



# The 2016 U.S. Open Data Roundtables:

Recommendations  
from Data Providers  
and Users



THE CENTER FOR  
**OPEN DATA ENTERPRISE**



THE CENTER FOR  
**OPEN DATA ENTERPRISE**



# Table of Contents

2

Executive Summary

4

Background

6

Methodology

8

Data Lifecycle  
Recommendations

34

Research Data  
Recommendations

60

Appendices

The Center for Open Data Enterprise thanks our Open Data Partner Microsoft and our Open Data Supporter Booz Allen Hamilton for supporting the Center's work on the Open Data Roundtables. We also thank the 290 open data experts from government, the private sector, nonprofit organizations, and academia who participated in the Roundtables and generously gave their time towards this effort.

# The 2016 U.S. Open Data Roundtables:

## Recommendations from Data Providers and Users

SEPTEMBER 2016

---

---

## Executive Summary

Open data from government is free, publicly available data that anyone can use and republish. This valuable public resource has helped citizens find better value for college, fair housing, and safer medicines. In addition, it helps government agencies operate more efficiently, share information, and engage the citizens they serve.

Although the value of open data is clear, much work still needs to be done to make open government data more accessible and usable. To develop ways to improve this valuable public resource in the United States, the White House Office of Science and Technology Policy (OSTP) and the Center for Open Data Enterprise co-hosted four Open Data Roundtables in 2016. The Open Data Roundtables focused on data privacy, data quality, sharing research data, and public-private collaboration. While each Roundtable addressed a particular topic, many common themes emerged throughout the series. This report organizes challenges and recommendations based on the relevant stages of the data lifecycle:

- Data collection: Gathering data to serve a government agency's or organization's mission
- Standardization: Developing common data definitions, formats, and metadata
- Managing privacy: Ensuring that personally identifiable information in datasets is not released to the public
- Data release: Publishing data in effective and appropriate ways
- Quality improvement: Making datasets more accurate, complete, and usable
- Curation, storage, and management: Managing datasets for maximum value
- Communities and collaboration: Building user communities and collaborations between government, the private sector, nonprofits, and academia

In addition to addressing these cross-cutting issues, several Roundtables included breakout sessions where experts from specific research domains discussed as a group. These research domains included Arctic, Cancer and Biomedical, Climate, Health, Infectious Disease, Materials Science, Oceans, Smart Cities, and Soil and Agriculture.

This report is a unique synthesis of insights from experts both inside and outside of government with a deep understanding of the challenges and opportunities in data use. It presents open data recommendations that the Center for Open Data Enterprise has synthesized from learned lessons at the four 2016 Open Data Roundtables and other research.

The report is divided into two sections: (1) Data Lifecycle Recommendations and (2) Research Data Recommendations. The first presents strategies and recommendations intended to be useful to a wide range of people working with open federal data. The second covers issues and strategies that are particular to select research areas. Throughout both sections, we have included case studies that provide examples of solutions to the data challenges described in this report.

This report is designed to be of use to government data providers, experts who collaborate with the government, and the individuals, businesses, and organizations that rely on open government data. This report's recommendations are largely addressed to federal agencies, but they contain a number of ideas that could be implemented by other stakeholders or through public-private collaboration. We encourage anyone concerned with the use of open data to consider ways to participate in improving government data resources and putting them to use.

# Background

The Obama Administration has made strong commitments to improving open government data and promoting its use. The administration has emphasized the use of open data as a public resource. While open data has long been seen as a critical tool for government accountability and transparency, this administration has also promoted open data as a resource for innovation, entrepreneurship, and improving public services. Businesses, nonprofits, governments and citizens are using open government data to launch new ventures, analyze trends, make data-driven decisions, and solve complex problems.

Since the Open Data Policy was established in May 2013, the federal government has established a Data Cabinet, a new Open Source policy, and a number of resources, programs, and initiatives supporting open data. At the same time, White House leadership has recognized that much needs to be done to make open government data more accessible and usable for wider use. The White House partnered with the Center for Open Data Enterprise to develop this Roundtable series as a rapid, inclusive way to address and help solve the challenges that are keeping open data from reaching its full potential. These Roundtables were designed to:

- Identify open data case studies, learned lessons, and best practices across the federal government;
- Strengthen a community of technical, legal, and policy experts in support of open data; and
- Support continuity and accelerate the progress of open data work.

## Over four months, the 2016 Roundtables addressed four open data challenges organized around the following questions:

**While the fourth** Roundtable was especially focused on collaboration, all the Roundtables underscored the importance of public-private collaboration and its potential to advance open data. DJ Patil, the U.S. Chief Data Scientist, has often said that “data science is a team sport.” For open data programs to succeed, they need to be supported by a community of stakeholders from business, academia, and civil society.

Throughout the Roundtable series, participants identified opportunities for collaboration on ways to manage and improve open data. These include technical contributions, such as data infrastructure and storage, search functionality, analytics and visualizations; user engagement through user interfaces, feedback channels, and outreach; and collaboration on developing data and metadata standards.

### Privacy [March 24]: How can we open granular information while protecting privacy?

As more open data has become available, data users have come up against a conundrum. Many datasets in health, education, housing, and other areas may have the most value when they are released with “microdata” that can be analyzed at the level of individual records. However, releasing data at that level carries the risk of exposing personally identifiable information (PII) that could threaten individuals’ privacy if it were released openly. Individual privacy should be treated in the context of public good, recognizing that many datasets with PII also contain information that provide great public benefit.

### Sharing Research Data [May 25]: How can we 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 a system of incentives that does not reward data sharing and may actually discourage it.

### Data Quality [April 27]: How can we improve data quality in efficient and scalable ways?

Organizations that want to use open government data face a number of obstacles as a result of quality issues with the data. Government agencies and their data users are now working to improve data quality by addressing issues such as timeliness, accuracy, precision, and interoperability.

### Public-Private Collaboration [June 15]: How can public-private collaboration support open data?

Many healthcare, agriculture, financial services, energy, and transportation companies, among others, use open government data as a key business resource. The private sector and government agencies have a mutual interest in helping to ensure that government data programs are high quality, easily accessible, and cost effective. In addition, open data stakeholders outside of government often have knowledge, expertise, resources, and processes that could benefit government data programs. Participants identified a number of specific opportunities for collaborations to help develop and disseminate open government data. The White House, federal agencies, and external data users are continuing to explore these opportunities.

# Methodology

The 2016 Open Data Roundtables were designed to take on four major issues in the use of open government data: privacy, quality, sharing research data, and public-private collaboration.

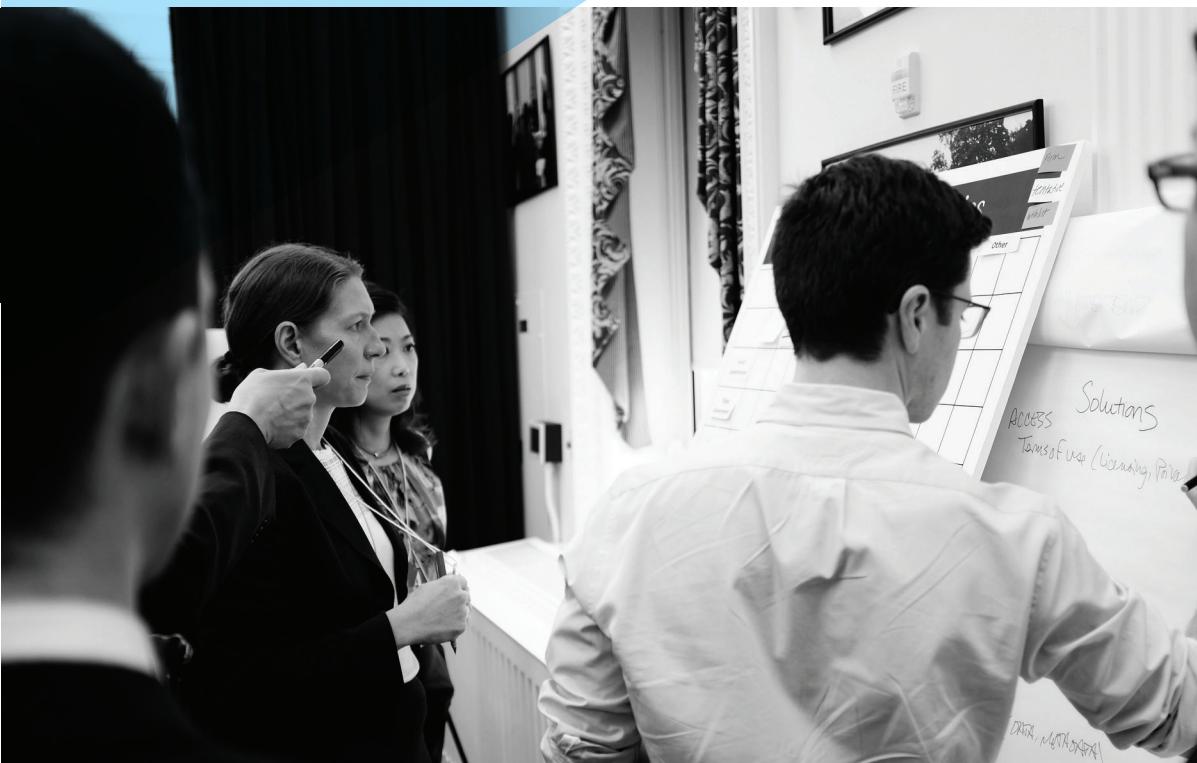
The Center for Open Data Enterprise used a multimethod approach in developing the recommendations described in this report, including desk research, the Open Data Roundtables themselves, solicitation of expert feedback, and interviews.

The sequence of work was as follows:

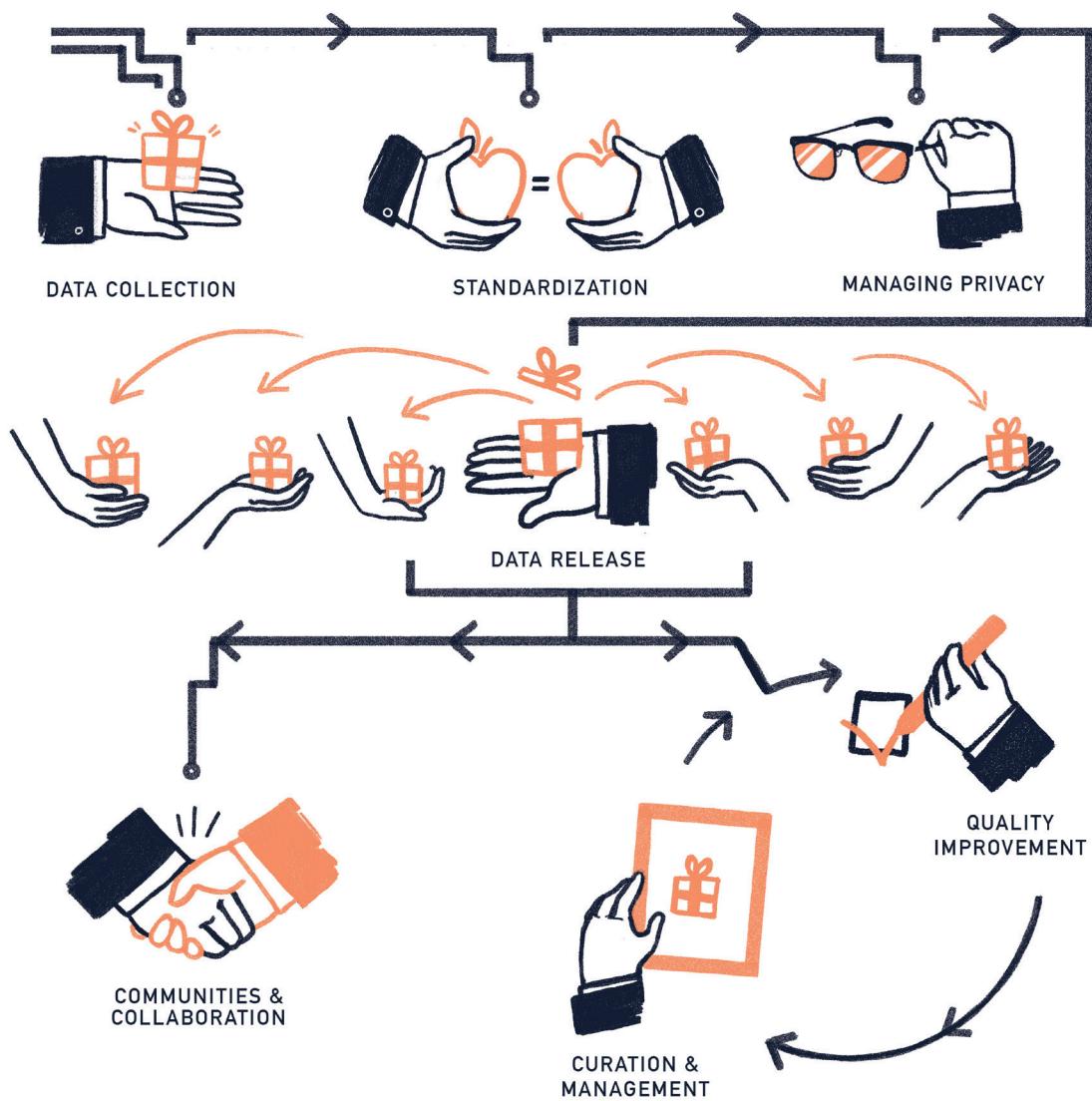
- 1. Review of existing literature to develop the initial framework for identifying the issues, solutions, and experts for each roundtable.**
- 2. Information collection through a public survey made available online. Questions assessed:**
  - Respondents' evaluation of the key challenges in each issue area
  - Effectiveness of current approaches used to address those challenges
  - Respondents' interest in participating in the Roundtable
- 3. Preparation of a Briefing Paper for background to each Open Data Roundtable, and participant review.**
- 4. Four Open Data Roundtables, held from March through June 2016, to address the issues.** These roundtables were facilitated discussions bringing together participants from federal agencies, academia, the private sector, and nonprofit organizations with technical, policy, and legal expertise. 290 experts participated. Notes were taken at each of these roundtable to capture everyone's input. Roundtable participants were not asked to develop consensus recommendations but to provide individual observations and suggestions.
- 5. Supplemental interviews and desk research.**

# Data science is a team sport.

- DJ Patil  
U.S. CHIEF DATA SCIENTIST



# Data Lifecycle Recommendations



Many data experts have stressed the importance of an integrated approach that addresses the entire data lifecycle. We have divided recommendations by the stages that are critical to making open data accessible and usable.

### **1. Data collection: Gathering data to serve a government agency's or organization's mission**

1. Focus on quality at the time of data collection, which is more efficient and effective than quality improvement at later stages.
2. Eliminate manual data entry as much as possible.
3. Use consumers and volunteers as data sources.

### **2. Standardization: Developing common data definitions, formats, and metadata**

1. Use common data standards and taxonomies for U.S. federal data.
2. Establish standards for international collaborations in different research domains.
3. In the absence of uniform standards, develop an additional "data layer" to enhance interoperability.

### **3. Managing Privacy: Ensuring that personally identifiable information (PII) in datasets is not released to the public**

1. Create customized privacy-protection programs based on risk assessment for each agency or program.
2. Consider a range of privacy protection strategies, and consider using them in combination.
3. Coordinate Disclosure Review Boards, Chief Privacy Officers, and other governance structures.
4. Build trust with the community around data use.

### **4. Data Release: Publishing data in effective and appropriate ways**

1. Release both raw data and improved data with transparency about quality and provenance.
2. Employ user-focused communication strategies to

encourage data dissemination and use.

3. Require data sharing and publication as a condition of research funding and help researchers meet that requirement.
4. Use new incentives to promote research data sharing more widely.
5. Develop collaborations and outreach to collect, manage, and publish data.

### **5. Quality Improvement: Making datasets more accurate, complete, and usable**

1. Develop user feedback systems to flag quality problems for agencies.
2. Use challenges and competitions to improve data quality.
3. Use crowdsourcing to improve data quality.

### **6. Curation, Storage, and Management: Managing datasets for maximum value**

1. Centralize and strengthen data governance at the agency level through Chief Data Officers.
2. Develop common tools, platforms, and catalogs for managing, sharing, and improving data.

### **7. Communities and Collaboration: Building user communities and collaborations between government, the private sector, nonprofits, and academia**

1. Use a range of approaches to identify user communities and measure data use.
2. Use public-private collaborations to find efficient, scalable solutions to data challenges.

# 1. Data Collection

The way data is initially collected can determine the quality and value of data throughout the lifecycle. Whether data is collected by satellites, sensors, or surveys, or through other means, quality control at the very beginning can prevent problems down the line.

## 1. Focus on quality at the time of data collection, which is more efficient and effective than quality improvement at later stages.

Organizations that want to use open government data face a number of obstacles in the quality of the data. Government agencies and their data users now see the need to address timeliness, accuracy, precision, interoperability, and other factors in open data. That effort will be directed most efficiently if it focuses on data at the time of collection. Quality safeguards can include:

- Standardizing data collection to avoid inconsistencies in data fields
- Systems to cross-check and validate data against existing datasets
- Formal quality control processes at the time of data collection

## 2. Eliminate manual data entry as much as possible.

Human error is a major cause of poor data quality. The move towards e-filing of government forms can help ensure better data. So can automating data-gathering through environmental sensors or other electronic means. Agencies are generally able to offer e-filing as an option for individuals and organizations that send them information. Although they may not be able to require it, incentives such as ease and speed of filing can encourage its use.

## 3. Use consumers and volunteers as data sources.

Individual citizens can help agencies collect data in different ways. Crowdsourcing programs for “citizen science,” which invite volunteers to collect data on the natural world or other phenomena, have been endorsed and encouraged by the White House Office of Science and Technology Policy. In a very different way, individuals can contribute to valuable datasets when they notify federal agencies about unfair business practices, unsafe products, or other consumer concerns. Several agencies, including the Consumer Financial Protection Bureau, have made consumer complaint databases open (with appropriate privacy protections) as a way to improve consumer markets.

## 2. Standardization

**Open data experts** stress the importance of developing and applying standards for data and metadata as early in the process as possible. Ideally, standards should be applied before data collection; converting to standards later on can lead to some loss of data quality. Better standardization can help improve data quality and in particular can help make different datasets interoperable. The Roundtables surfaced several suggestions for improving data collection and for applying data standards early on and throughout the data lifecycle.

### 1. Use common data standards and taxonomies for U.S. federal data.

Standards are needed to provide a basis for assessing data quality, for comparing datasets to each other to cross-validate them, and to make datasets interoperable. Different sectors in the federal government should approach standardization in the ways that are most appropriate to their data, in collaboration with academia, industry, and other outside experts. One example: The Departments of Justice, Homeland Security and Health and Human Services collaboratively created the National Information Exchange Model (NIEM) as a standard for data exchange. Organizations that join the NIEM community commit to following agreed upon terms when opening data.

