THE CENTER FOR
OPEN DATA ENTERPRISE

Briefing Paper on Open Data for Public-Private Collaboration

This paper is one in a series of four Briefing Papers on key issues in the use of open government data. The U.S. federal government, like many governments around the world, releases "open data" that can help the public find better value in education, fair housing, and safer medicines, and has a wide range of other social and economic benefits. Open data also helps government agencies themselves operate more efficiently, share information, and engage the citizens they serve. Under the U.S. government's Open Data Policy,[1] all federal agencies "must adopt a presumption in favor of openness to the extent permitted by law and subject to privacy, confidentiality, security, or other valid restrictions." While this paper focuses on policies and examples from the U.S., it is meant to be useful to open data providers and users in other countries as well.

Introduction

Many healthcare, agriculture, financial services, energy, and transportation companies, among others, use open government data as a key business resource.[2] 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.

This paper presents an overview of the issue and possible solutions for government data stewards and other stakeholders interested in the application of open government data.[3] While it focuses on policies and examples from the U.S., it is meant to be useful to open data providers and users in other countries as well.

Prioritizing and Improving Datasets

The simple act of releasing data publicly, allowing stakeholders to use, understand, and help improve the data, can provide opportunities to improve data quality.[4] Businesses can help government agencies

[1] Sylvia M. Burwell, Steven VanRoekel, Todd Park, Dominic J. Mancini, "M-13-13, Open Data Policy - Managing Information As an Asset", Executive Office of the President, Office of Management and Budget, May 9, 2013, https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf (accessed April 20, 2016).
[2] Open Data 500, "Open Data 500 U.S.: Open Data Compass", http://www.opendata500.com/us, GovLab, (accessed May 13, 2016).
[3] This paper was originally prepared as background for an Open Data and Public-Private Collaboration Roundtable, co-hosted by the White House Office of Science and Technology Policy and the Center for Open Data Enterprise on June 15, 2016.
[4] Open Data Institute, "Environment Agency: Going Open" Open Data Institute, http://theodi.org/ea-going-open-benefits-for-ea (acccessed May 17, 2016), Benefits for EA.

address data quality issues by identifying the most important datasets for improvement and providing feedback on how to improve them.

The Obama Administration has launched a number of communication avenues to gather this feedback as part of its open data efforts.[5] The Demand-Driven Open Data (DDOD) program launched in the Department of Health and Human Services, and the Patent Office project PatentsView, were described in the second Roundtable in this series. Both engage data users to help improve data quality and prioritize high-value datasets.[6] The DDOD provides a model that can be used in other agencies as well to engage stakeholders and enable them to "tell public data owners what's most valuable".[7]

Standards and Interoperability

Open data is provided by government agencies and offices in a wide range of formats, making it difficult to compare and combine different datasets. Collaboration between the public and private sectors can help develop common data standards that ensure the data is interoperable and thus more usable.[8] Standards can be designed to be broadly applicable or targeted at a specific industry or data type. One example is the Building and Land Development Specification (BLDS), "a collaborative effort by civic technology companies, governments and other interested parties to create a shared data specification for building and construction permit data."[9]

Other collaborations have been launched to help individuals more easily access and use their own personal data.[10] The Blue Button initiative is a "partnership between the healthcare industry and the Federal Government that aims to empower all Americans with access to their own electronic health information." The initiative began by making individual medical data from the Department of Veterans Affairs available to vets, but has been expanded to serve a much wider constituency. [11] The initiative makes it easier for citizens to access their health records and facilitates sharing those records among various medical providers.

Similarly, the nonprofit Green Button Alliance was launched in February 2015 to help improve consumers' access to their energy data.[12] The Alliance was started in coordination with the National Institute of Standards and Technology and the Department of Energy to help give consumers access to information about their energy usage.[13] The Alliance brings together government and industry to

