that they will spread.

Proposed solutions include:

- Develop processes and agreements to share data during infectious outbreaks to help identify the focus of intervention, determine which interventions are working, and detect when pathogens change genetically.
- Work with the World Health Organization (WHO) to expand open data practices and capacities in low to middle income countries and facilitate discussions between countries.
- Use well-established organizations and institutions with data management capabilities to facilitate data-sharing. For example, the National Library of Medicine facilitates data reuse and allows for replication of studies. Open data shared by the Centers for Disease Control allows for assessing emerging infectious risks, possible interventions, quantitative evidence, and capacity for response.

## **MATERIALS SCIENCE**

The table was comprised of participants with expertise in sharing materials science research data in an effort to accelerate the pace of discovering, manufacturing, and deploying advanced materials. See: [Materials Genome Initiative](#).

Data challenges include a lack of data standards, poor metadata, and inconsistent indexing. Additionally, participants noted that there is no sustainable business model for a sharing infrastructure.

To tackle these issues, participants suggested adding data sharing plans into public governance models. Participants also discussed strengthening open data mandates by providing government resources and working to change the culture and status quo of research data.

## **SMART CITIES, TRANSPORTATION, AND ENERGY**

The table was comprised of participants with expertise in transportation and energy research data for improving local services. See: [Smart Cities Initiative](#).

Researchers working to develop smart cities face challenges specific to the use of localized data. Most state and local governments do not have the skills or infrastructure needed to manage research data. Additionally, some states have legal or copyright constraints on opening their data.

Participants developed a solution that would give cities the capacity to manage data more effectively by creating standardized data repositories. This would enable networks of researchers and practitioners to have greater impact with fewer resources. The repositories would consist of the following elements: access control, sharing aggregated data, sharing and bulk purchase, local access to global data, and links to training.

## **SOIL AND AGRICULTURE**

The table was comprised of participants with expertise in soil and agricultural data. See: [Global Open Data for Agriculture and Nutrition Initiative \(GODAN\)](#).

Participants noted that limited funding impacts the quality of data infrastructure, delivery, and maintenance. They also discussed privacy issues pertaining to soil data. Some researchers are required to remove location information before sharing data, which limits its value.

Solutions included:

- Develop strategies for creating public buy-in to encourage private landowners to share data.
- Build agency-based repositories for grantee-generated research and evaluation data, and making the data available to researchers.
- Streamline data publishing process to decrease barriers by developing a platform for curating data and providing agency assistance to principal investigators in storing data.

## **NEXT STEPS**

In addition to domain-specific solutions, these table discussions produced ideas for systemic changes to encourage data sharing in all research domains. Three ideas had broad support from participants:

**Improve data governance to encourage data-sharing.** Participants suggested the following approaches:

- Enforce existing federal data sharing mandates.
- Provide a template for data management plans for federal grantees, who are required to prepare these plans as a condition of funding.
- Develop a federal “data lake,” a repository for storing vast amounts of raw data in its native format. Individual databases could be made available through separate “buckets,” each with its own security policies.

**Develop incentives to promote research data sharing more widely.** Participants recommended these steps:

- Ensure that researchers are rewarded for releasing and sharing datasets by using citation systems similar to citations for published papers. Data citation systems could help researchers gain credit for their work and measure the impact of their research.
- Take a systems approach to incentivizing data-sharing. This could incentivize multiple actors to increase data sharing, and provide more open data equitably with systematic and mutually reinforcing incentives. The approach should include participation from stakeholders including international governments, science agencies, funders, and university administrators.
- Require data sharing as a condition of federal funding for research.
- Create a database of data-sharing success stories as models for academics and the private sector.
- Consider the use of open data shared by researchers in decisions about their tenure.

**Provide resources to help develop sustainable models for data-sharing infrastructure and data management.** Participants suggested the following strategies:

- Explore “Why can’t we do in common that which is commonly done?” For example, data collection and processing may be different for each agency, but we could find common means of data dissemination.
- Create a virtual commons for data sharing, with shared network infrastructure for access and a registry to find datasets.
- Build data infrastructure once for use across agencies instead of having each agency build its own.
- Develop agency data management plans using the National Institute of Standards and Technology as a model.
- Use GitHub repos for groups that need to host all data in a public, open space and track version control.

**ABOUT THE CENTER FOR OPEN DATA ENTERPRISE**

The Center for Open Data Enterprise is an independent nonprofit 501(c)(3) organization, based in Washington, DC, whose mission is to maximize the value of open government data as a public resource, by focusing on data users. We thank our Open Data Partner Microsoft and Open Data Supporter Booz Allen Hamilton for supporting the Center's work on the Open Data Roundtables. We welcome feedback on this report; please send comments and inquiries to Katherine Garcia, Director of Communications and Outreach, at [katherine@odenterprise.org](mailto:katherine@odenterprise.org).