### 2. Establish standards for international collaborations in different research domains.

In research domains including climate, Arctic research, and genomics, scientists must analyze very large quantities of complex data. Collecting and managing that amount of data requires international collaboration among many research groups. That collaboration, in turn, will be greatly facilitated by internationally accepted standards that both help ensure data quality and support interoperability.

### 3. In the absence of uniform standards, develop an additional “data layer” to enhance interoperability.

In areas where uniform data standards have not yet been established, it's possible to create a “data transformation layer” that provides meaningful information for developers to facilitate both the finding and the connecting of data within and across federal and private sector data producers and consumers. The CitySDK project, developed by the Census Bureau, is one example: It makes Census data available to developers so that they can create solutions for cities and communities using a simple open source toolkit. The CitySDK is that data transformation layer that is delivered through an Application Programming Interface (API) to help standardize the data mash-up or interoperability with a host of federal and private sector data. This kind of open source solution could be applied in many areas where different kinds of users need to discover, access and connect disparate standardized datasets.

## Case Study, NIEM:

A community-driven, standards-based approach to information exchange

The National Information Exchange Model (NIEM) brings together communities of people who have similar information needs but may speak different "languages". The NIEM website explains it this way:

» Maybe you drink "soda", but I drink "pop". The problem is—even though we're both talking about sweet fizzy beverages, our computers treat my "pop" as entirely different from your "soda".

This is where a standards-based approach to exchanging information comes in: You don't have to change your language. You can still call that fizzy beverage "soda" which is great for legacy systems with 30 years of data about "soda". But if you want to exchange information with me about my "pop", we'll each have to agree to call this fizzy beverage the same word, such as "soft drink".

NIEM has three "stewards" in the federal government: The Department of Justice, the Department of Homeland Security, and the Department of Health and Human Services. All 50 states and most Federal work with NIEM. When an organization buys into NIEM it agrees to a "data-dictionary of agreed-upon terms, definitions, relationships and formats -- independent of how information is stored in individual systems ..." This model can help produce data that is higher quality, machine readable, more interoperable, and real-time.

The NIEM is composed of a Core -- data elements that are commonly understood across domains -- and a number of Domains, made up of mission specific data that is managed through independent communities of interest. NIEM has continued to expand in scope, moving beyond its initial focus on justice and homeland security. NIEM communities currently exist around Agriculture, Emergency Management, Immigration, International Trade, Intelligence, Surface Transportation, and more.



NATIONAL INFORMATION EXCHANGE MODEL

[niem.gov](http://niem.gov)

## Case Study, The CitySDK: Combining Census data with other datasets

**U.S. Census Bureau** data from the Decennial Census and American Community Survey is the authoritative source for describing our nation's people and economy. This data helps city planners and leaders better allocate their investments, businesses optimize their operations, and nonprofits understand if their programs are being successful. To make its data as easy as possible to use, the Census Bureau built the CitySDK (software development kit), inspired by a project developed in the European Union.

The CitySDK is designed to help users combine different data sets to generate new insights. For example, combining Census data on median income with USDA data on the Supplemental Nutrition Assistance Program (SNAP) can help show how effective that program is and who it is reaching. From a technical perspective, the City SDK allows for data retrieval from different silos, geocoding, caching of large datasets, and mapping – all within a simple programming approach. It is now hosted on Amazon Web Services, adding support for all programming languages. By using the City SDK, developers, data scientists, and civic innovators can now get insights from open data or build products faster than ever before.

In rolling out the CitySDK, Census set out to build a community to make data easier to use and improve the feedback loop between the public and Census. This included launching the Census Open Source program and published guidelines to encourage community participation. The Census Bureau has also partnered with civic tech organizations like Code for America, and partners across the federal government, to help transform how Census engages with data users and invite them to help co-create an open-source product.



[census.gov](https://census.gov)

### 3. Managing Privacy

Before government data can be published as open data, agencies need to be sure they have masked or removed personally identifiable information (PII) effectively. There is continuing concern about the mosaic effect, through which disparate datasets can be combined to identify individuals. It can be a particular problem when groups of individuals are being studied. Combining different datasets from different agencies can raise this concern. Potential negative impacts include breaches of individual privacy and the chance that data will be used in a discriminatory way.

Data scientists are also realizing the limits of de-identification technology, which is a useful approach but not a complete solution. While technologies exist to remove identifying information from datasets, they are not fully effective. The technology is difficult to apply to the range of data now available, including geospatial, medical and genomic, body-camera, and other data. Finally, even if it is possible to de-identify data today, it's impossible to predict whether it will become possible to re-identify individuals as technology evolves in the future.

In this context, the best approaches seem to acknowledge the risk that individuals may be linked with their data, minimize that risk as much as possible, and look at the risk in the context of the potential both for individual harm and public good.

#### 1. Create customized privacy-protection programs based on risk assessment for each agency or program.

Although there are risks to opening data, policymakers can create programs and assessment tools that reduce these risks. Round-table participants noted that data-sharing culture should recognize and assess the actual risk for releasing a given dataset under different conditions. The potential damage from someone breaking the code and learning where an individual went to college, for example, is much less than the potential harm from revealing that same person's medical history. For that reason, each agency should assess

the true risk for every dataset that contains PII and choose strategies for managing those datasets accordingly.

When truly sensitive data is at stake, agencies or cross-agency programs will need to develop thorough, coordinated plans for privacy protection. For example, the Precision Medicine Initiative, which is intended to help patients personalize their health care, has developed a framework for protecting privacy without inhibiting this scientific work.

#### 2. Consider a range of privacy protection strategies, and consider using them in combination.

The Roundtable discussions included a number of potentially complementary strategies for privacy protection. Key strategies include:

- **Balancing strategies.** Agencies can balance the risks of releasing data against the potential for public good. This is the approach the Consumer Financial Protection Bureau (CFPB) is planning to use to release data under the Home Mortgage Disclosure Act (HMDA). The CFPB is statutorily mandated to publicly disclose data under HMDA while developing appropriate protections for borrower privacy in light of HMDA's purposes. Following a recent rulemaking, the CFPB will use a "balancing test" with public input to determine the right balance of serving the public good and protecting individual privacy in this data release.

- **Differential access.** It may be necessary to consider gradations of openness under different circumstances. For example, some kinds of data could be made "open" only for sharing between federal agencies under certain conditions, or sharing only with qualified and vetted researchers, rather than opening it to the public at large. Approaches include:

- Interagency transfer of data that is controlled and kept securely between the two agencies involved.
- Federated model using a cloud repository and limiting access to trusted users. Requires a secure way to upload data as well as secure ways to share it.

- Tiered access data-sharing programs to allow levels of access to multiple types of users.
- Opt-in and permission-based mechanisms that enable individuals to make their data more widely available if they choose to. For example, individual patients have an incentive to share data about their condition in the hope that it will be used to find better treatments.
- **De-identification.** It may be technically impossible to create a method of de-identification, removing PII from public datasets, that retains the full value of the data and is completely effective at anonymizing it. However, there are many situations where a high level of de-identification is sufficient, even if it does not provide absolute, 100% privacy protection. Conversely, it may be possible to completely de-identify data if researchers can accept less-than-perfect accuracy in the result. Approaches include:
  - Identifying individuals with unique ID numbers that make it possible to connect data about them in different datasets without revealing their identity.
  - Dropping non-critical information to make re-identification more difficult. For example, one regular practice is to drop the last three digits of an individual's zip code.
  - Using differential privacy and synthetic data. Participants noted, however, that more data scientists and resources are needed to apply sophisticated tools like these.

**Individual privacy should be treated in the context of public good.**

### 3. Coordinate Disclosure Review Boards, Chief Privacy Officers, and other governance structures.

New data governance structures can help manage privacy concerns. Many agencies now handle privacy issues through a Chief Privacy Officer, a Disclosure Review Board, or other offices and organizational structures. To make these as effective as possible, their work needs to be integrated and aligned with the agency's goals for data release. Options include:

- Use the office of the Chief Data Officer to centralize each agency's management of open government data, and address privacy concerns. The Chief Data Officer is in the best position to lead a process that coordinates the different individuals and groups working on privacy-related issues, including the Chief Privacy Officer, Disclosure Review Board, and others.
- Strengthen the role of Disclosure Review Boards within agencies and the way they operate, including participation from the General Counsel's office and subject matter experts. Develop a core set of policies and procedures and a framework for Disclosure Review Boards' operations that can be customized for each agency.
- Create model infrastructure – a virtual central data hub where access to data and APIs is managed by a common set of metadata (security, definitional, sharing licences) and user agreements.

### 4. Build trust with the community around data use.

Individual privacy should be treated in the context of public good. Many datasets that include PII also include information that can have great public benefit. In these cases, it will be essential to craft approaches to privacy protection that respect individuals' rights while also making data available to the public, or to selected researchers, in a way that supports social and scientific goals.

It is also essential to communicate the goals of open data, and privacy safeguards for the data, to the community and individuals that have provided it. Individuals are understandably concerned that data about their health, education, employment, financial status, or other sensitive data should not be exposed or misused. Federal agencies and others that plan to use the data with appropriate privacy protections will need to be sure that the communities involved understand and are satisfied with their approach.

## Case Study, HMDA Home Mortgage Data: Balancing privacy with public need-to-know

The Home Mortgage Disclosure Act (HMDA) was enacted in 1975 to help prevent discrimination in lending. The Consumer Financial Protection Bureau (CFPB), which now implements HMDA, has taken a number of steps to improve HMDA data and to make it more useful. The agency hosts a variety of online tools to help users explore the data and has built a public API, giving advanced users the ability to use HMDA data in their own external applications and projects.

In addition to helping prevent discrimination, releasing HMDA data serves several purposes. It can provide the public and public officials with sufficient information to enable them to determine whether institutions are serving the housing needs of the communities in which they are located. It is also designed to help public officials distribute public sector investments in a manner that will improve the private investment environment. HMDA data can only serve these purposes if it is made publicly available at a sufficient level of detail.

The CFPB is statutorily mandated to release HMDA data while developing appropriate protections for borrower privacy in light of HMDA's purposes. In a recent amendment to the HMDA, the CFPB has stated that it will use a "balancing test" and public input to determine how to best serve the public good while simultaneously protecting individual privacy. Balancing tests can be useful for agencies when considering the risks associated with data release in comparison to the potential for public good. The Census bureau has also utilized this approach.



[consumerfinance.gov](http://consumerfinance.gov)

## Case Study, Police Data Initiative: Managing sensitive data by working with the community

In May 2015 the White House launched the Police Data Initiative (PDI) with an initial group of 21 police departments from across the country along with a range of partners. The PDI strives to leverage data and technology to build community trust and increase internal accountability while decreasing the inappropriate use of force. As of April 22, 53 jurisdictions, representing more than 40 million people, have signed on to participate in the initiative.

Through the PDI, police departments are working with data and technology partners to overcome technical and other hurdles and improve data sharing and analysis. Partners include Code for America, The Police Foundation, ESRI, CI Technologies, Socrata, Operation Spark, The Sunlight Foundation, The Southern Coalition for Social Justice, The International Association of Police Chiefs, The University of Chicago, and Stanford University.<sup>1</sup>

Working with police data poses challenges to security and privacy, including concerns about releasing data on potential perpetrators, victims, and individual officers' actions. Several police departments have taken this challenge as an opportunity to work with the community to find solutions together. For example, "the New Orleans Police Department...previewed policing datasets with a group of young coders and their tech mentors [and] the Orlando Police Department worked with sexual assault and domestic violence victim advocates to figure out how to balance transparency with victim privacy..." By taking this kind of approach, a number of "communities and police departments [are] using data as a way to engage in dialogue and build trust."<sup>2</sup>

In a little more than a year the PDI has achieved several concrete milestones. The Police Foundation launched its Public Safety Open Data Portal, pulling in data released by participating law enforcement agencies.<sup>3</sup> Around 100 datasets have been released and ingested in various ways.<sup>4</sup> The number of jurisdictions, non-governmental partners, and specific commitments tied to the PDI continues to grow.<sup>5</sup>

1. The White House, "Launching the Police Data Initiative"

2. The White House, "The Police Data Initiative Year of Progress"

3. Police Foundation, "Public Safety Open Data Portal", [publicsafetydataportal.org](http://publicsafetydataportal.org) (accessed July 1, 2016)

4. Wardell, Clarence and Ross, Denice. "The Police Data Initiative Year of Progress: How We're Building on the President's Call to Leverage Open Data to Increase Trust between Police and Citizens." The White House, 22 April 2016. [medium.com/the-white-house/the-police-data-initiative-year-of-progress-how-we-re-building-on-the-president-s-call-to-leverage-3ac86053e1a9#.e9w-jbqdh6](https://medium.com/the-white-house/the-police-data-initiative-year-of-progress-how-we-re-building-on-the-president-s-call-to-leverage-3ac86053e1a9#.e9w-jbqdh6) and Ortellado, Damian. "Sunlight adds more than 100 new datasets to our Hall of Justice database." Sunlight Foundation. 26 May 2016. [sunlightfoundation.com/blog/2016/05/26/sunlight-adds-more-than-100-new-datasets-to-our-hall-of-justice-database](http://sunlightfoundation.com/blog/2016/05/26/sunlight-adds-more-than-100-new-datasets-to-our-hall-of-justice-database).

5. The White House, "Fact Sheet: White House Police Data Initiative Highlights New Commitments", April 21, 2016, [whitehouse.gov/the-press-office/2016/04/22/fact-sheet-white-house-police-data-initiative-highlights-new-commitments](http://whitehouse.gov/the-press-office/2016/04/22/fact-sheet-white-house-police-data-initiative-highlights-new-commitments) (accessed July 1, 2016).

## 4. Data Release

**Despite federal mandates** requiring open data, some government agencies may still be reluctant to share their data freely. Their reluctance may include concerns over data quality or the ways that data may be interpreted.

In addition, many scientific researchers are hesitant to share their data. Federal policies now require an open-science approach, which would provide greater access to publicly funded research data. Despite the clear benefits of open science, researchers face many challenges in sharing data, including the need for data platforms and concerns about data privacy, confidentiality, and intellectual property rights. Researchers may also worry about attribution for their data and interpretation of the data by other researchers.

Developing new incentives for researchers may be the most important strategy to promote data-sharing. Researchers now have many reasons to control their own data until they have gotten the maximum publication value from it, and few incentives to share the data. Career advancement depends more on publishing scientific papers than on creating the datasets those papers are based on. New incentives could credit and reward data sharing in ways that encourage open science.

Data governance and technical initiatives are also needed for an integrated approach to data sharing. Implementation of technical standards, data platforms, and tools for interoperability will all help promote data-sharing. So will new approaches to data governance that manage research data from different sources in an integrated way.

### 1. Release both raw data and improved data with transparency about accuracy, quality and provenance.

Raw data is more likely to have quality problems but is also timelier than data that has been vetted and cleaned. Despite its imperfections, raw data may lead to innovative insights and uses. Strategies for releasing both

raw and improved data can give data users more options, as long as the provenance and limitations of the datasets are clear, and as long as the “authoritative” dataset is clearly identified. Roundtable participants recommended releasing both raw data and improved versions as they are developed, and keeping both versions available. This strategy meets data users’ needs for both timeliness and quality.

Datasets do not have to be perfect to be usable, but it’s important to know the strengths and weaknesses of data in order to be able to use it. All data should be released together with transparent information about its accuracy, quality, and provenance. Some participants recommended conducting pre-release testing to assess data quality as an additional safeguard. This could include developing an experimental space to give users access to work with data so they can test it before it is made public.

### 2. Employ user-focused communication strategies to encourage data dissemination and use.

The case for open data becomes more compelling if there are clear examples that show its value to users. Federal data stewards, as well as independent organizations, can collect and publish case studies that show how open government data is being applied, and can help inspire others to release their data as well. Agencies can develop tutorials, customized user interfaces, and other tools to make their data easier to use. Hackathons and challenges can engage the open data community and others to find new ways to put data to use.

### 3. Require data sharing and publication as a condition of research funding and help researchers meet that requirement.

While federal guidelines now include an expectation of data-sharing for federally funded research projects, the guidelines could be made stronger and more specific. Guidelines now require grantees to develop data management plans with an expectation that, at

# Funding, tenure, and other career incentives could encourage researchers to share data.



minimum, the data underlying publications will be made accessible and shared. Federal funders could tie grants to clearer, binding requirements to adopt open standards and share data publicly to the greatest extent possible, taking privacy and other concerns into consideration.

At the same time, federal grant-makers can provide positive incentives and help researchers meet the data-sharing requirement. They can value open data more highly in funding decisions, giving extra points to grant applicants who are committed to sharing their data. Funders can also provide sample data management plans for federal grantees. While grant applicants are now required to develop these plans, it would be beneficial to clearly encourage data sharing within the goals and recommendations for data management and data infrastructure development.

#### 4. Use new incentives to promote research data sharing more widely.

Currently, there are many incentives against sharing research data and few that support it. The challenges include both cultural and pragmatic obstacles. The current scientific culture is not to share data, but for individual researchers to hold datasets for their own use. The academic model does not reward data sharing. Since academics are rewarded for publishing peer-reviewed articles more than for publishing datasets, researchers want to get maximum publication value out of their data before releasing it. They may also worry about attribution for their data and interpretation of the data by other researchers. In addition, data sharing can be expensive, and it is not clear how to fund it.

New ways to reward data-sharing through funding, tenure decisions, and other career incentives could significantly increase data-sharing by researchers. A key is to ensure that researchers receive systematic and meaningful credit for sharing their data. Data citation systems, similar to the citations for published papers, could help researchers gain credit for their work, measure the impact of their research, and advance professionally. They could form the basis for “report cards” that researchers can access to see how their data is being used. This would be similar to the way some organizations now support the use of open source software.<sup>6</sup>

While focused on researchers, improved citation systems for data could also help federal agencies and research institutions track the use and impact of the data they produce. This effort could be supported by the GSA, data.gov, and OMB guidance to federal agencies, in partnership with organizations that support open science.

#### 5. Develop collaborations and outreach to collect, manage, and publish data.

Scientists in many research domains have begun to form wide-ranging collaborations around data. These collaborations make it possible for researchers to draw on each others’ work to accelerate the pace of science. They also enable scientists to draw on diverse groups for data collection: for example, tribal researchers in the Arctic, patients who volunteer data in health studies, and the population at large for a number of citizen science projects. Some examples, described in case studies following and in the Research Domain Recommendations section, are:

- **The National Institutes of Health Data Commons**, a shared virtual space where scientists can find, deposit, manage, share, and reuse data and metadata.
- **The Cancer Moonshot Initiative**, a \$1 billion initiative to “eliminate cancer as we know it” with open data and robust data-sharing infrastructures.
- **Study of Environmental Arctic Change (SEARCH)**, a U.S. program that provides scientific information to Arctic stakeholders, policy-makers, and the public.
- **The Materials Genome Initiative**, a multi-agency initiative to help discover, manufacture, and deploy advanced materials rapidly and cost-effectively.
- **The BRAIN Initiative** (Brain Research through Advancing Innovative Neurotechnologies), a bold new initiative focused on revolutionizing our understanding of the human brain.