[5] "Open Data Engagement, Project Open Data, https://project-open-data.cio.gov (accessed May 17, 2016).
[6] "Briefing Paper on Open Data and Improving Data Quality", Center for Open Data Enterprise, 2016, http://www.opendataenterprise.org/reports/BriefingPaperonOpenDataandImprovingDataQuality.pdf.
[7] "Demand Driven Open Data for the U.S. Department of Health and Human Services", U.S. Department of Health and Human Services, last modified July 5, 2015, http://ddod.healthdata.gov/wiki/Main_Page (accessed May 13, 2016).
[8] Steven Adler, https://www.linkedin.com/pulse/open-data-standards-steven-adler, LinkedIn Pulse, February 26, 2015, (accessed May 17, 2016).
[9] "BLDS Data Specification", BLDS, http://permitdata.org (accessed May 17, 2016).
[10] Kristen Honey, Phaedra Chrousos, and Tom Black, "My Data: Empowering All Americans with Personal Data Access", The White House, March 15, 2016, https://www.whitehouse.gov/blog/2016/03/15/my-data-empowering-all-americans-personal-data-access (accessed May 17, 2016).
[11] Nick Sinai and Adam Dole, "Leading Pharmacies and Retailers Join Blue Button Initiative", HealthIT Buzz, February 7, 2014, http://www.healthit.gov/buzz-blog/consumer/leading-pharmacies-retailers-join-blue-button-initiative, (accessed May 17, 2016).
[12] "An overview of the Green Button initiative", Green Button, http://www.greenbuttondata.org/learn (accessed May 17, 2016).
[13] Green Button Initiative Makes Headway with Electric Industry and Consumers https://www.whitehouse.gov/blog/2015/07/22/green-button-initiative-makes-headway-electric-industry-and-consumers.

develop standards for energy information, increase customers' access to data about their energy use, and encourage "analysis of energy usage data, customized heating and cooling based on the data, analysis of energy costs, and more."[14]

Infrastructure and Hosting

Infrastructure to host and distribute data can be difficult and expensive to develop, especially when it needs to support large or very widely used datasets. The private sector has the resources and technology to securely host, distribute, and analyze open government data and is increasingly doing so. While some of this work is being done through traditional contractual relationships, companies are also volunteering their resources to help make open government data more accessible and useful.

Thematic programs focused on difficult problems may include many private-sector partners to reach their goals. The Climate Data Initiative, launched by the Obama Administration in 2014, includes over a dozen companies who have pledged cloud storage and computing power for climate data and those working to make sense of it.[15] Similar efforts can help develop data for smart cities, preserving the Arctic and oceans, and other large-scale issues.

Data Access and Use

Several companies have taken government data sets, hosted them, and added value to them in ways that include building custom interfaces for users and developers. The USDA, for example, has worked with the private sector to develop portals for key datasets.

At the same time that they broaden the audience for usable data, organizations that help provide access to open government data can help ensure that sensitive data is protected. An earlier Open Data Roundtable explored privacy issues in the release of detailed open data and solutions to address them.[16] The private sector can help ensure that data is managed securely to protect privacy – for example, by using anonymization techniques, by providing "data enclaves" to allow data analysis under secure conditions, or by restricting access to data in other ways.

Strategies for Collaboration

There are often good reasons for the private sector to voluntarily collaborate with government agencies even if they are not paid for their services through grants or contracts. By offering their services, companies can visibly the government in a way that supports the public interest. Many strategies also ensure that both government agencies and their private-sector partners get tangible benefits from the collaboration.

[14] "Data for Our Energy Future: A Roundtable with the U.S. Department of Energy", Center for Open Data Enterprise, 2015, https://s3.amazonaws.com/odenterprise/DOE+Roundtable+Report.pdf (accessed May 17, 2016), 4-5.
[15] Holdren, John. Podesto, John. "Climate Data Initiative Launches with Strong Public and Private Sector Commitments." *White House Blog.* https://www.whitehouse.gov/blog/2014/03/19/climate-data-initiative-launches-strong-public-and-private-sector-commitments. For an example, see Microsoft Azure for Research. *Microsoft.* http://research.microsoft.com/en-us/projects/azure/cdi.aspx (accessed May 13, 2016).
[16] "Briefing Paper on Open Data and Privacy", Center for Open Data Enterprise, 2016, http://www.opendataenterprise.org/reports/BriefingPaperonOpenDataandPrivacy.pdf.

Published by The Center for Open Data Enterprise, www.OpenDataEnterprise.org (2016)

Multi-Company Collaborations

If a government agency collaborates with a single company or organization, both parties may risk the appearance of favoritism or "buying" influence. For example, a company could theoretically take a government data set, clean it, and then gift it back to the government. However, this simple arrangement could cause two problems. It could give the appearance that the government was endorsing that company or favoring it over its competitors by using its data.[17] It could also give the appearance that the company was hoping to influence the government agency, perhaps in an upcoming regulatory proceeding, by providing a valuable service for free.