Despite its imperfections, raw data may lead to innovative insights and uses.

6. See [Depsy.org](http://Depsy.org), for example.

## Case Study, The Commons: One approach to health data sharing

The NIH Commons is being developed to help researchers share their data so it can be widely leveraged. In more detail:

» The Commons is defined as a shared virtual space where scientists can find, deposit, manage, share and reuse data, software, metadata and workflows – the digital objects of biomedical research. It is a digital ecosystem that supports open science and leverages currently available computing platforms in a flexible and scalable manner to allow researchers to transparently find and use computing services and tools they need, access large public data sets and connect with other resources associated with scholarly research...

The Commons uses cloud services to make data accessible, and ensures that datasets adhere to standards that make them findable, accessible, interoperable, and reproducible (FAIR).

Several pilots are being conducted to test and evaluate the technical components of the Commons. They include a “cloud credits” business model that gives researchers credits to choose the best cloud service for their needs, and a Human Microbiome Project (HMP) that uses cloud hosting along with a suite of tools and APIs to facilitate data access and use.



[commons.era.nih.gov](http://commons.era.nih.gov)

## Case Study, My Data Initiative: Giving people access to their data securely and safely

In 2010 the White House launched a series of My Data initiatives “to ensure all Americans have easy and secure access to their own personal data, whether related to health, energy, finance, or education.” The goal of My Data is to empower consumers with their own data to put them “in the driver’s seat to make informed choices.”<sup>7</sup>

High profile examples of My Data include Blue Button and Green Button. The Blue Button initiative began by making individual medical data from the Department of Veterans Affairs available to veterans, but has since expanded to serve a much wider constituency. The initiative is designed to help individuals track their health, correct errors in their medical records, and improve the sharing of health information between doctors, patients, and families. The Green Button initiative provides data about residential and commercial energy usage to help save money on energy costs. Other notable My Data initiatives include My Social Security and My Student Data.<sup>8</sup>

My Data initiatives do not simply make it easier for citizens to access their personal information. They also “raise the bar for both public and private organizations, empowering consumers through public-private data interoperability, security and access.” For example, Green Button data is reviewed by the National Institute of Standards and Technology’s cyber-security team to ensure that it does not contain any personally identifiable information.<sup>9</sup> The General Services Administration is now working on a service that will let individuals sign on only once to get access to their data from several different Federal websites or services.

7. The White House, “My Data: Empowering All Americans with Personal Data Access”, March 15, 2016, [whitehouse.gov/blog/2016/03/15/my-data-empowering-all-americans-personal-data-access](https://whitehouse.gov/blog/2016/03/15/my-data-empowering-all-americans-personal-data-access) (accessed July 1, 2016).

8. Social Security, “My Account”, <https://www.ssa.gov/myaccount> and Federal Student Aid, “My StudentData Download”, [studentaid.ed.gov/sa/resources/mystudentdata-download](https://studentaid.ed.gov/sa/resources/mystudentdata-download)

9. Green Button, “An Overview of the Green Button Initiative”, [greenbuttondata.org/learn](http://greenbuttondata.org/learn) (accessed July 1, 2016).

# 5. Quality Improvement

**Improving data quality** leads to a number of social, scientific, economic, and operational benefits. High quality data can lead to better public services, the advancement of scientific research, and an increase in business opportunities. Operational benefits that result from improved data quality include more usability and reusability, better organizational efficiency, increased collaboration with and within organizations, and higher organizational value from increased trust.

Data quality is comprised of many factors that can be addressed individually or in combination. These include data accuracy, metadata, machine-readability, timeliness, and interoperability, among others. Federal agencies, private-sector companies, and researchers have developed solutions that can improve each of these factors throughout the data lifecycle.

## 1. Develop user feedback systems to flag quality problems for agencies.

User feedback is critical to improve data quality. Data users can identify quality problems and can often help improve quality by providing data of their own. Community management and feedback can help find and eliminate data quality problems in the same way that it can address bugs in open-source software. In addition to providing channels for feedback, data stewards could proactively invite the developer community to evaluate data according to basic quality requirements.

Roundtable participants specifically suggested developing stronger feedback channels for data.gov. While the website has a high volume of traffic, it receives very little feedback through the simple “report a problem” buttons that are on each page. Data.gov could develop a public interface for feedback, market the program, and engage the private sector, academics, and others to analyze and address the feedback. In addition to quality improvement, this could help data.gov and participating agencies determine what kinds of data are highest priority for data users.

Government agencies could also adapt the model that the Department of Health and Human Services has developed in its Demand-Driven Open Data (DDOD) project. The HHS IDEA Lab created DDOD to give stakeholders from industry, academia, nonprofits, and other government organizations a feedback pathway to share concerns about HHS data. This systematic approach to gathering, tracking, and acting on user feedback can be a scalable model for data quality improvement. Benefits include prioritizing resources to improve quality of the most widely used and valuable datasets.

## 2. Use challenges and competitions to improve data quality.

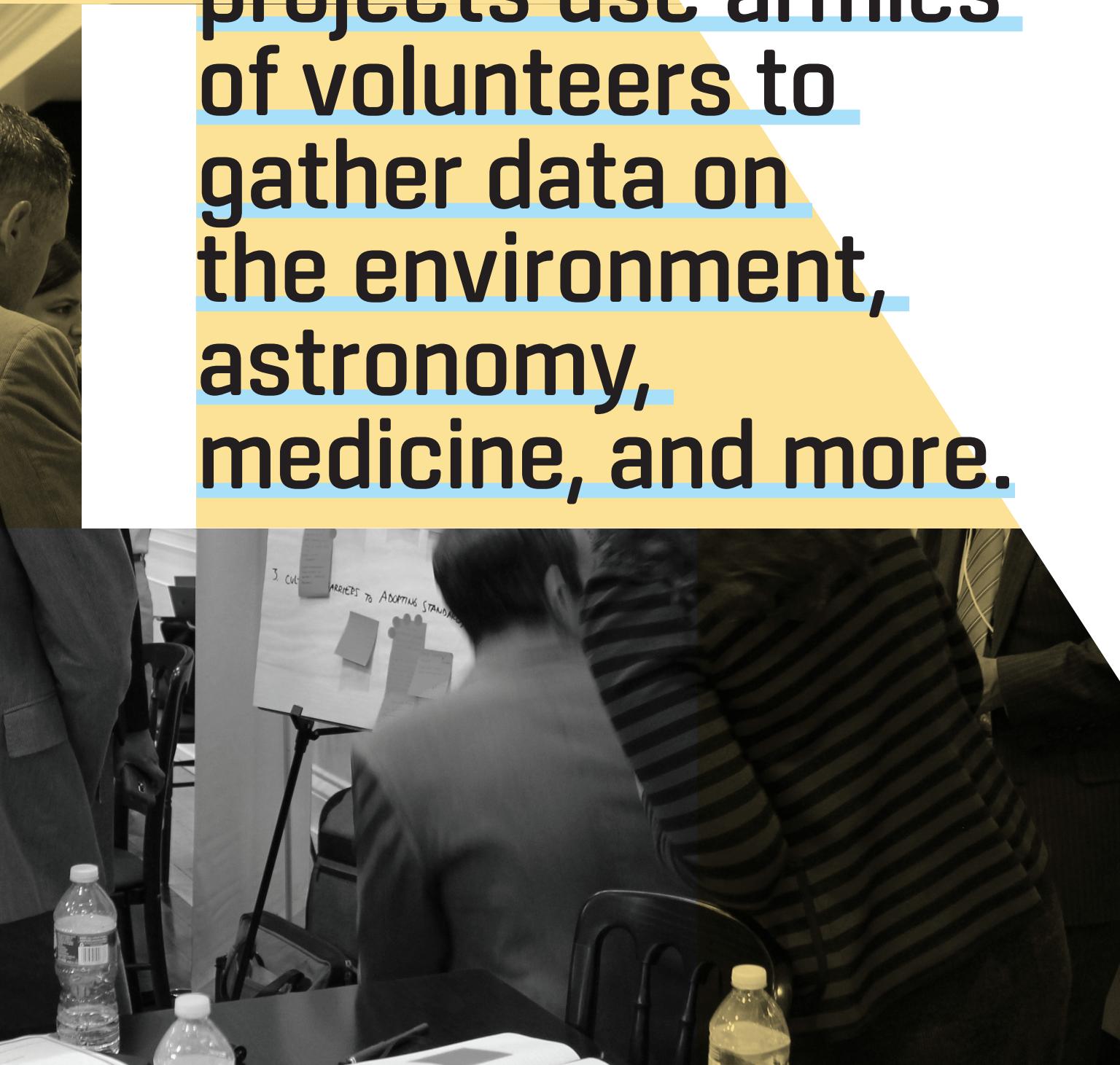
Data quality problems often have technical solutions. Algorithms can be developed to find and clarify instances where data includes some ambiguity, for example, when an individual or organization is not identified consistently. In one recent example, the U.S. Patent and Trademark Office put out a technical challenge for “disambiguation” of patent data to improve its data stores. The result has been PatentsView.org, a prototype data visualization and analysis platform that allows users to interact with 40 years of patent data, improved to make the data more informative and high quality.

## 3. Use crowdsourcing to improve data quality.

Many “citizen science” projects now use armies of volunteers to gather or categorize data on the environment, astronomical phenomena, medical images, and more. The White House Office of Science and Technology Policy has recognized crowdsourcing as a legitimate and valuable way to help build scientific data resources. In a similar way, agencies can use crowdsourcing to help improve data quality, by inviting citizens to review their data and contribute corrections. This approach can be helpful, for example, in improving geospatial data so that it reflects real-world locations.



# "Citizen science" projects use armies of volunteers to gather data on the environment, astronomy, medicine, and more.



## Case Study, Demand Driven Open Data Program: A user-focused method for quality improvement

The Demand Driven Open Data (DDOD) program grew out of a desire to maximize the value of existing data assets across the Department of Health and Human Services. It has led to improvements in the quality, relevance, and usefulness of HHS' open data offerings, and provides a model that other government agencies could use as well.

DDOD works by giving stakeholders from industry, academia, nonprofits, and other government organizations a pathway to tell HHS what data they need and by creating a feedback loop that ensures follow-up and follow-through. It is more consistent and transparent than other engagement options such as one-off events and regulatory approaches.<sup>10</sup> It also helps ensure that HHS is releasing its most relevant data by asking stakeholders what data they need that isn't being released and how data that is released could be improved.

DDOD has had a range of positive effects on data quality, including improving machine-readability, helping identify and eliminate manual mistakes, and surfacing opportunities for standardization. A prime example has been the transformation of Medicaid Managed Care Organizations (MCO) data. Various related datasets were being published in PDF format, not being aggregated in a useful way, and not necessarily up to date. Thanks to DDOD, these issues were identified and simple solutions were put into place to ensure that the public has timely access to machine-readable MCO data.<sup>11</sup>



[ddod.healthdata.gov](http://ddod.healthdata.gov)

10. "Intro to Demand Driven Open Data for Data Users", Slideshare, [slideshare.net/DavidPortnoy/intro-to-demanddriven-open-data-for-data-users](http://slideshare.net/DavidPortnoy/intro-to-demanddriven-open-data-for-data-users) (accessed July 1, 2016). Consumer Financial Protection Bureau, "Public API", [consumerfinance.gov/data-research/hmda/api](http://consumerfinance.gov/data-research/hmda/api) (accessed July 1, 2016).

11. Demand Driven Open Data, "Use Case 46: Medicaid MCO Data", U.S. Department of Health and Human Services, [ddod.healthdata.gov/wiki/Use\\_Case\\_46:\\_Medicaid\\_MCO\\_Data](http://ddod.healthdata.gov/wiki/Use_Case_46:_Medicaid_MCO_Data) (accessed July 1, 2016).

## Case Study, Patents View:

### Making 40 years of patent data accessible with better quality

**PatentsView** is a prototype “data visualization and analysis platform intended to increase the value, utility, and transparency of US patent data.” It allows users to interact with 40 years of data. The platform grew out of an initiative by the Chief Economist at the U.S. Patent and Trademark Office (USPTO), in partnership with the Center for the Science of Science and Innovation Policy at the American Institutes for Research, New York University, the University of California at Berkeley, Twin Arch Technologies, Periscopic, and the U.S. Department of Agriculture (USDA).

PatentsView has several underlying goals -- to democratize patent data, to reduce redundancy, to facilitate linking, and to improve data quality. A high priority is “data disambiguation,” that is, solving the ambiguity that results when one inventor or organization uses different names or the same name is shared by different inventors or organizations. The patent database, for example, includes no fewer than five different inventors names associated with one and the same Steve Jobs. Similarly, there are 15 distinct organization names associated with Google that can be disambiguated into one entity.

Disambiguation can make all kinds of data more useful and informative. While the patents database is plagued by inventor confusion, other government databases may face similar problems when companies or corporations are named ambiguously, or even when state names are recorded in different ways. The collaboration behind PatentsView took a technical approach that could turn out to be applicable in other contexts. As the PatentsView website describes it, “The PatentsView API provides access to a newly developed data source that uses probabilistic methods to determine whether or not inventors with the same name are indeed the same person and generates disambiguated inventor identifiers...”

To tackle the disambiguation problem, USPTO has turned to the research community. In September 2015 the USPTO hosted an “Inventor Disambiguation Technical Workshop” to “find creative new approaches to get better information on innovators and the new technologies they develop by disambiguating inventor names.”<sup>12</sup> The winning team hails from The University of Massachusetts Amherst.

By making better-quality and more patent data public and open in one central location, PatentsView is intended to save time and money by “eliminating the wasteful and redundant cleaning, converting and matching of these data by many individual researchers.”



12. United States Patent and Trademark Office, “PatentsView Inventor Disambiguation Technical Workshop”, [patentsview.org/workshop](http://patentsview.org/workshop) (accessed July 1, 2016).

# 6. Curation, Storage and Management

While **open data** has become a high priority for the federal government, resources to support it are scarce. Agencies need to include data management strategies in their plans and make the best use of the resources, staff, and tools they have. By improving data governance, sharing tools across agencies, and collaborating with the private sector, agencies can make the most of the resources they do have for data curation, storage, and management.

## 1. Centralize and strengthen data governance at the agency level through Chief Data Officers.

Agencies can improve data quality throughout the data lifecycle by strengthening data governance. The introduction of Chief Data Officers into federal agencies provides an opportunity for stronger data governance, particularly if he or she is empowered to lead the collection, management, and dissemination of data across the agency. The Chief Data Officer is in the best position to lead a process that coordinates the different individuals and groups working on open data and related issues such as privacy and quality. Participants noted that more effective data governance could help throughout the data lifecycle. Some recommended using principles of Total Quality Management and improving data quality continually by focusing on data customers' needs.

## 2. Develop common tools, platforms, and catalogs for managing, sharing, and improving data.

Participants recommended using a number of tools including data catalogs, knowledge bases (e.g. MediaWiki), platforms such as Github and/or tools like Hackpad to share information about datasets and flag data quality issues. Some suggested creating a “Pinterest for data” to make it possible to see what’s being used and shared to determine what’s important to data customers. The data.gov platform can also be leveraged to learn how agencies are sharing data. Data providers could learn from successful examples in and outside of government such as Open EI (for energy), Globus (for research data), healthdata.gov, and the Open Medicine Foundation. Much can be done with participation from the private sector. For example, companies could join with government to make data more discoverable through search engines.

Adding Chief Data Officers to federal agencies can strengthen data governance - if they're empowered to lead data collection, management, and dissemination across the agency.

## 7. Building Data Communities and Collaboration

**Government agencies have** an opportunity to develop communities of practice around open data that include stakeholders both within and outside of government. Many healthcare, agriculture, financial services, energy, and transportation companies, among others, use open government data as a key business resource.

The private sector and government agencies have a mutual interest in ensuring that government data programs are high quality, easily accessible, and cost effective. Government data stewards can do much to inform potential users what data is available and how it can be used. In addition, open data stakeholders outside of government often have knowledge, expertise, resources, and processes that could benefit government data programs.

### 1. Use a range of approaches to identify user communities and measure data use.

Most federal agencies do not know exactly who their data users are or how they are using their data. Systems for tracking API usage can be helpful if they can be implemented without impinging on users' privacy. Web analytics and demographic analysis can also provide insight into open data users. At the same time, metrics to assess the use of open data can help show how effectively agencies' and researchers' data is being used and can help agencies prioritize the value of different datasets.

### 2. Use public-private collaborations to find efficient, scalable solutions to data challenges.

There is currently not enough federal funding to do everything necessary to improve federal open data, and private-sector companies often need federal data to meet their business goals. Public-private collaborations with a wide range of stakeholders can provide resources and expertise to make the country's open data more accessible and useful for all parties. From major companies to small nonprofits, many types of private-sector partners can work with the federal government to achieve their mutual open data goals. These stakeholders outside of government, as well as local and state governments, can work with federal data leaders to improve access, discovery, ease of use, and quality of open data.

**The private sector and government have a mutual interest in high-quality, accessible, cost-effective data programs.**

## Case Study, NOAA Big Data Partnership: Expanding data access through private sector collaboration

The National Oceanic and Atmospheric Administration (NOAA) Big Data Partnership (BDP) is an approach to publishing the agency's vast data resources with cost-efficient high performance computing, analytic and storage services. Working with private sector cloud infrastructure providers Amazon Web Services, Google, IBM, Microsoft, and the Open Commons Consortium, NOAA is leveraging its data to "create a sustainable, market-driven ecosystem that lowers the cost barrier to data publication."

NOAA collects data from the bottom of the ocean to the surface of the sun. The data comes from 10 satellites, more than 150 weather radars, 3 buoy networks, over 200 tide gauges, human observers, ships, aircraft, and more. NOAA gathers more than 20 terabytes of data every day -- adding up to more than 6 petabytes in FY2014 -- and user demand for access to that data is growing rapidly. But despite the growing demand, not all of NOAA's data is easily accessible to the public, especially in the case of very large datasets.

To change this, NOAA is partnering with private sector organizations that have the infrastructure and technical capacity to deal with the volume and complexity of NOAA's data. These organizations host selected NOAA datasets in the cloud, providing users remote access along with the ability to perform computations directly on the data. These partnerships are officially established through Cooperative Research and Development Agreements (CRADA's), a vehicle for collaboration that other research-driven agencies can also use.



[noaa.gov](http://noaa.gov)

## Case Study, Opportunity Project:

Working together with open data to increase opportunity locally

The Opportunity Project, developed by the White House, the U.S. Census Bureau, and the U.S. Department of Housing and Urban Development, makes it easy to combine key federal and local datasets to answer questions that provide opportunity for local residents. Participants in the project include Zillow, Measure of America, ESRI, PolicyLink, Socrata, Redfin, and more. These organizations have created tools that are aimed at helping local populations make smarter, more efficient decisions about their lives. They also often provide value to the organizations that built them. Examples include:

- Affordable Housing Finder: A tool that helps low income citizens with children find opportunities for affordable housing.
- Diversitydatakids.org: Maps that provide visibility into the relative affordability of opportunity for children in U.S. metro areas based on indicators including health, education, residential segregation, and more.
- GreatSchools: A collaboration with Zillow that makes it easy to search for housing near high quality schools that provide high levels of opportunity to students of various incomes and racial backgrounds.

The Opportunity Project makes it simple to download data for different uses. The Project is open to the public at large and its managers are actively seeking participants and case studies to publicize and share.



[opportunity.census.gov](http://opportunity.census.gov)

## Case Study, Green Button Initiative: Giving consumers access to their utility usage data

The Green Button Initiative helps “Americans access their detailed household or building electricity usage from their utilities online, facilitating virtual energy audits to identify inefficiencies and save money for both residential and commercial customers.”<sup>13</sup>

The initiative stemmed from a 2011 challenge by Aneesh Chopra, then the U.S. Chief Technology Officer, calling on utility companies to provide customers access to their energy-usage information in a simple, common format. The initiative existed as a Public Private Partnership until 2014 before a nonprofit was created in 2015 to help ensure compliance and provide an ongoing, open forum for public input and enhancement. The Green Button Alliance was launched “to foster the development, compliance, and wide-spread adoption of the Green Button standard.”<sup>14</sup>

Participating organizations agree to make energy meter data -- the data used by energy companies to calculate energy use and cost -- available to consumers for analysis and planning. Green Button platforms make use of a common data exchange standard, ensuring “energy information can easily be exchanged without requiring developers to invest time and money to support proprietary metered data conversion technolo-

gies...” and making it easier for developers to build functional, consumer-facing tools that utilize the data.

The combined efforts of government agencies and energy providers have ensured adoption of Green Button solutions. Over 60 million households and businesses in the U.S. can currently access energy usage data via Green Button and that number is likely to grow as more energy providers embrace the program.

To further boost confidence in the ecosystem and encourage adoption, the Green Button Alliance has launched a Green Button Certification program for implementations that allow the downloading of usage data and is currently working on a certification program for the connection of data directly to third-party companies authorized by a consumer to obtain those data on their behalf. “The presence of a Green Button certification mark on an energy provider’s website enables consumers and application developers to confidently rely on the data they receive being Green Button standard compliant.”



[greenbuttondata.org](http://greenbuttondata.org)

13. The White House, “My Data: Empowering All Americans with Personal Data Access”, March 15, 2016, [whitehouse.gov/blog/2016/03/15/my-data-empowering-all-americans-personal-data-access](http://whitehouse.gov/blog/2016/03/15/my-data-empowering-all-americans-personal-data-access) (accessed July 1, 2016).

14. Green Button Alliance, [greenbuttonalliance.org](http://greenbuttonalliance.org) (accessed July 1, 2016).

---

# Research Data Recommendations

While many recommendations can apply to all open government data, there are specific issues and opportunities for data across research domains. The Roundtables included data-driven experts who work in sector-specific groups to identify issues and make recommendations with a particular focus on scientific research. This section presents their recommendations and several case studies of successful data programs, organized by research domain.

## 1. Arctic

1. Coordinate engagement with tribal researchers across agencies.
2. Share data across institutions and initiatives to validate data by comparing multiple datasets and to reduce any redundancies in data collection.
3. Fund cloud-based management tools to match the pace of data collection.
4. Prioritize government leadership and funding for Arctic data.

## 2. Cancer and Biomedical

1. Use high-level government leadership and public-private collaborations to develop platforms for sharing biomedical research data.

## 3. Climate

1. Establish a paradigm of moving analytic services to the data through cloud hosting rather than downloading data for localized analysis.

## 4. Health

1. Develop clear guidelines at the outset of a project to manage and protect privacy.
2. Incentivize and enable patients to share their data voluntarily.
3. Develop a trusted ecosystem to make it possible to share sensitive data and patient data.

## 5. Infectious Disease

1. Work with the World Health Organization to expand open data practices and capacities in low to middle income countries and facilitate discussions between countries.
2. Share data during research through well-established organizations and institutions.
3. Share data during infectious outbreaks to help identify the focus of intervention, determine which interventions are working, and detect when pathogens change genetically.

## 6. Materials Science

1. Add data sharing plans into public governance models.
2. Strengthen open data mandates by providing government resources and working to change the culture and status quo of research data.

## 7. Oceans

1. Draw on the private sector, federal government, state/local governments, and nonprofits, academics, and developers to improve data access, discovery, quality, and use.

## 8. Smart Cities

1. Support research across geographies; share data and research internationally.
2. Create incentives for university-city pairs.
3. Use common information models (like the transportation XML standard) and simplified metadata standards.

## 9. Soil and Agriculture

1. Create public buy-in to encourage landowners to share data.
2. Use common metadata standards like DCAT (Data Catalog Vocabulary) to make it possible to discover and aggregate data from diverse repositories.
3. Develop collaborations for multi-state, multi-university projects.
4. Build agency-based repositories for grantee-generated research and evaluation data, and make the data available to researchers.
5. Streamline data publishing process to decrease barriers by developing a platform for curating data and providing agency assistance to principal investigators in storing data.

# Arctic

The Roundtables included participants with expertise in the Arctic region and an interest in using satellite, energy, and biodiversity data to advance community sustainability and resilience. Arctic researchers face a number of data challenges in scale, time, and coordination. They study an immense region, making it difficult to acquire data comprehensively. They are studying rapid climate-related changes, making it critical to capture, use, and share data in real time. Their work involves collaborating with a diverse range of international stakeholders, which can be difficult to manage. More locally, they collaborate with local tribes, a set of alliances that requires cultural sensitivity and careful coordination. At present, the level of demand from various different research groups risks overwhelming and alienating tribal partners.

Participants also identified a need for 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.

In addition, the participants identified a number of recommendations for Arctic research:

- **Coordinate engagement with tribal researchers across agencies.** OSTP could lead tribal engagement relating to research, check on number and types of engagements, and ensure that tribal outreach is coordinated.
- **Share data across institutions and initiatives to validate data by comparing multiple datasets and to 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. This could be done with private-sector support from geospatial companies.
- **Prioritize government leadership and funding for Arctic data.** Have OSTP design a template and guidance for agencies to open Arctic data. Provide more resources for Arctic data through USG policies/budget.

**Needed:  
An initial  
agreement  
between  
countries  
that  
advances  
open data  
in Arctic  
science.**

## Case Study, Study of Environmental Arctic Change (SEARCH): A national program with global impacts

The Study of Environmental Arctic Change (SEARCH) is a U.S. Interagency Arctic Research Policy Committee (IARPC) program with a mission of global importance. SEARCH "provides scientific information to Arctic stakeholders, policy-makers, and the public to help them understand and respond to Arctic environmental change... facilitates research activities across local-to-global scales, with an emphasis on addressing needs of decision-makers..." and collaborates with relevant national and international science programs. SEARCH receives most of its support from the National Science Foundation, with additional support from other partners inside government and academia. This funding enables SEARCH to share data collected as part of this program to benefit a larger scientific community and other stakeholders.

Over the past several years, the Arctic research community has increasingly implemented data sharing and open data through SEARCH and other government programs. The International Arctic Buoy Programme maintains "a network of drifting buoys in the Arctic Ocean to provide meteorological and oceanographic data."<sup>15</sup> The U.S. Navy has also worked with international partners -- including France, South Korea, and the UK -- to study the atmosphere, ice, ocean and sea using robotic technologies to gather relevant data.<sup>16</sup>



[arcus.org/search-program](http://arcus.org/search-program)

15. International Arctic Buoy Program, "Overview", [iabp.apl.washington.edu](http://iabp.apl.washington.edu) (accessed July 1, 2016).

16. United States Navy, "#NavyInnovates in the Arctic", March 9, 2015, [navylive.dodlive.mil/2015/03/09/navyinnovates-in-the-arctic](http://navylive.dodlive.mil/2015/03/09/navyinnovates-in-the-arctic) (accessed July 1, 2016).

# Cancer and Biomedical

**T**he challenge of sharing research data, a major focus of the Roundtables, has been central to many new efforts in the biomedical sciences. It's core to the work of the Cancer Moonshot Initiative, a new \$1 billion research program that promises to boost the production, use, and sharing of cancer-related research data. Large, collaborative research programs face many challenges in developing effective ways to share the research data essential to their progress. Research groups use different formats, structures, processes, and standards in collecting and managing their data, making it difficult to synthesize their work. Despite their differences, such groups share a common goal:

- Use high-level government leadership and public-private collaborations to develop platforms for sharing biomedical research data.

Three case studies demonstrate how biomedical research data can be shared effectively: The **Cancer Moonshot Initiative**, the **BRAIN Initiative**, and the **National Microbiome Initiative**.

New programs on  
cancer, the brain, and  
microorganisms rely  
on sharing  
research data.

## Case Study, Cancer Moonshot Initiative: A \$1 billion fight against cancer based on data sharing

Cancer will kill an estimated 600,000 Americans in 2016 and affect the lives of an untold number of families. In February, the President announced \$1 billion to kickstart an initiative to fight cancer and ultimately “eliminate cancer as we know it.” The data-driven national Cancer Moonshot initiative “will work to accelerate...research efforts and break down barriers to progress by enhancing data access, and facilitating collaborations with researchers, doctors, philanthropies, patients, and patient advocates, and biotechnology and pharmaceutical companies.” This effort is being coordinated by a White House Cancer Moonshot Taskforce, and will leverage the power of the federal government as well as academia and the private sector.<sup>17</sup>

Open data and robust data-sharing infrastructures will be key to achieving the goals of the Cancer Moonshot Initiative. The Cancer Genome Atlas, a collaborative effort between the National Cancer Institute and the National Human Genome Research Institute, has mapped key genomic changes in 33 types of cancer. Its publicly available and widely used dataset measures in at 2.5 petabytes.<sup>18</sup> Though TCGA is coming to a close in early 2017, new NCI genomics initiatives, run through the Center for Cancer Genomics (CCG), will continue to use the same model of collaboration for large-scale genomic analysis and make the genomics

data publicly available.<sup>19</sup> The International Cancer Genome Consortium has a similar goal – “obtain a comprehensive description of genomic, transcriptomic and epigenomic changes in 50 different tumor types and/or subtypes which are of clinical and societal importance across the globe” – and operates across the globe. The ICGC has 79 projects and a presence in 17 countries.<sup>20</sup>

At the same time, a comprehensive effort to characterize the proteogenomics of primary tumors – an approach that combines studying proteins and genetic material – was launched in 2011 by the National Cancer Institute. Open data on 375 patients were released in 2013 with landmark papers being published in May and June of 2016.<sup>21</sup> As part of the Cancer Moonshot, a collaborative effort between the National Cancer Institute, Veterans Affairs, and Department of Defense to conduct proteogenomics characterization of 8,000 patient was announced at the Vice President’s Cancer Moonshot Summit.<sup>22</sup> This was quickly followed-up in July 2016 by an international collaboration announced by the Vice President in Australia where a combined 60,000 patients will have proteogenomics characterization done jointly by the two countries and data sets made publicly available.<sup>23</sup>



[whitehouse.gov/CancerMoonshot](http://whitehouse.gov/CancerMoonshot)

17. The White House, “Fact Sheet: Investing in a National Cancer Moonshot”, February 1, 2016, [whitehouse.gov/the-press-office/2016/02/01/fact-sheet-investing-national-cancer-moonshot](http://whitehouse.gov/the-press-office/2016/02/01/fact-sheet-investing-national-cancer-moonshot) (accessed July 1, 2016).
18. National Institutes of Health, “Program Overview”, The Cancer Genome Atlas, [cancergenome.nih.gov/abouttcga/overview](http://cancergenome.nih.gov/abouttcga/overview) (accessed July 1, 2016).
19. International Cancer Genome Consortium, [icgc.org/icgc](http://icgc.org/icgc) (accessed July 1, 2016).
20. [proteomics.cancer.gov/news/events/newsannouncements/archive/2016/Breast-Cancer-Provides-Therapeutic-Insight](http://proteomics.cancer.gov/news/events/newsannouncements/archive/2016/Breast-Cancer-Provides-Therapeutic-Insight); <http://proteomics.cancer.gov/news/events/newsannouncements/archive/2016/Ovarian-Cancer-Study-Uncovers-New-Biology>
21. [medium.com/cancer-moonshot/apollo-network-canserve-cancer-patients-through-collaboration-15a3530ed57e#.c2uqvqfc6](http://medium.com/cancer-moonshot/apollo-network-canserve-cancer-patients-through-collaboration-15a3530ed57e#.c2uqvqfc6)
22. [whitehouse.gov/the-press-office/2016/07/16/fact-sheet-victoria-comprehensive-cancer-center-vice-president-biden](http://whitehouse.gov/the-press-office/2016/07/16/fact-sheet-victoria-comprehensive-cancer-center-vice-president-biden)

## Case Study, BRAIN Initiative: Accelerating neuroscience with data-sharing platforms

In April 2013, President Obama announced the launch of the BRAIN (Brain Research through Advancing Innovative Neurotechnologies) Initiative® – a bold new initiative focused on revolutionizing our understanding of the human brain. The BRAIN Initiative has the potential to do for neuroscience what the Human Genome Project did for genomics by supporting the development and application of innovative technologies that can create a dynamic understanding of brain function. It aims to help researchers uncover the mysteries of brain disorders, such as Alzheimer's and Parkinson's diseases, depression, and traumatic brain injury (TBI).

The initiative will accelerate the development and application of new technologies that will enable researchers to produce dynamic pictures of the brain that show how individual brain cells and complex neural circuits interact at the speed of thought. These technologies will open new doors to explore how the brain records, processes, uses, stores, and retrieves vast quantities of information, and shed light on the complex links between brain function and behavior.<sup>23</sup>

Since its launch in April 2013, the BRAIN Initiative has grown to include investments from five Federal agencies: the Defense Advanced Research Projects Agency (DARPA), the National Institutes of Health (NIH), the National Science Foundation (NSF), Intelligence Advanced Research Projects Activity (IARPA), and the Food and Drug Administra-

tion (FDA). Federal agencies are supporting the initiative by investing in promising research projects aimed at revolutionizing our understanding of the human brain, developing novel technologies, and supporting further research and development in neurotechnology. The President's 2017 Budget also proposes funding for the Department of Energy (DOE) to join DARPA, NIH, NSF, IARPA, and FDA in advancing the goals of the BRAIN Initiative. Major foundations, private research institutions, and companies including the Howard Hughes Medical Institute, Allen Institute for Brain Science, the Kavli Foundation, the Simons Foundation, GE, GlaxoSmithKline, as well as patient advocacy organizations and universities, have also answered the President's call to action.

In 2013 NIH convened a working group of the Advisory Committee to the Director to develop a plan for the NIH's portion of the BRAIN Initiative. That report, BRAIN 2025: A Scientific Vision, identified seven core principles for the NIH's portion of the BRAIN Initiative – one of which was to establish platforms for sharing data because "[p]ublic, integrated repositories for datasets and data analysis tools, with an emphasis on ready accessibility and effective central maintenance, will have immense value."<sup>24</sup>



[braininitiative.nih.gov](http://braininitiative.nih.gov)

23. The White House. "The BRAIN Initiative", [whitehouse.gov/BRAIN](http://whitehouse.gov/BRAIN) (accessed July 1, 2016).

24. National Institutes of Health. "The BRAIN Initiative", [braininitiative.nih.gov/2025](http://braininitiative.nih.gov/2025) (accessed July 1, 2016).

## Case Study, National Microbiome Initiative:

### Cross sector support to understand, protect and restore healthy function

The hidden world of microbiomes is essential to life in ways that are just now being recognized. The White House has described their significance this way:

- » Microbiomes are the communities of microorganisms that live on or in people, plants, soil, oceans, and the atmosphere. Microbiomes maintain healthy function of these diverse ecosystems, influencing human health, climate change, food security, and other factors. Dysfunctional microbiomes are associated with issues including human chronic diseases such as obesity, diabetes, and asthma; local ecological disruptions such as the hypoxic zone in the Gulf of Mexico; and reductions in agricultural productivity.<sup>25</sup>

The National Microbiome Initiative (NMI), launched in May 2016, is a collaboration between the White House Office of Science and Technology Policy, various federal agencies, and private-sector stakeholders that “aims to advance understanding of microbiome behavior and enable protection and restoration of healthy microbiome function.”

The launch of the NMI has sparked significant funding for microbiome research inside and outside of the federal government. Agencies involved in the effort include the Department of Energy, NASA, the National Institutes of Health, the National Science Foundation, and the U.S. Department of Agriculture. Non-governmental engagement is coming from the Bill and Melinda Gates Foundation, JDRF, the University of California San Diego, One Codex, the BioCollective LLC, the Health Ministries Network, and the University of Michigan.

Programs related to microbiome research are encouraging data sharing, improving coordination among federal agencies, boosting interdisciplinary research, sequencing microbiome and other genomic data for public benefit, and more. For instance, as part of its Data Commons project, the National Institutes of Health is running a pilot to host 20 terabytes of microbiome data on the AWS Cloud. The data comes from the Human Microbiome Project (HMP), an NIH-funded effort “to generate data and resources to characterize the commensal microbiota present in the human body.”<sup>36</sup>



25. The White House, “Fact Sheet: Announcing the National Microbiome Initiative”, May 13, 2016, [whitehouse.gov/the-press-office/2016/05/12/fact-sheet-announcing-national-microbiome-initiative](http://whitehouse.gov/the-press-office/2016/05/12/fact-sheet-announcing-national-microbiome-initiative) (accessed July 1, 2016).

26. National Institutes of Health, “Data Science Commons”, [datascience.nih.gov/commons](http://datascience.nih.gov/commons) (accessed July 1, 2016).

# Climate

**M**ajor 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 at the Roundtables proposed to:

- Establish a paradigm of moving analytic services to the data through cloud hosting rather than downloading data for localized analysis.

This strategy, which can help remove barriers to data use while creating incentives for data sharing, could be advanced by a coordinated effort to identify test cases, fund data storage, and provide training. Potential contributors include such climate research partnerships as the Earth Science Information Partners and Partnership for Resilience and Environmental Preparedness.

Climate scientists at the Roundtables proposed a number of steps to improve their data resources – some of which are current recommendations, and some of which will require new efforts. These include:

- Develop international standards for weather data.
- Develop metadata standards, including for data used in climate risk management.
- Bridge the data divide between data for climate mitigation and data for climate resilience.
- Use corroborative sources of climate data to verify quality.
- Aggregate and integrate data from different sources (federal, local, private) and between government departments.
- Understand users' needs and develop targeted products that meet them.

For climate studies, scientists plan to “move the analysis to the data” rather than the other way around.

## Government agencies at all levels, the private sector, nonprofits, and academics can help advance climate science.

- **Create greater demand for climate data** and improve public understanding through communication about:
  - Specific types of climate data available.
  - Difference between weather and climate data (timescale).
  - Types of data needed for different purposes.

Some scientists at the Roundtables proposed a number of ways that different stakeholders can help advance climate science:

- **Private sector:** Help improve data quality, specifically improving global weather data for Numerical Weather Prediction; develop consistent APIs to open data; fund partnerships with academia and nonprofits.
- **Federal government:** Facilitate and create online services and resources (e.g. Global Geospatial Consortium); engage with the public on Climate Resilience Toolkit; work with international stakeholders to develop an open climate data repository; implement data access directive (2015), data documentation directive, metadata quality checker, data quality policy.
- **State/local governments:** Survey users to define needs for improved climate data services.
- **Nonprofits, academics, developers:** Develop an Open Commons Consortium; apply digital identifiers to tag data so that it can be accessed in the same way even if it is located in different data commons; build on the Ontario Climate Change Symposium, which focused on agriculture to reduce greenhouse gasses. Help build user-centric data-driven applications.

## Case Study, U.S. Climate Resilience Toolkit: Tools, training, and data to help alleviate climate risk

The U.S. Climate Resilience Toolkit is designed to help individuals, businesses, and communities identify and address climate vulnerabilities. It lays out a five-step process for project development and implementation while providing access to helpful tools, expert advice, training, case studies and climate-related data. While the Toolkit is currently focused on U.S. federal government resources, it will soon be expanded across sectors "to include information and decision support resources from state and local governments, businesses, and academia and other non-governmental organizations."

The Toolkit describes five "steps to resilience" that include exploring climate threats, assessing vulnerability and risks, investigating options, prioritizing actions, and taking action. To facilitate these steps, the Toolkit also highlights a variety of tools, eases access to climate data from across the federal government, and provides access to trainings from government agencies and non-governmental organizations.

Additionally, the Toolkit features a Climate Explorer application. Climate Explorer was "built to support the U.S. Climate Resilience Toolkit...and offers interactive visualizations for exploring maps and data related to the toolkit's Taking Action case studies. It was built with data from ESRI Web Services and climate.data.gov.

The Toolkit is designed not only to provide information but to facilitate the public-private collaborations needed for success. It aggregates information about funding opportunities that "offer financial and technical resources to advance local adaptation and mitigation efforts." These opportunities come from governmental sources, including the USDA, the Department of Housing and Urban Development, the EPA, state governments including the Commonwealth of Massachusetts, and more. Private foundations and nonprofits including the Rockefeller Foundation, the Kresge Foundation, and the Wildlife Conservation Society are also funding programs in this area.



[toolkit.climate.gov](http://toolkit.climate.gov)

# Health

**T**he personal nature of health data makes sharing it especially challenging. Individuals are concerned about the privacy of their health information, and insuring the privacy of individual data can be time consuming and expensive. Obtaining informed consent for data release can help solve privacy concerns in principle but can be difficult to implement. Moreover, institutional review boards (IRBs) for different institutions do not use uniform consent forms, making it difficult to share and upload data.

Concerns over privacy are not only ethical but legal. There are various interpretations of the Health Insurance Portability and Accountability Act (HIPAA) and what it restricts. Researchers may need to get legal review on these concerns before using data, which can delay research for months.

Additional challenges include:

- Lack of consistency within patient records. They may contain information on such factors as gender and histology, but treatment plan data and clinical data are either missing or not coded consistently.
- Ownership issues around mortality data and medical claims data that can hinder data sharing between different health systems.
- Need to repurpose clinical data to use it as research data. Much clinical or health data is generated for billing or administrative purposes, not for science, and may not be of high enough quality for scientific research.

Health experts participating in the Roundtables identified several recommendations:

- **Develop clear guidelines at the outset of a project to manage and protect privacy.** The Precision Medicine Initiative, which will analyze patient records and genomic information in the interest of finding more individualized and effective treatments, has developed a framework for protecting privacy without inhibiting this scientific work.
- **Incentivize and enable patients to share their data voluntarily.** Patients could share their health information through an established process where all stakeholders recognize the value of their data. 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.
- **Develop a trusted ecosystem to make it possible to share sensitive data and patient data.** Look to other models, e.g. banking, for ways to build individuals' trust in institutions' ability to manage their sensitive data.

**Privacy and trust are central in managing personal health data.**

## Case Study, Precision Medicine Initiative: Privacy and Trust Principles for medical data

The Precision Medicine Initiative (PMI) is part of a new approach to disease treatment and prevention that “takes into account individual differences in people’s genes, environments, and lifestyles.” The goals of the PMI -- and precision medicine more broadly -- will enable “patients, researchers, and providers to work together toward development of individualized care” and will rely heavily on patient participation.<sup>27</sup>

These goals can only be met through the analysis of individual medical data. For that reason, privacy and security are core issues for the PMI. Two key documents have set the stage for addressing these issues. The PMI Privacy and Trust Principles lay the groundwork for public trust while the PMI Data Security Policy Principles and Framework establishes security expectations for participating organizations.<sup>28</sup>

The Privacy and Trust Principles “articulate a set of core values and responsible strategies for sustaining public trust and maximizing the benefits of precision medicine”. Developed by an interagency working group with expert consultation, they are broken down into six key areas: Governance; Transparency; Respecting Participant Preferences; Participant Empowerment through Access to Information; Data Sharing, Access, and Use; and Data Quality and Integrity.

The Data Security Policy Principles and Framework builds off the Privacy and Trust Principles to provide “a broad framework for protecting participants’ data and resources in an appropriate and ethical manner that can be tailored to meet organization-specific requirements,” while recognizing that there is no “one size fits all” approach and that data security is a constantly evolving field. The Data Security Policy Principles and Framework were developed pursuant to a collaborative process similar to that of the Privacy and Trust Principles’. The Framework is broken down into five key areas: Identify, Protect, Detect, Respond, and Recover.



[nih.gov/precision-medicine-initiative-cohort-program](http://nih.gov/precision-medicine-initiative-cohort-program)

27. National Institutes of Health, “Precision Medicine Initiative Cohort Program”, [nih.gov/precision-medicine-initiative-cohort-program](http://nih.gov/precision-medicine-initiative-cohort-program) (accessed July 1, 2016).

28. The White House, “Precision Medicine Initiative: Privacy and Trust Principles”, November 9 2015. [whitehouse.gov/sites/default/files/microsites/finalpmiprivacyandtrustprinciples.pdf](http://whitehouse.gov/sites/default/files/microsites/finalpmiprivacyandtrustprinciples.pdf) and The White House, “Precision Medicine Initiative: Data Security Policy Principles and Framework,” May 25 2016. [whitehouse.gov/sites/whitehouse.gov/files/documents/PMI\\_Security\\_Principles\\_Framework\\_v2.pdf](http://whitehouse.gov/sites/whitehouse.gov/files/documents/PMI_Security_Principles_Framework_v2.pdf)

# Infectious Disease

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. (Much of the data from low-income countries is created outside the scientific community.) 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.

Public health surveillance research should be standardized with frameworks and agreements on how to get, store, and use data before an outbreak. Public health agencies can be incentivized to share data and develop use cases to show the value of this data. At the same time, public health officials in low- and middle-income countries should be able to access research and use it to inform public health policies more easily. It’s especially important to share data during infectious outbreaks to help identify the focus of intervention, determine which interventions are working, and detect when pathogens change genetically.

Proposed recommendations include:

- **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.** The WHO can build on the existing influenza pandemic framework and expand it to other pathogens.
- **Share data during research through well-established organizations and institutions.** Examples: 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.
- **Share data during infectious outbreaks to help identify the focus of intervention, determine which interventions are working, and detect when pathogens change genetically.** Make data on biological agents more available and enable quick sharing of samples to allow better understanding and support vaccine development.

During outbreaks, sharing data is critical to public health.

# Materials Science

**R**esearch data in this field can accelerate the pace of discovering, manufacturing, and deploying advanced materials. 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. Roundtable participants in materials science suggested two approaches:

- Add data sharing plans into public governance models.
- Strengthen open data mandates by providing government resources and working to change the culture and status quo of research data.

The Materials Genome Initiative draws on many agencies to help discover, manufacture, and deploy advanced materials.

## Case Study, Materials Genome Initiative: Building materials science across agencies

The Materials Genome Initiative (MGI) is a multi-agency initiative designed to speed up and lower the cost of efforts to discover, manufacture, and deploy advanced materials. Despite being essential for “economic security and human well being, with applications in industries aimed at addressing challenges in clean energy, national security, and human welfare,”, the process of moving a new material to market can be excruciatingly slow. The MGI aims to change that by enabling a paradigm shift in culture; integrating experiments, computation, and theory; facilitating access to materials data; and equipping the next-generation materials workforce.<sup>29</sup>

To this end, there are no less than 14 projects underway aimed at facilitating materials data sharing and access. They are designed to “identify best practices for implementation of a materials data infrastructure [and] support creation of accessible materials data repositories.”<sup>30</sup> A diverse group of agencies is participating in specific efforts to ensure data sharing as well as ensuring the success of the MGI more broadly. Many have formed unique partnerships to achieve the initiative’s strategic goals. Some examples focused on data sharing include these:

- The Air Force Research Laboratory along with NIST and the National Science Foundation held a Materials Science and Engineering Data Challenge seeking “solutions for new uses of publicly accessible digital data to advance materials science and engineering knowledge to accelerate the transition to industrial applications.”<sup>31</sup>
- The Department of Commerce and NIST built a “Materials Data Curation System” that provides “a means for capturing, sharing, and transforming materials data into a structured format that is XML based [and] amenable to transformation to other formats.”
- The Department of Energy’s Office of Energy Efficiency and Renewable Energy “maintains a publicly accessible database of the material properties of hydrogen storage materials” designed to facilitate the development of hydrogen storage materials.

Materials Genome Initiative

mgi.gov

29. The Materials Genome Initiative: The First Five Years, [whitehouse.gov/blog/2016/08/01/materials-genome-initiative-first-five-years](http://whitehouse.gov/blog/2016/08/01/materials-genome-initiative-first-five-years) (accessed August 24, 2016).

30. Materials Genome Initiative, “Facilitate Access to Materials Data”, [mgi.gov/strategic-goals/facilitate-access-materials-data](http://mgi.gov/strategic-goals/facilitate-access-materials-data) (accessed July 1, 2016).

31. Materials Genome Initiative, “AFRL, NIST, and NSF announce materials Science and Engineering Data Challenge awardees”, [mgi.gov/content/afrl-nist-and-nsf-announce-materials-science-and-engineering-data-challenge-awardees](http://mgi.gov/content/afrl-nist-and-nsf-announce-materials-science-and-engineering-data-challenge-awardees) (accessed July 1, 2016).

# Oceans

**A**s scientists work to preserve the world's oceans, they face challenges in data access, discovery, quality, and ease of use. Oceans experts at the Roundtables identified several ways that key stakeholders can address these issues:

- **Private sector:** Provide expertise in ocean data formats and structures to increase cloud offerings; provide tools to facilitate metadata compliance; develop topical data aggregation platforms; enable cloud computing storage.
- **Federal government:** Connect experts from oceans sector to marine renewable energy data; support and maintain Quality Assurance for Real Time Ocean Data (QARTOD); support agencies in implementing open data policy; enforce open data and open science policies.
- **State/local governments:** Provide support for Northeast Regional Ocean Data portal; develop citizen engagement initiatives focused on specific topics.
- **Nonprofits, academics, developers:** Collect and monitor data about fishing and fisheries; empower coastal communities for climate change resilience with decision support tools (e.g. [coastalresilience.org](http://coastalresilience.org)); map marine-protected areas; conduct outreach to ocean science community through professional organizations and conferences; curate datasets for oceans challenges.

Global  
Fishing  
Watch  
enables  
anyone with  
an internet  
connection  
to monitor  
commercial  
fishing - and  
potential  
overfishing  
- around the  
world.

## Case Study, Global Fishing Watch: Building healthy fisheries through open data

**Global Fishing Watch** is the result of a technology partnership between Google, SkyTruth, and Oceana. It is designed to show all of the trackable fishing activity in our oceans and is presented so “anyone with an internet connection [can] monitor when and where commercial fishing is happening around the globe. Citizens can use the tool to see for themselves whether their fisheries are being effectively managed. Seafood suppliers can keep tabs on the boats they buy fish from. Media and the public can act as watchdogs to improve the sustainable management of global fisheries. Fisherman can show that they are obeying the law and doing their part. Researchers will have access to a multi-year record of all trackable fishing activity.”

Global Fishing Watch relies on Automatic Identification System (AIS) location data that is broadcast by ships and collected by satellite. Using this data, observers can follow individual ships, track entire fleets, and look for global patterns. Many governments and regulatory bodies require fishing vessels to use AIS systems, in order to provide consistent and trustworthy data.

Even bad actors attempting to avoid scrutiny will turn on their AIS systems for safety reasons when they are close to port or in heavily trafficked areas. A larger number of eyes scrutinizing AIS data makes it more likely that the strange behavior of these outliers will be noticed and reported. The partnership behind Global Fishing Watch envisions a future with lower stress on our fisheries and fewer illegal fishing operations.



[globalfishingwatch.org](http://globalfishingwatch.org)

# Smart Cities

**U**rban populations are growing rapidly across the globe, providing economic and environmental opportunities but also ushering in new challenges. Meanwhile, as the White House has written, “technology is creating new opportunities to reduce traffic congestion, fight crime, foster economic development, reduce greenhouse gases, and make local governments more open, responsive, and efficient.” However, new technologies alone are unlikely to solve these problems without community cooperation and coordination.<sup>32</sup>

The Smart Cities Initiative is designed to harness “the growing data revolution, low-cost sensors, and research collaborations” to support community efforts to solve everyday problems. The initiative includes significant investment by federal partners like the National Science Foundation, the National Institute of Standards and Technology, the Department of Homeland Security, the Department of Transportation, and the Department of Commerce. The initiative also includes investment by several private-sector partners.<sup>33</sup>

To be successful, Smart Cities initiatives need to create data-sharing ecosystems within cities that run efficiently and securely while protecting privacy. As one model, the Department of Transportation’s Smart City Challenge has just awarded \$40 million to Columbus, Ohio to test smart city technologies. The city will develop a data management plan as part of the award.<sup>34</sup> The plan will address “how data will be collected, managed, integrated, and disseminated before, during, and after” the pilot. Ideally, lessons learned from this program will be applicable to many urban planning challenges, not just transportation.

Smart cities will serve as laboratories generating data and insights for everything from reducing energy use to fighting traffic congestion or improving air and water quality. Shar-

ing research data will be central to all these goals. Some notable projects under the Smart Cities Initiative banner include a Smart Grid Integration Challenge for Cities, offering at least \$1 million in funding from the Department of Energy’s Office of Electricity Delivery and Energy Reliability, and an expansion of the U.S. Census Bureau’s open-source CitySDK project featuring a “pilot program of data innovation workshops delivered in close collaboration with city experts to help solve the most pressing local issues.” To support data-driven local energy planning, the Department of Energy recently awarded nearly \$1.3 million to cities as part of their Cities Leading Through Energy Analysis and Planning (Cities-LEAP) project.<sup>35</sup>

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. Cities operate differently and one city’s data is often not comparable to another’s. Within a city, data is often not interoperable between departments. Additionally, some states have legal or copyright constraints on opening their data or may be hampered by proprietary systems. Finally, while the federal government has data relevant to cities, agencies are not generally set up to work well with cities on data issues.

At several of the Roundtables, experts on smart cities explored ways to manage the data that is central to improving city use of data to improve quality of life. One recommendation for a new approach is to create data repositories that would give cities the capacity to manage data more effectively. 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.

At present, there are several approaches that help advance data-driven smart cities work. Roundtable participants suggested that smart cities programs can:

- **Support research across geographies.** Share data and research internationally.
- **Create incentives for university-city pairs.** Institutionalize these partnerships along with regional partnerships.
- **Use common information models (like the transportation XML standard) and simplified metadata standards.**
- **Create other data-sharing tools for cities,** including:
  - Infrastructure for secure data access, sharing, and combining
  - Support for bulk/shared data purchase
  - Tools for local use of global international datasets

The cities experts at the Roundtables also identified a number of data issues that can be addressed through collaboration, with contributions from different groups of stakeholders:

- **Private sector:** Integrate streaming with tabular data for comparative analytics; facilitate easier sharing of private-sector data; develop partnerships to scale up individual communities' programs; help scientific community produce basic information products; organize startup showcase for White House Open Data Summit.
- **Federal government:** Create more programs and platforms to make federal data applicable at a local level (e.g. CitySDK); develop smart data grid standards; use DOT Smart Cities grant winner, Columbus, Ohio, as a model for engaging small businesses around open data; coordinate relevant funding and technical assistance programs across agencies.
- **State/local government:** Build on the USDS eight-state initiative; reduce redundancy in data collection; develop catalogs of sensor network data; develop tools, analytics, and visualizations and case studies that showcase open data use.
- **Nonprofits, academics, developers:** Identify best practices and lessons learned for publishing data that users can easily discover and understand; promote best practices for web analytics that encourage data use and increase ROI on data investments.

## Tech and data programs can improve traffic, public safety, green energy, and more.

32. The White House, "Launching a Smart Cities Initiative to Tackle City Challenges with Innovative Approaches", September 16, 2015, [whitehouse.gov/blog/2015/09/16/launching-smart-cities-initiative-tackle-city-challenges-innovative-approaches](http://whitehouse.gov/blog/2015/09/16/launching-smart-cities-initiative-tackle-city-challenges-innovative-approaches) (accessed July 1, 2016).
33. The White House, "Fact Sheet: Administration Announces New 'Smart Cities' Initiative to Help Communities Tackle Local Challenges and Improve City Services", September 14, 2015, [whitehouse.gov/the-press-office/2015/09/14/fact-sheet-administration-announces-new-smart-cities-initiative-help](http://whitehouse.gov/the-press-office/2015/09/14/fact-sheet-administration-announces-new-smart-cities-initiative-help) (accessed July 1, 2016).
34. U.S. Department of Transportation, "U.S. Department of Transportation Announces Columbus as Winner of Unprecedented \$40 Million Smart City Challenge", June 23, 2016, [transportation.gov/briefing-room/us-department-transportation-announces-columbus-winner-unprecedented-40-million-smart](http://transportation.gov/briefing-room/us-department-transportation-announces-columbus-winner-unprecedented-40-million-smart) (accessed July 7, 2016).
35. U.S. Department of Energy Office of Energy Efficiency & Renewable Energy, "Cities Leading through Energy Analysis and Planning", [energy.gov/eere/cities-leading-through-energy-analysis-and-planning](http://energy.gov/eere/cities-leading-through-energy-analysis-and-planning)

# Soil and Agriculture

**S**oil and agriculture data has gotten increased attention through the Global Open Data for Agriculture and Nutrition initiative (GODAN) and the development of “precision agriculture” by private companies. Researchers in this area also face challenges. Limited funding impacts the quality of data infrastructure, delivery, and maintenance. Privacy issues also pertain to soil data. Some researchers are required to remove location information before sharing data, which limits its value. The Roundtable participants in this field suggested several recommendations and steps to take:

- Create public buy-in to encourage private landowners to share data.
- Use common metadata standards like DCAT (Data Catalog Vocabulary) to make it possible to discover and aggregate data from diverse repositories.
- Develop collaborations for multi-state, multi-university projects.
- Build agency-based repositories for grantee-generated research and evaluation data, and make 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.

**The GODAN initiative and “precision agriculture” are new, data-driven approaches to help feed the world.**

## Case Study, Global Open Data for Agriculture and Nutrition: Sharing data to ensure food security and feed the hungry

The Global Open Data for Agriculture and Nutrition (GODAN) Initiative “supports the proactive sharing of open data to make information about agriculture and nutrition available, accessible and usable for unrestricted use worldwide to deal with the urgent challenge of ensuring world food security.” GODAN is a growing group of over 350 partners representing governments, donors, international and not-for-profit organizations, and businesses, who have committed to a joint statement of purpose to promote the proactive sharing of open data to make information about agriculture and nutrition available to all. Partner organizations hail from every continent (except Antarctica). GODAN is convinced that the solution to Zero Hunger lies within existing, but often unavailable, agriculture and nutrition data. The goal is to make all agriculture and nutrition data open for better policy and decision making to ensure no one goes hungry.

By building high-level support among policymakers and diverse stakeholders, GODAN promotes collaboration among partners to harness the growing volume of data generated by new technologies. GODAN welcomes any organization that supports open access to agriculture and nutrition data, and creates opportunities for partners to develop open data resources by hosting events and conversations, conducting research, and convening working groups around specific topics

The latest of these meetings, the 2016 GODAN Summit, occurred on September 15–16, 2016 in New York City. With addresses from Ministers of Agriculture, diverse presentations and exhibits from open data practitioners, an Open Data Challenge, and a hackathon, the GODAN Summit was the largest open data summit supporting the agriculture and nutrition fields ever held.<sup>36</sup> Through venues such as these, GODAN continues to lead and petition for the role of open data as a powerful tool that supports sustainable development.<sup>37</sup>



[godan.info](http://godan.info)

36. Global Open Data for Agriculture and Nutrition Summit, [summit.godan.info](http://summit.godan.info) (accessed September 1, 2016).

37. GODAN Petition, [summit.godan.info/petition](http://summit.godan.info/petition) (accessed September 1, 2016)

## Case Study, The USDA-Microsoft Innovation Challenge: Finding ways to increase food resilience

**Many government agencies** have used data challenges to spark innovation. They release key datasets in an accessible and usable form, invite the public to use the data for a particular public purpose, and award and publicize the best solutions. In 2015, Microsoft and the USDA partnered to use a data challenge for a major national goal: Improving food resilience, the ability of the American food supply to withstand climate change.

For this challenge, the USDA released more than 100 years of crop and climate data through Microsoft Azure, Microsoft's cloud computing platform. The release included data on the farm economy, production, and the health of crops around the country. To make the data accessible and usable, Microsoft designed an entire portal where multiple datasets could be found together and queried in common languages.

The challenge, which awarded \$63,000 in cash and prizes, was created in support of the President's Climate Data Initiative. It simultaneously helped develop new approaches to food resilience, demonstrated a model for public-private collaboration around data, and showed how combining cloud computing resources with government data can have a positive impact.

The winning solutions included apps to show farmers what is being grown in nearby farms to help them evaluate their own planting strategy; a dashboard to help farmers visualize production, economic, livestock, and commodity data; a tool to analyze the resources needed to meet specific agricultural goals; and a tool to visualize crowd-sourced pricing data from around the world. The winners came from all over the country, including California, Tennessee, Washington state, Nebraska, and Brooklyn, New York – a demonstration of the nationwide talent that can be tapped to solve public problems with data.





---

# Appendices

[Appendix A: Additional Resources](#)

[Appendix B: About the Center](#)

[Appendix C: Program Support](#)

[Appendix D: Roundtable Agendas](#)

[Appendix E: Participants](#)

[Appendix F: Acknowledgements](#)

# Appendix A: Additional Resources

The Center for Open Data Enterprise has published a number of resources that supplement this report at [opendataenterprise.org/convene](http://opendataenterprise.org/convene). These include:

- **Interagency Roundtables: Briefing Papers.** In preparation for each of the 2016 Open Data Roundtables, the Center prepared a briefing paper to frame discussion and provide background and resources on the issue under discussion. These four Briefing Papers, on privacy, quality, sharing research data, and public-private collaboration, are designed to be stand-alone references for anyone interested in working on these open data issues.
- **Interagency Roundtables: Key Takeaways.** These brief summaries recap each individual Roundtable and key points of discussion.
- **Federal Agency Roundtables.** The website includes agendas and extensive reports from each of eight Roundtables that the Center's team held with federal agencies from mid-2014 through 2015. Like the 2016 Roundtables, these events brought together government data providers with their data users to identify challenges in using open government data and approaches to make the data more accessible and usable. Rather than focus on specific data issues like privacy or quality, these Federal Agency Roundtables explored the range of issues that each Agency or Office faces. The federal organizations the Center partnered with, in order, included:
  - U.S. Department of Commerce
  - U.S. Department of Agriculture
  - U.S. Patent and Trademark Office
  - U.S. Department of Transportation
  - U.S. Department of Energy
  - U.S. Department of Veterans Affairs
  - U.S. Department of the Treasury
  - U.S. Department of Labor

We encourage interested readers to use these resources for more background, insight, and ideas on the Open Data Roundtables and their results. Like all of the Center's publications, they are published under a Creative Commons Share-Alike License.

# Appendix B: About the Center for Open Data Enterprise

The Center for Open Data Enterprise is an independent nonprofit organization that develops smarter open data strategies for governments, businesses, and nonprofits by focusing on data users. Our mission is to maximize the value of open data as a public resource.

Our Open Data Roundtables in the U.S. and abroad help identify high-value datasets, find solutions to data problems, and establish new collaborations. We work with public and private sector partners to develop solutions to key data challenges informed by user feedback.

## CONTACT US

For general inquiries, contact Katherine Garcia at [katherine@odenterprise.org](mailto:katherine@odenterprise.org).

For partnership opportunities, contact Laura Manley at [laura@odenterprise.org](mailto:laura@odenterprise.org).

Learn more at [OpenDataEnterprise.org](http://OpenDataEnterprise.org).

# Appendix C: Program Support

The Center for Open Data Enterprise thanks our Open Data Partner, **Microsoft** and our Open Data Supporter, **Booz Allen Hamilton** for supporting the Center's work on the Open Data Roundtables.

## OPEN DATA PARTNER

Microsoft is a worldwide supplier of devices and services that help people and businesses realize their full potential. Customers use Microsoft's products to find creative solutions to business problems, develop breakthrough ideas, and stay connected to what's most important to them. For more information, see [microsoft.com](http://microsoft.com).



## OPEN DATA SUPPORTER

Booz Allen Hamilton provides management and technology consulting and engineering services to leading Fortune 500 corporations, governments, and not-for-profits across the globe. Booz Allen partners with public and private sector clients to solve their most difficult challenges through a combination of consulting, analytics, mission operations, technology, systems delivery, cybersecurity, engineering, and innovation expertise. For more information, visit [boozallen.com](http://boozallen.com).

Booz | Allen | Hamilton

# Appendix D: Open Data Roundtable Agendas

## Open Data and Privacy Roundtable

March 24, 2016 | 12:30 PM – 6:00 PM

### Welcome

**Kristen Honey**  
*Policy Advisor, OSTP*  
AND  
**Nancy Weiss**  
*Senior Advisor to the CTO, OSTP*

### Opening Keynote

**Megan Smith**  
*U.S. Chief Technology Officer, OSTP*  
WITH  
**Alexander Macgillivray**  
*U.S. Deputy Chief Technology Officer, OSTP*

### 21st Century Government with Responsible Open Data

**Erie Meyer,**  
*Digital Services Expert, U.S. Digital Services,  
White House Office of Management and Budget (OMB)*

### Lightning Talks

**Simson Garfinkel**  
*Senior Advisor, National Institute of Standards and Technology*  
**Ren Essene**  
*Program Manager, Consumer Financial Protection Bureau*

**Melissa Goldstein**  
*Assistant Director, Precision Medicine and Bioethics, OSTP*

### Structure of the Roundtable

**Joel Gurin**  
*President and Founder, Center for Open Data Enterprise*

#### Breakout session 1:

**Goals and Approaches**  
Followed by Report Out

#### Breakout session 2:

**Exploring Solutions**  
Followed by Report Out

#### Breakout session 3:

*Cross-Agency Lessons Learned and Takeaways*

### Highlights Presented to White House Leadership

### Closing Keynotes

**Tony Scott**  
*U.S. Chief Information Officer, OMB*  
**Marc Groman**  
*Senior Advisor to the Director for Privacy, OMB*

### Discussion and Next Steps

**Jason Schultz**  
*Senior Advisor, OSTP*  
AND  
**Jerry Sheehan**  
*Assistant Director, Scientific Data and Information, OSTP*

# Open Data and Improving Data Quality Roundtable

April 27, 2016 | 1:00 PM – 6:00 PM

## Welcome

**Kristen Honey**

*Policy Advisor, OSTP*

AND

**Jerry Sheehan**

*Assistant Director, Scientific Data and Information, OSTP*

**Mike Hulme**

*Co-Chair, NIEM Technical Architecture Committee*

## Opening Keynotes

**Megan Smith**

*U.S. Chief Technology Officer, OSTP*

**Tom Kalil**

*Deputy Director for Technology and Innovation, OSTP*

## Structure of the Roundtable

**Erie Meyer,**

*Digital Services Expert, U.S. Digital Services,  
White House Office of Management and Budget (OMB)*

## Lightning Talks

**Joel Gurin**

*President and Founder, Center for Open Data Enterprise*

## Breakout session 1:

**Goals and Approaches**

**Tables Report Out**

**Mike Hulme**

*Co-Chair, NIEM Technical Architecture Committee*

## Breakout session 2:

**Exploring Solutions**

**Followed by Report Out**

## Breakout session 3:

**Cross-Agency Lessons Learned and Takeaways**

**Followed by Report Out**

## Highlights Presented to White House Leadership

## Closing Keynotes

**Margie Graves**

*Acting Deputy Federal Chief Information Officer, White House  
Office of Management and Budget*

**DJ Patil**

*Deputy Chief Technology Officer for Data Policy and  
Chief Data Scientist, OSTP*

## Next Steps

**Nancy Weiss**

*Senior Advisor to the CTO, OSTP*

AND

**Katherine Garcia**

*Communications Manager, Center for Open Data Enterprise*

## #OpenData Bright Spots and Solutions

**David Portnoy**

*Fellow and Founder of Demand-Driven Open Data, HHS IDEA Lab*

**Alan Marco**

*Chief Economist, U.S. Patent and Trademark Office*

## Open Data Roundtable Agendas

---

# Open Data Roundtable for Sharing and Applying Research Data

May 25, 2016 | 12:30 PM – 6:00 PM

**Open Data Table Topics Featuring:** Arctic, Cancer/Biomedical, Climate, Health/My Data, Infectious Disease, Materials Science, Smart Cities/Transportation/Energy, Soil/Agriculture

### Welcome

**Kristen Honey**  
Policy Advisor, OSTP

AND

**Jerry Sheehan**

Assistant Director, Scientific Data and Information, OSTP

### Opening Keynotes

**Cristin Dorgelo**  
Chief of Staff, OSTP

**Greg Simon**

Executive Director of the Cancer Moonshot Task Force,  
Office of the Vice President

### Structure of the Roundtable

**Kristen Honey**  
Policy Advisor, OSTP

AND

**Joel Gurin**

President and Founder, Center for Open Data Enterprise

### Breakout session 1:

Goals and Challenges  
Followed by Report Out

### #OpenData Bright Spots and Solutions

**Ariel Gold**

Data Program Manager, Intelligent Transportation Systems,  
U.S. Department of Transportation

**Philip Bourne**

Associate Director for Data Science, National Institutes of Health

### Breakout session 2:

Exploring Solutions  
Followed by Report Out

### #OpenData Bright Spots and Solutions

**Martin Jeffries**

Assistant Director, Polar Sciences, OSTP

### Breakout session 3:

Cross-Agency Proposals and Takeaways

### Open Data for Open Science, Knowledge, Action, and Impact

**Jo Handelsman**

Associate Director for Science, OSTP

### Presentation of Highlights

### Closing Keynote

**Megan Smith**

U.S. Chief Technology Officer, OSTP

### Next Steps

**Nancy Weiss**

Senior Advisor to the CTO, OSTP

AND

**Joel Gurin**

President and Founder, Center for Open Data Enterprise

## Open Data Roundtable Agendas

### Open Data Roundtable for Public-Private Collaboration

June 15, 2016 | 12:00 PM – 4:00 PM

**Open Data Table Topics Featuring:** Open Data Big Think, Climate Big Think, Oceans Big Think, Smart Cities Big Think

#### Welcome

**Kristen Honey**

*Policy Advisor, OSTP*

#### Opening Keynote

**John Holdren**

*Assistant to the President for Science and Technology,  
Director, OSTP*

#### Structure of the Roundtable

**Kristen Honey**

*Policy Advisor, OSTP*

AND

**Joel Gurin**

*President and Founder, Center for Open Data Enterprise*

#### Breakout session 2:

##### Opportunities and Possibilities

Followed by Report Out

#### Presentation of Highlights

#### Closing Keynote

**Amy Luers**

*Assistant Director, Climate Resilience and Information, OSTP*

#### Next Steps

**Joel Gurin**

*President and Founder, Center for Open Data Enterprise*

### #OpenData Bright Spots and Solutions

**Jeff de La Beaujardiere**

*Data Management Architect, National Oceanic and Atmospheric Administration (NOAA)*

**Jeremy Roberts**

*General Manager, Green Button Alliance*

**Alex Dehgan**

*Co-founder, Conservation X Labs*

**Katie Matthews**

*Deputy Chief Scientist, Oceana*

#### Breakout session 1:

##### Wicked Problems and Next-Gen Solutions

Followed by Report Out

# Appendix E: Participating Organizations

## COMPANIES, NONPROFIT, ACADEMIC AND RESEARCH ORGANIZATIONS

Participants are noted by the roundtables they attended in 2016 (**RT1** is the Open Data and Privacy Roundtable, **RT2** is the Open Data and Improving Data Quality Roundtable, **RT3** is the Open Data Roundtable for Sharing and Applying Research Data, **RT4** is the Open Data Roundtable for Public-Private Collaboration).

**Amazon Web Services** provides cloud computing services to a range of clients.

Chris Gorski, Solutions Architect, **RT2**  
John Stephenson, Manager, **RT1**  
Jed Sundwall, Global Open Data Lead, **RT3**  
Lena Trudeau, Practice Manager, **RT4**

**Amida Technology Solutions** develops patient-centered solutions based on the Blue Button personal health record (PHR).

Peter Levin, Co-Founder and CEO, **RT1**

**The Anthropocene Institute** partners with private companies, academic institutions, NGO's, and government agencies to nurture new technologies and cultural institutions required for the long term viability of humanity and preservation of biological diversity.

Virgil Zetterlind, Manager, **RT4**

**Association of Public Data Users** is a national network that links users, producers and disseminators of government statistical data.

Katherine Pettit, Senior Research Associate, **RT2**

**The Association of Research Libraries** influences the changing environment of scholarly communication and the public policies that affect research libraries and the diverse communities they serve.

Caile Morris, Law & Policy Fellow, **RT1**

**Autodesk** is an American multinational software corporation that makes software for the architecture, engineering, construction, manufacturing, media, and entertainment industries.

Jonathan Knowles, Explorer in residence, **RT4**

**Booz Allen Hamilton** is a leading provider of management consulting, technology, and engineering services to the US government in defense, intelligence, and civil markets, and to major corporations, institutions, and not-for-profit organizations.

RK Paleru, Principal, **RT2**  
Bryce Pippert, Principal, **RT1, RT3, RT4**  
John Sandoval, Lead Associate, **RT4**  
Sahil Sanghvi, Senior Associate, **RT1, RT2**  
Jin Shao, Principal, **RT2**  
Greg Wenzel, Executive Vice President, Strategic Innovation Group (SIG) Digital Initiative, **RT1**

**Boston Children's Hospital**, one of the largest pediatric medical centers in the United States, offers a complete range of health care services for children from birth through 21 years of age.

Gaurav Tuli, Postdoctoral Research Fellow, **RT3**

**The Bridge@USC** is an initiative that brings together some of the most creative minds in chemistry, biology, medicine, mathematics, engineering, physics and nanosciences – as well as experts in such areas as animation and cinematography – to build the first atomic resolution structure of man.

Peter Kuhn, Professor, **RT4**

**Caprivity** is a cloud-native data management platform that captures and converts data generated by paper forms – including handwritten – into business-ready information.

Brian Busch, Public Sector Lead, **RT2**

**The Center for Advanced Transportation Technology (CATT) Laboratory** at the University of Maryland deploys innovative technology and user-centered design of software and information visualization sys-

tems to support national, state, and local efforts to solve important transportation, safety, and security problems.

Michael Pack, Director, **RT4**

**Center for Data Intensive Science**'s mission is to pioneer translational data intensive science to advance biology, medicine, healthcare and the environment.

Robert Grossman, Director, **RT4**

**Center for Democracy & Technology** is a champion of global online civil liberties and human rights, driving policy outcomes that keep the Internet open, innovative, and free.

Michelle De Mooy, Deputy Director, Privacy & Data Project, **RT1**

**The Center for Open Data Enterprise** is an independent nonprofit organization that develops smarter open data strategies for governments, businesses, and other nonprofits by focusing on data users.

Audrey Ariss, Director of Research & Design, **RT1, RT2, RT3, RT4**  
Joohee Choi, Research Intern, **RT4**  
Katherine Garcia, Communications Director, **RT1, RT2, RT3, RT4**  
Theo Goetemann, Research Intern, **RT4**  
Joel Gurin, President & Founder, **RT1, RT2, RT3, RT4**  
Stephanie Huang, Research Intern, **RT1, RT2, RT3, RT4**  
Laura Manley, Director of Partnerships and Programs, **RT1, RT2, RT4**  
Pooja Singh, Tech Fellow, **RT4**  
Hanlin Song, Tech Fellow, **RT4**  
Vinayak Pande, Tech Fellow, **RT4**

**The Children's Hospital of Philadelphia** is the nation's first hospital devoted exclusively to the care of children. The Hospital has fostered medical discover-

## Participating Organizations

### COMPANIES, NONPROFIT, ACADEMIC AND RESEARCH ORGANIZATIONS

ies and innovations that have improved pediatric healthcare and saved countless children's lives.

Jena Lilly, Director of Operations and Strategic Planning, RT2

Adam Resnik, Director of Center for Data Driven Discovery in Biomedicine, RT2

**Citrine Informatics** provides a machine learning-based platform that mines vast quantities of data about materials, chemicals, and processes to help clients reach R&D and manufacturing targets in half the time.

Greg Mulholland, Chief Operating Officer, RT3  
Christina Hartman, Business Manager, RT3

**Clarkson University** offers over 50 programs of study in engineering, business, arts, education, sciences and the health professions to build powers of observation, challenge the status quo and connect discovery and innovation with enterprise.

Michelle Meyer, Assistant Professor of Bioethics, RT1

**Climate Data Solutions** has the expertise to help businesses and communities better formulate their climate problems, design solutions that minimize risk and enhance opportunities, and identify and access the right data and tools for the tasks.

Ana Pinheiro Privette, Chief Data Scientist, RT2

**Cloudera** is a provider of Apache Hadoop-based software and services and works with customers in financial services, web, telecommunications, government and other industries.

Suzanne McIntosh, Computer Scientist, RT2, RT4

**Conservation X Labs** harnesses exponential technologies, open innovation, and a for-profit entrepreneurship financial model to improve the efficacy, cost, speed, sustainability, and scale of conservation efforts.

Paul Bunje, Co-founder, RT4

Alex Dehgan, Co-founder, RT4

Barbara Martinez, Open Innovation Director, RT4

Clare Fiese, Fellow, RT4

**Consortium of Universities for the Advancement of Hydrologic Science Inc (CUAHSI)** represents over 130 U.S. universities and international water science-related organizations to develop infrastructure and services for the advancement of water science in the United States.

Jon Pollak, Program Manager, RT2

**The D'Amore-McKim School of Business at Northeastern University** integrates strong academics with authentic experiential learning opportunities to meet the evolving needs of today's business student and the global marketplace.

Yang W Lee, Professor, RT2

**Dana-Farber/Broad Institute** is a community of scientists working at the interface of computational cancer biology and experimental cancer biology.

Stephanie Mullane, Senior Data Research Specialist, RT3

**Data Blueprint** is a data management consulting firm that uses core competencies in both foundational and technical data management practices to help clients leverage data as a strategic investment and become data-driven.

Catheryn Denton, Vice President of Finance, RT4

**The Data Coalition** advocates on behalf of the private sector and the public interest for the publication of government information as standardized, machine-readable data.

Christian Hoehner, Director of Policy, RT2  
Hudson Hollister, Founder, RT2

**Data Quality Campaign** is a national, nonprofit organization leading the effort to bring every part of the education community together to empower educators,

parents, and policymakers with quality information to make decisions that ensure students excel.

Rachel Anderson, Senior Associate, Policy and Advocacy, RT1

**Deloitte's** dedicated professionals in independent firms throughout the world collaborate to provide audit, consulting, financial advisory, risk management, tax and related services to select clients.

Will Sarni, Director and Practice Leader, RT2  
Akshai Prakash, Deloitte Digital Manager, RT4

**Descartes Labs** uses satellite imagery and analysis to enable real-time global awareness in areas such as food production, energy infrastructure, the growth of cities, and the environment and assesses Earth's changes over time.

Steven Brumby, Chief Technology Officer, RT3

**Drinker Biddle** serves clients in matters ranging from billion-dollar deals to complex class actions, across a broad spectrum of industries.

Jason R. Baron, Of Counsel, RT1

**The Earth Genome** connects the world's best earth data, scientists and technology providers to drive insight and innovation for governments, NGOs, corporations and investors.

Dan Hammer, Chief Data Scientist, RT2

**Elsevier Research Data Management** publishes academic journals and books about discoveries in science, health, and technology to provide information solutions for better decision-making.

Anita de Waard, Vice President for Research Data Collaborations, RT3

**Enigma** helps organizations and individuals fuse, organize, and explore data to make smarter decisions.

Eve Ahearn, Data Project Manager, RT2

## Participating Organizations

### COMPANIES, NONPROFIT, ACADEMIC AND RESEARCH ORGANIZATIONS

**ESRI** provides GIS (geographic information systems), a mapping software that visualizes, questions, analyzes, and interprets data to understand relationships, patterns, and trends.

Pat Cummins, Government Policy & Strategy, **RT3**  
 Lauren Lipovic, GIS and Policy Making, **RT4**  
 Andrew Turner, Chief Technology Officer, **RT4**

**The Federation of Earth Science Information Partners (ESIP)** is a nonprofit organization that advances the use of Earth science data through collaborations around topics such as data stewardship, information technology and interoperability, and application areas like disaster response, climate, energy and agriculture

Erin Robinson, Executive Director, **RT3**

**Forum One** is a full service digital agency that works with mission-driven organizations to create stunning designs, smart messaging, and custom built technology tools.

Chris Wolz, CEO, **RT4**

**General Dynamics Information Technology** provides information technology (IT), systems engineering, professional services and simulation and training to customers in the defense, federal civilian government, health, homeland security, intelligence, state and local government and commercial sectors.

Debbie Dean, Senior Director of Research Services, **RT4**

**Genetic Alliance** defines the intersection of health and genetics. Originally founded as an alliance for support groups, their work has evolved along with the growing health advocacy movement and the rapid advancement of genetic technology.

Sharon Terry, President & CEO, **RT1**

**Georgetown University** is an academic and research institution that offers an educational experience that prepares the next

generation of global citizens to lead and make a difference in the world.

Theo Goettmann, Student, **RT1**

**Google's** technology and continued innovation serve the company's mission of "organizing the world's information and making it universally accessible and useful."

Jennifer Austin, **RT4**  
 Valliappa Laksmanan, Program Manager, **RT2**

**Green Button Alliance** is a non-profit U.S. corporation formed in 2015 to foster the development, compliance, and widespread adoption of the Green Button standard.

Jeremy Roberts, General Manager, **RT4**

**The Harvard John A. Paulson School of Engineering and Applied Sciences (SEAS)** works to create collaborative bridges across Harvard and educate the next generation of global leaders.

Ben Green, PhD Candidate, **RT1**

**Harvard Medical School's** mission is to create and nurture a diverse community of the best people committed to leadership in alleviating human suffering caused by disease

Ateev Mehrotra, Associate Professor, **RT3**

**Hortonworks** creates, distributes and supports enterprise-ready open data platforms and modern data applications that deliver actionable intelligence from all data: data-in-motion and data-at-rest. Hortonworks provides the expertise, training and services that allow customers to unlock transformational value for their organizations across any line of business.

Henry Sowell, Technical Director, **RT2**

**IBM** is a multinational corporation offering a range of business consulting services and technological products.

Steve Adler, Chief Data Strategist, **RT4**  
 Chid Apte, Director, **RT4**

**IDA's Science and Technology Policy Institute** provides rigorous and objective analysis for the formulation of national science and technology policy, supporting the White House Office of Science and Technology Policy, the National Science Foundation, the National Science Board, and other offices and councils within the executive branch of the federal government.

Emily Sylak-Glassman, Science and Technology Policy Analyst, **RT2, RT3, RT4**

**IMS Health** is a global information and technology services company providing clients in the healthcare industry with end-to-end solutions to measure and improve their performance.

Kimberly Gray, Chief Privacy Officer, Global, **RT1**

**Indiana University** is a multi-campus public research institution, grounded in the liberal arts and sciences, and a world leader in professional, medical, and technological education.

David Fidler, James Louis Calamaras Professor of International Law, **RT3**

**Kaggle** is a platform for data-related competitions. The platform allows companies, researchers, government and other organizations to post their modeling problems and have data professionals and researchers compete to produce the best solutions.

Anthony Goldbloom, Co-founder & CEO, **RT2**

**The Laboratory for Advanced Computing (LAC)** is a research group at the University of Chicago that focuses on big data and its applications, including applications in biomedical informatics, information sciences and imaging.

Maria Patterson, Research Scientist, **RT4**

## Participating Organizations

### COMPANIES, NONPROFIT, ACADEMIC AND RESEARCH ORGANIZATIONS

**Link Digital** is a full service digital agency delivering high quality design, consulting and development services to Government and the private sector to promote government transparency and a global market.

Steven De Costa, Executive Director, RT4

**Massive Connections** specializes in the art and science of Identifying relationships between actors, human and otherwise, in complex systems; architecting strategic initiatives that expertly weave through science, technology, and policy; strategizing how data can be used to understand and manage complex social-ecological systems; and engaging government, NGO, and private sector constituents in crafting solutions to wicked problems.

Brian Wee, Founder and Managing Director, RT3

**Mayo Clinic** aims to inspire hope and contribute to health and well-being by providing the best care to every patient through integrated clinical practice, education and research.

Veronique Roger, MD, MPH, RT1, RT3

**Mercercomp LLC** provides accounting and technology services focusing on XBRL and open data implementation, structured data and data quality.

Ronald Schechter, Consultant, RT4

**Microsoft** develops, manufactures, licenses, supports and sells computer software, services, devices and solutions that help people and businesses realize their full potential.

James Dugan, Regional Architect, RT1, RT2

Jeffrey Friedberg, Chief Trust Architect, RT1

Elizabeth Grossman, Tech & Civic Engagement, RT1, RT2, RT3, RT4

Michael Pizzo, Principal Software Architect, RT2

**The MIT Information Quality Program** equips professionals with the understanding and means to significantly improve

their organization's information and to use that information as a core strategic tool.

Richard Wang, Director, MIT Chief Data Officer and Information Quality Program, RT2

**The Moore Foundation** is a private grant making organization that fosters path-breaking scientific discovery, environmental conservation, patient care improvements, and preservation of the special character of the San Francisco Bay Area.

Chris Mentzel, Program Director, RT3

**mySociety** is a not-for-profit social enterprise, based in the UK and working internationally. It builds online technologies that give people the power to get things changed, and shares these technologies so that they can be used anywhere.

Emily Shaw, US Civic Technologies Researcher, RT1

**The Nature Conservancy** is a conservation organization working around the world to protect ecologically important lands and waters for nature and people.

Jay Odell, Mid-Atlantic Marine Program Director, RT4

**National Opinion Research Center (NORC) at the University of Chicago** is an independent research institution that delivers reliable data and rigorous analysis to guide critical programmatic, business, and policy decisions.

Tim Riddle, Director, RT4

**NuCivic** provides full-service open source software-as-a-service solutions to public sector institutions around the world and are the lead developers of the open source open data platform.

Andrew Hoppin, President, RT4

**New York University Center for Data Science** is a leader in educating researchers and professionals to harness the power of big data and open data. Owing to the

interdisciplinary nature of data science, the Center collaborates on data science projects in a wide range of research areas.

Suzanne McIntosh, Adjunct Professor, RT2, RT4

**Oceana** is an international organization that focuses solely on oceans, dedicated to achieving measurable change by conducting specific, science-based campaigns with fixed deadlines and articulated goals.

Katie Matthews, Deputy Chief Scientist, RT4  
Jacqueline Savitz, Vice President for U.S. Oceans, RT4

**OceanElders** is an independent group of global leaders who have joined together to serve as a catalyst in the conservation and protection of the ocean and its wildlife.

Gigi Brisson, Founder & CEO, RT4

**Open Commons Consortium** is a not for profit that manages and operates cloud computing and data commons infrastructure to support scientific, medical, health care and environmental research.

Walter Wells, Director of Operations, RT4

**Open Data Nation** creates plug-and-play, productivity-based solutions that combine open, public data with data science techniques to increase transparency and productivity of public agencies.

Carey Nadeau, Founder and CEO, RT4

**OpenDataSoft** is a cloud-based turn-key platform for data publishing and API management, designed for business users, to allow for innovation to be easily built around data.

Jason Hare, Open Data Evangelist, RT2, RT4

**OpenGov** is setting a new standard across the country for how governments analyze, share, and compare financial data.

Joel Natividad, Director of Open Data, RT4

## Participating Organizations

### COMPANIES, NONPROFIT, ACADEMIC AND RESEARCH ORGANIZATIONS

**Open Health Systems Laboratory**, located at the Johns Hopkins University, is an independent non-profit organization that encourages and incubates ideas for innovative problem solving in the area of life sciences research through collaborations.

Kabeer Minhas, Vice President of Special Projects, RT4  
Anil Srivastava, President, RT4

**Open Medicine Institute** has created a neutral, safe place for patients to collect, curate and share their own medical data and more while helping to contribute (if they desire) to a community of science.

Andy Kogelnik, Founder and Director, RT2

**Panasonic Corporation** is a worldwide leader in the development of diverse electronics technologies and solutions for customers in the consumer electronics, housing, automotive, enterprise solutions and device industries.

Neil Jacobs, Chief Atmospheric Scientist, RT4

**The Patient-Centered Outcomes Research Institute** is an independent nonprofit, nongovernmental organization whose mandate is to improve the quality and relevance of evidence available to help patients, caregivers, clinicians, employers, insurers, and policymakers make informed health decisions.

Rachael Fleurence, Program Director, RT3

**Personal.com** provides individuals with secure digital vaults for storing, sharing, importing, and reusing their important data and files, including for auto form-filling.

Josh Galper, Chief Policy Officer, RT1

**Personal Genome Project** is the founding pilot project in PGP network. The project hosts publicly shared genomic and health data from thousands of participants.

Madeleine Ball, Director of Research, RT1

**The Pew Charitable Trusts** is an independent nonprofit organization informed by the founders' interest in research, practical knowledge and robust democracy.

Mark Richardson, Senior Associate, RT4

**Planet OS** creates a suite of powerful cloud-based solutions that allows companies to securely integrate open and commercial data with their proprietary data.

Rainer Sternfled, CEO, RT4

**PricewaterhouseCoopers** is a network of firms in 158 countries that delivers quality in assurance, tax and advisory services.

Abdul Shaikh, Director, RT4

**Public Library of Science** is a nonprofit open access scientific publishing project aimed at creating a library of open access journals and other scientific literature under an open content license.

Larry Peiperl, Chief Editor, RT3

**RPS Group** is a multinational energy resources and environmental consultancy company that advises its clients on the built and natural environment across diverse economic sectors, both public and private.

Eoin Howlet, Managing Director, RT4  
Brian McKenna, Senior Programmer, RT4

**Sage Bionetworks** is a non-profit research organization that seeks to develop predictors of disease and accelerate health research through the creation of open systems, incentives, and standards.

John Wilbanks, Chief Commons Officer, RT4

**The Schmidt Ocean Institute** works to advance the frontiers of global marine research by providing state of the art operational, technological, and informational support to the pioneering ocean science and technology development projects at sea.

Allison Miller, Research Program Manager, RT4

**SecondMuse** is an innovation and collaboration agency that co-creates prosperity by applying the art and science of collaboration to solve complex problems.

Neisan Massarrat, Director, RT4

**SiteCompli** provides breakthrough technology, expertise and resources to analyze, alert and report on its customers' critical NYC real estate compliance information.

Nicholas O'Brien, Director of Data Insights, RT4

**Socrata** provides cloud solutions for federal, state, and local governments to transform data into actionable insights for public and government use.

Ken Melero, Federal Director, RT4  
Joe Pringle, Director, RT3

**SPARC (the Scholarly Publishing and Academic Resources Coalition)** works to enable the open sharing of research outputs and educational materials in order to democratize access to knowledge, accelerate discovery, and increase the return on our investment in research and education.

Nick Shockey, Director of Programs, RT3

**Sunlight Foundation** is a national, non-partisan, nonprofit organization that uses the tools of civic tech, open data, policy analysis and journalism to make our government and politics more accountable and transparent to all.

John Wonderlich, Executive Director, RT1  
Emily Shaw, Senior Analyst, RT3

**Syapse** drives healthcare transformation through precision medicine, enabling provider systems to improve clinical outcomes, streamline operations, and shift to new payment models.

Jonathan Hirsch, Founder, RT1

## Participating Organizations

### COMPANIES, NONPROFIT, ACADEMIC AND RESEARCH ORGANIZATIONS

**Tableau** is a computer software company that produces a family of interactive data visualization products focused on business intelligence.

Helen Xing, Strategic Alliances and Channel, RT4

**Trea** is a service where people can explore innovation relationships and map ideas to the unified knowledge graph.

Max Yuan, Founder, RT2

**The University of Arkansas at Little Rock's Graduate Certificate in Technology Innovation program** is intended for working professionals and post-baccalaureate students who are interested in the development, evaluation and implementation of original ideas for existing businesses and new enterprises.

John R. Talburt, Professor and Acxiom Chair of Information Quality, RT2

**The University of Maryland** is a public research institution committed to educating students and advancing knowledge in areas of importance to the State, the nation, and the world

Lissa Snyders, Presidential Management Fellow, RT3

Stephanie Yarwood , Assistant Professor, RT3

Lola Taiwo, PhD Student, RT3, RT4

**West Big Data Innovation Hub**, one of four National Science Foundation funded Regional Innovation Hubs, builds multi-sector and multi-state partnerships to address societal challenges with Big Data innovation.

Meredith Lee, Director, RT2

**Western Pennsylvania Regional Data Center** provides a shared technological and legal infrastructure to support research, analysis, decision making, and community engagement.

Bob Gradeck, Project Manager, RT2

**World Resources Institute** is a global research organization that focuses on six critical issues at the intersection of environment and development: climate, energy, food, forests, water, and cities and transport.

Carolyn Savoldelli, Research Analyst, RT4

## Participating Organizations

### GOVERNMENT AGENCIES AND OFFICES

Participants are noted by the roundtables they attended in 2016 (RT1 is the Open Data and Privacy Roundtable, RT2 is the Open Data and Improving Data Quality Roundtable, RT3 is the Open Data Roundtable for Sharing and Applying Research Data, RT4 is the Open Data Roundtable for Public-Private Collaboration).

**The City of Seattle, Department of Information Technology** is the central information technology organization for Seattle. It partners with internal and external customers to provide reliable and secure information and communication solutions that make technology work for the city.

Michael Mattmiller, Chief Technology Officer, RT1

**The City of Syracuse's Department of Innovation** works to identify, examine, and resolve infrastructure-related challenges. It works within the city and with the community to make large, impactful changes, and are guided by data to make evidence-based decisions.

Sam Edelstein, Chief Data Officer, RT2

**The Committee on the Marine Transportation System** is a Federal interdepartmental committee whose mission is to create a partnership of Federal departments and agencies with responsibility for the Marine Transportation System (MTS).

Helen Brohl, Executive Director, RT3

**The Commonwealth of Massachusetts' Office of Information Technology's** mission is to drive the digital business of state government to meet the continually evolving needs of constituents, schools, businesses, and local government.

Holly St. Clair, Director of Enterprise Data Management, RT2

**The Office of the Manhattan Borough President** is responsible for advising the Mayor and City Council on borough concerns, commenting on all land-use matters in the borough, advocating for the borough in the , municipal budget process, and appointing members of Manhattan's 12 Community Boards.

William Colegrove, Senior Technology Advisor, RT2

**The Centers for Disease Control and Prevention** is dedicated to protecting health & promoting quality of life through prevention and control of disease, injury, and disability.

Michael Johansson, Biologist, RT2

**The Consumer Financial Protection Bureau** works to make markets for consumer financial products and services work for Americans. It arms people with the information they need to make smart financial decisions, and protects them from unfair, deceptive, and abusive practices.

Ren Essene, Program Manager,  
Mortgage Data Assets, RT1

Linda Powell, Chief Data Officer, RT1

Claire Stapleton, Chief Privacy Officer, RT1

**Defense Advanced Research Projects Agency** works with academic, corporate and governmental partners to make pivotal investments in breakthrough technologies for national Security.

Matt Hepburn, Program Manager, RT3

**The Defense Health Agency** is a joint, integrated Combat Support Agency that enables the Army, Navy, and Air Force medical services to provide a medically ready force and ready medical force to Combatant Commands in both peacetime and wartime.

Jean-Paul Chretien, Lead, Innovation & Evaluation,  
RT2, RT3

**The Federal Communications Commission** regulates interstate and international communications by radio, television, wire, satellite, and cable in all 50 states, the District of Columbia and U.S. territories.

Jon Minkoff, Chief Data Officer, RT1

**The Federal Deposit Insurance Corporation (FDIC)** is a recognized leader in promoting sound public policies, addressing risks in the nation's financial system, and carrying out its insurance, supervisory, consumer protection, resolution planning, and receivership management responsibilities.

Mark Montoya, Chief of Data Strategy, RT2

**The Federal Trade Commission's** mission is to prevent business practices that are anticompetitive or deceptive or unfair to consumers; to enhance informed consumer choice and public understanding of the competitive process; and to accomplish this without unduly burdening legitimate business activity.

Alex Tang, Attorney, RT1

**The General Services Administration's** mission is to deliver the best value in real estate, acquisition, and technology services to government and the American people.

Phil Ashlock, Chief Architect, RT3, RT4

Bob Ballance, Presidentialial Innovation Fellow, RT2, RT4

Tom Black, Digital Services Consultant, RT3

Johan Bos-Beijer, Director, RT2, RT4

John Jediny, Chief Data Engineer for Data.gov,  
RT2, RT3, RT4

Hyon Kim, Program Director, RT1, RT3, RT4

Kris Rowley, Chief Data Officer, RT1, RT2, RT4

Robert Wuhrman, CTO, RT2

**18F's** team of designers, developers, and product specialists inside the GSA enable agencies to rapidly deploy tools and services that are easy to operate, cost efficient, and reusable.

Jeremy Canfield, Service & Experience Designer, RT2

## Participating Organizations

### GOVERNMENT AGENCIES AND OFFICES

**The National Aeronautics and Space Administration's** mission is to pioneer the future in space exploration, scientific discovery and aeronautics research.

David Meyer, Program Manager, RT3, RT4  
Beth Beck, Open Innovation Program Manager, RT3

Michael Gangl, RT3

Kevin Murphy, Earth Science Data Systems Program Executive, RT2  
Margaret Roberts, Attorney-Advisor, RT4

**The National Archives and Records Administration** is the nation's record keeper. Of all documents and materials created in the course of business conducted by the United States Federal government, only 1%-3% are so important for legal or historical reasons that they are kept forever.

Hannah Bergman, Assistant General Counsel, RT1  
Gary Stern, General Counsel, RT1

**The National Defense University** provides rigorous Joint Professional Military Education to members of the U.S. Armed Forces and select others in order to develop leaders that have the ability to operate and creatively think in an unpredictable and complex world.

Libbie Prescott, Deputy Director, RT1

**The National Information Exchange Model (NIEM)** is a community-driven, standards-based approach to exchanging information. It includes a data model, governance, training, tools, technical support services, and an active community—that assists users in adopting a standards-based approach to exchanging data.

Mike Hulme, Co-Chair, RT2

**The National Institute of Standards and Technology's** mission is to promote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life.

Simson Garfinkel, Computer Scientist, RT1  
Robert Hanisch, Director, Office of Data and Informatics, RT2  
Sokwoo Rhee, Associate Director of Cyber-Physical Systems Program, RT3  
Jim Warren, Director, RT3  
Henry Wixon, Chief Counsel, RT3

**The National Science Foundation's** mission includes support for all fields of fundamental science and engineering, except for medical sciences. It is tasked with keeping the United States at the leading edge of discovery in areas from astronomy to geology to zoology.

Chaitan Baru, Senior Advisor for Data Science, RT3  
Peggy Ann Hoyle, Deputy General Counsel, RT4  
Beth Linas, AAAS Fellow, RT3  
Amy Northcutt, Senior Agency Official for Privacy and Chief Information Officer, RT1, RT4  
Eva Zanzerkia, Program Director, RT3

**The Office of Management and Budget's** mission is to serve the President of the United States in implementing his vision across the Executive Branch.

Claire Ehmann, White House Leadership Development Program Fellow, RT1, RT3, RT4  
Margie Graves, Policy Analyst, RT2, RT3, RT4  
Justin Grimes, Policy Analyst, RT2, RT4  
Jimena Luna, Senior Policy Analyst, RT2, RT3, RT4  
Margo Schwab, Science Policy Analyst, RT2, RT4  
Bob Sivinski, Statistician, RT2

**The Office of the Vice President's** mission is to serve the Vice President of the United States.

Jerry Lee, Health Sciences Director, RT3, RT4

**U.S. Agency for International Development** is the lead U.S. Government agency that works to end extreme global poverty and enable resilient, democratic societies to realize their potential.

Gayle Girot, Chief Innovation Counsel, Assistant General Counsel for the Global Development Lab and the Bureau for Economic Growth, Education, and the Environment, RT1  
Brandon Pustejovsky, Chief Data Officer, RT1  
Ruth Yodaiken, Senior Attorney, RT1

**The U.S. Arctic Research Commission** is an independent agency that advises the President and Congress on domestic and international Arctic research through recommendations and reports.

John Farrell, Executive Director, RT3  
Ruth Cooper, Intern, RT3

**U.S. Department of Agriculture** is the federal executive department responsible for developing and executing federal government policy on farming, agriculture, forestry, and food.

Samuel Crowell, Advisor, RT2  
Kelvin Fairfax, Chief Privacy Officer, RT1  
Brooke Geller, Attorney-Advisor, RT3  
Bobby Jones, Sr. Advisor / Acting Chief Data Officer, RT1, RT4  
David Kingbury, Soil Survey Regional Director, RT3  
Melissa McClellan, General Attorney, Office of the General Counsel, RT1, RT4  
Shawn McGruder, Associate General Counsel, General Law and Research, RT1, RT3

**The U.S. Department of Commerce's** mission is to create the conditions for economic growth and opportunity. The Department works with businesses, universities, communities, and the Nation's workers to promote job creation, economic growth, sustainable development, and improved standards of living for Americans.

Justin Antonipillai, Senior Advisor, RT4  
Robin Bachman, Chief, Policy Coordination Office and Chief Privacy Officer, RT1  
Avi Bender, Chief Technology Officer, Information Technology Directorate, RT2, RT3, RT4  
Jeff Chen, Chief Data Scientist, RT2, RT3, RT4  
Colin Holmes, Senior Advisor to the General Counsel, RT2  
Negar Kalbasi, Web Developer, RT4  
David Langdon, Economist and Senior Policy Advisor, RT4  
Sara Zdeb, Senior Counsel to the General Counsel, RT4

**The Bureau of Economic Analysis (BEA)** promotes a better understanding of the U.S. economy by providing the most timely, relevant, and accurate economic

## Participating Organizations

### GOVERNMENT AGENCIES AND OFFICES

accounts data in an objective and cost-effective manner.

Andrea Julca, Economist, RT3

**The Economics and Statistics Administration's** mission is to provide high-quality economic analysis and foster the missions of the Census Bureau and the Bureau of Economic Analysis.

Sabrina Montes, Economist, RT2

**The National Oceanic and Atmospheric Administration's** mission is to understand and predict changes in the Earth's environment, from the depths of the ocean to the surface of the sun, and to conserve and manage our coastal and marine resources.

Jen Arrigo, Federal Program Officer, RT3, RT4

Jeff de La Beaujardiere, Data Management Architect, RT4

Brian Eiler, Senior Advisor, RT4

Ed Kearns, Technical Lead, NOAA's Big Data Partnership & Chief, RT2

Barb Kirkpatrick, Board of Directors and Executive Committee, RT4

Jeremy Mathis, Director, RT3

Derrick Snowden, Chief, RT4

Glenn Tallia, Section Chief, RT3

**The National Technical Information Service's** mission is to promote the Commerce Department's and Federal data priorities, including open access and open data, by providing information and data services to the public, industry, and other federal agencies in ways that enable U.S. innovation and economic growth.

Derssie Mebratu, Deputy Chief Scientist, RT4

Paul Weston, RT4

**The U.S. Patent and Trademark Office (USPTO)** advises the president of the United States, the secretary of commerce, and U.S. government agencies on intellectual property (IP) policy, protection, and enforcement; and promotes the stronger and more effective IP protection around the world.

Alan Marco, Chief Economist, RT2, RT3

**The U.S. Department of Education's** mission is to promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access.

Liz Albro, Associate Commissioner, Teaching and Learning Division, RT1

Jay Chen, Senior Counsel, RT4

Deborah Friendly, Attorney, RT1

Michael Hawes, Statistical Privacy Advisor, RT1

Ron Petracca, Senior Counsel, RT3

Ross Santi, Associate Commissioner, RT2

Kristina Spencer, General Attorney, RT4

**The U.S. Department of Energy's** mission is to ensure America's security and prosperity by addressing its energy, environmental and nuclear challenges through transformative science and technology solutions.

John Lucas, Acting Deputy General Counsel, RT3

Alison LaBonte, Interdisciplinary Scientist, RT4

**The National Petroleum Council** is a federally chartered and privately funded advisory committee that advises, informs and makes recommendations to the Secretary of Energy with respect to any matter relating to oil and natural gas, or to the oil and gas industries submitted to it or approved by the Secretary.

John Guy, Deputy Executive Director, RT3

**The National Renewable Energy Laboratory (NREL)** focuses on creative answers to today's energy challenges. From breakthroughs in fundamental science to new clean technologies to integrated energy systems that power our lives, NREL researchers are transforming the way the nation and the world use energy.

Debbie Brodt-Giles, Digital Assets Manager, RT2

Jeff Gonder, Section Supervisor, RT3

Jon Weers, Data Scientist / Web Strategist, RT2

**The U.S. Department of Health and Human Services** is the U.S. government's principal agency for protecting the health of all Americans and providing essential

human services, especially for those who are least able to help themselves.

Margeaux Akazawa, Presidential Management Fellow, RT3

George Chambers, Executive Director, RT2

**The Centers for Disease Control and Prevention** is dedicated to protecting health and promoting quality of life through prevention and control of disease, injury, and disability.

Michael Johansson, Biologist, RT3

**The IDEA Lab** was established to improve how HHS delivers on its mission. This effort was started as a response to input from the workforce and public to promote advances in organizational management centered around three core beliefs: every individual has the ability to improve the health and well-being of Americans, people are more powerful when working together; and there is a solution to every problem.

David Portnoy, HHS IDEA Lab Fellow, RT2, RT3

**The National Institutes of Health**, a part of the U.S. Department of Health and Human Services, is the nation's medical research agency — making important discoveries that improve health and save lives.

Vivien Bonazzi, Senior Advisor, RT4

Philip Bourne, Associate Director, RT3

Mike Huerta, Associate Director, RT4

Elizabeth Kittrie, Senior Advisor, RT4

Jennie Larkin, Senior Advisor, RT4

Dina Paltoo, Director, Scientific Data Sharing Policy Division, RT1, RT3, RT4

**The U.S. Food and Drug Administration** is responsible for protecting the public health by assuring the safety, efficacy, and security of human and veterinary drugs, biological products, medical devices, our nation's food supply, cosmetics, and products that emit radiation.

Ryan Ortega, Commissioner's Fellow, RT3

## Participating Organizations

### GOVERNMENT AGENCIES AND OFFICES

**U.S. Department of Interior** protects America's natural resources and heritage, honors our cultures and tribal communities, and supplies the energy to power our future.

Teri Barnett, Departmental Privacy Officer, **RT1**  
Thomas "Tod" Dabolt, Acting Geographic Information Office, **RT2**

Lin Zhang, Senior Enterprise Architect, Office of the CIO, **RT1, RT2**

**The U.S. Department of Justice's** mission is to enforce the law and defend the interests of the United States according to the law; to ensure public safety against threats foreign and domestic; to provide federal leadership in preventing and controlling crime; to seek just punishment for those guilty of unlawful behavior; and to ensure fair and impartial administration of justice for all Americans.

Christopher Egan, Director of Information Management, Tax Division, **RT1, RT4**

**The U.S. Department of Labor's** mission is to foster, promote, and develop the welfare of the wage earners, job seekers, and retirees of the United States; improve working conditions; advance opportunities for profitable employment; and assure work-related benefits and rights.

Joseph Plick, Counsel for FOIA & Information Law, **RT1**

**The U.S. Department of State's** mission is to shape and sustain a peaceful, prosperous, just, and democratic world and foster conditions for stability and progress for the benefit of the American people and people everywhere.

Sharett Diggs, Enterprise Architect, **RT1**  
Ramesh Ramakrishnan, Data Architect, **RT1**  
Paul Zeitz, Director, **RT4**

**U.S. Department of Transportation** serves the United States by ensuring a fast, safe, efficient, accessible and convenient transportation system that meets our vital national interests and enhances the quality of life of the American people, today

and into the future.

Claire Barrett, Chief Privacy Officer, **RT1**  
Leighton Christiansen, Data Curator, **RT3**  
Ariel Gold, Data Program Manager, **RT3**  
Dan Morgan, Chief Data Officer, **RT4**

**The U.S. Digital Service** is using the best of product design and engineering practices to transform the way government works for the American people. In every corner of government, we join forces with the many passionate and talented tech professionals within agencies who are dedicated to public service.

Alberto Colon-Viera, Digital Services Expert, **RT2**  
Poulomi Damany, Digital Services Expert, **RT1**

**The U.S. Environmental Protection Agency's** mission is to protect human health and the environment. EPA works with other federal agencies, states, tribes and local communities to improve the health of American families and protect the environment across the country.

Andy Dupont, Environmental Engineer, **RT2**  
Judy Earle, Agency Privacy Officer, **RT1**  
Kevin Kirby, Enterprise Data Architect, **RT2**

**The U.S. Geological Survey** is a science organization that provides impartial information on the health of our ecosystems and environment, the natural hazards that threaten us, the natural resources we rely on, the impacts of climate and land-use change, and the core science systems that help us provide timely, relevant, and useable information.

Peter Ruhl, Biological Data Manager, **RT2**  
Jonathan Smith, Program Coordinator, **RT3**  
Lynda Lastowka, Supervisory Geophysicist, **RT4**

**The U.S. Office of Personnel Management** provides human resources, leadership, and support to federal agencies and helps the federal workforce achieve their aspirations as they serve the American people.

Bernhard Kluger, Deputy Performance Improvement Officer, **RT2**

The Social Security Administration's mission is to deliver Social Security services that meet the changing needs of the public.

Ozlen Luznar, Economist/Statistician, **RT1**  
Linda McCaw, Specialist, **RT4**  
Mary Ann Zimmerman, Supervisory Government Information Specialist, **RT1**

**The U.S. Department of Health and Human Services** is the U.S. government's principal agency for protecting the health of all Americans and providing essential human services, especially for those who are least able to help themselves.

Lucia Savage, Chief Privacy Officer, **RT1**

**The White House, Office of Science and Technology Policy's** mission is: first, to provide the President and his senior staff with accurate, relevant, and timely scientific and technical advice on all matters of consequence; second, to ensure that the policies of the Executive Branch are informed by sound science; and third, to ensure that the scientific and technical work of the Executive Branch is properly coordinated so as to provide the greatest benefit to society.

Noel Bakhtian, Senior Policy Advisor, **RT3, RT4**  
Austin Brown, Assistant Director for Clean Energy and Transportation, **RT3, RT4**  
Dan Correa, Senior Advisor, Innovation Policy, **RT3, RT4**

Melissa Goldstein, Assistant Director for Bioethics and Privacy, **RT1, RT2, RT3, RT4**  
Natalie Evans Harris, Senior Policy Advisor, **RT2, RT4**  
Andrew Hanus, Fellow, **RT3**

Kristen Honey, Policy Advisor, **RT1, RT2, RT3, RT4**  
Martin Jeffries, Assistant Director, **RT3, RT4**

Kelly Kryc, Senior Policy Analyst, **RT2**  
Fabien Laurier, Senior Policy Advisor and Interim Director, **RT2, RT3, RT4**

Amy Luers, Assistant Director, **RT2, RT3, RT4**  
Jason Schultz, Senior Advisor, **RT1, RT2, RT4**

Jerry Sheehan, Assistant Director, Scientific Data and Information, **RT1, RT2, RT3**

Nancy Weiss, Senior Advisor to the CTO, Innovation and IP, **RT1, RT2, RT3, RT4**  
Marc Wynne, Fellow, White House Leadership Development Program, **RT1, RT2**

# Appendix F: Acknowledgements

Thank you to everyone that supported the 2016 Open Data Roundtables.

Program Leads: Katherine Garcia, Joel Gurin, Kristen Honey

Planning team: Audrey Ariss, Noel Bakhtian, Austin Brown, Dan Correa, Claire Ehmann, Melissa Goldstein, Natalie Evans Harris, Stephanie Huang, Martin Jeffries, Kelly Kryc, Fabien Laurier, Amy Luers, Jimena Luna, Laura Manley, David Meyer, Ana Pinheiro Privette, Jason Schultz, Jerry Sheehan, Nancy Weiss

Briefing Papers Lead Author: Matt Rumsey

Facilitators, Notetakers and Volunteers: Prue Adler, Jennifer Arrigo, Bob Ballance, Tom Black, Jimmy Catania, Joohee Choi, Ruth Cooper, Theo Goetemann, Ben Green, Andrew Hanus, Sophie Hu, Teresa Kim, Jerry Lee, Beth Linas, Caile Morris, Ryan Ortega, Vinayak Pande, Brandi Schottel, Emily Shaw, Pooja Singh, Lissa Snyders, Emily Sylak-Glassman, Lola Taiwo, Marc Wynne

Photographers: Oliver Contreras, Hanlun Song

Designer: Zak Bickel

Illustrators: Jon Marchione, Chelsea Beck



THE CENTER FOR  
**OPEN DATA ENTERPRISE**