Some government agencies are avoiding such concerns by setting up collaborations that include many companies within an industry. Cooperative Research and Development Agreements (CRADA's) can be used in this way; they can be structured to bring together "participating organizations across the public and private sectors." One successful CRADA is the National Oceanic and Atmospheric Administration's (NOAA) partnership with Amazon Web Services, Google Cloud Platform, IBM, Microsoft, and the Open Cloud Consortium. It was developed to "research and test solutions for bringing [NOAA's] vast information to the cloud."[18]

Both government and private-sector actors benefit from this relationship. NOAA uses its partners' resources to provide data publicly for free, and the companies benefit from having data in a form that makes it easier for them to sell their customers analytic tools to put the data to use.

While the NOAA partnership operates with data at a global level, another government initiative, the Opportunity Project, uses data that is highly localized. This project, led by the U.S. Census, is designed to make it easier for communities to use both federal and local data for new insights and action, putting "data and digital tools in the hands of families, communities, and local leaders…" The Opportunity Project enlists companies and non-profits to build tools that make it easier for local governments to leverage census data while also improving participants' existing products and services.[19]

Private Sector Data Sharing

A number of companies and industries are embracing open principles as part of their larger business strategies, rather than as one-off opportunities. While only a relatively small number of companies are sharing their data so far, private sector data sharing can be beneficial to governments, individual organizations, entire industries, and the general public. A recent report by the Open Data Institute explores three companies are "embedding open principles into their operations to gain competitive

[17] Joel Gurin, Audrey Ariss, Katherine Garcia, and Laura Manley, "Report of Findings from an Open Data Roundtable with the U.S. Patent & Trademark Office", Center for Open Data Enterprise, April 2015, https://s3.amazonaws.com/odenterprise/PTO+Roundtable+Report.pdf (accessed May 17, 2016), 14.
[18] "U.S. Secretary of Commerce Penny Pritzker Announces New Collaboration to Unleash the Power of NOAA's Data", U.S. Department of Commerce, April 21, 2015, https://www.commerce.gov/news/press-releases/2015/04/us-secretary-commerce-penny-pritzker-announces-new-collaboration-unleash (accessed May 26, 2016).
[19] Opportunity Project, http://opportunity.census.gov (accessed June 8, 2016).

advantage."[20] Pharmaceutical companies are beginning to share research data both to help accelerate research overall and to speed approval for new drugs.[21] Companies in the financial and insurance industries could also benefit by sharing their data with government, and vice versa, to make it easier for both the companies and government to prevent and detect fraud and abuse.

An even more altruistic approach is "data philanthropy," in which private-sector organizations allow access to their data holdings for the public good. This emerging concept was first proposed at the World Economic Forum in Davos in 2011 and popularized through a UN data project, Global Pulse.[22] While acknowledging potential complications, including privacy risks, proponents of data philanthropy argue that the rapidly growing quantities of "data exhaust" held by the private sector can be shared and leveraged to "protect communities against the impacts of fast-moving crises and keep global development on track."[23] Some companies have begun to explore the possibilities of sharing their data for the public good. For example, in 2013 "France Telecom-Orange...made anonymized records of five million mobile phone users in Cote d'Ivoire available to the research community" as part of a challenge aimed at tackling development issues in the country[24]

Despite the benefits, many businesses may be reluctant to share data that they view to be proprietary, out of concern that government agencies would release it publicly and erase a potential market advantage. For example, a recent report detailed Panasonic's proprietary weather data model, which offers significant improvements on government weather data. The company shares its data with the government during emergency situations, but expressed concern about "the legal implications of sharing proprietary data with the government—since a public agency is obligated to make all data it acquires public, and that could eliminate Panasonic's proprietary forecast advantage."[25]

Data Intermediaries

A growing number of companies exist "to make it easier for other businesses to use Open Data...These companies provide platforms and services that make open government data easier to find, understand, and use." [26] These "data intermediaries" could help take some pressure off of government by improving the quality of open data, making it more discoverable, and helping to connect it with other useful data.

[20] "Open Enterprise: How three big businesses create value with open innovation", The Open Data Institute, 2016, http://theodi.org/open-enterprise-big-business (accessed May 25, 2016).
[21] "Principles for Responsible Clinical Trial Data Sharing", PhRMA, http://www.phrma.org/phrmapedia/responsible-clinical-trial-data-sharing (accessed May 25, 2016).
[22] Stefaan Verhulst, "Mapping the next frontier of open data: Corporate data sharing", http://thegovlab.org/mapping-the-next-frontier-of-open-data-corporate-data-sharing, GovLab, September 16, 2014, (accessed May 13, 2016).
[23] Robert Kirkpatrick, "Data Philanthropy: Public & private sector data sharing for global resilience", United Nations Global Pulse", September 16, 2011, http://www.unglobalpulse.org/blog/data-philanthropy-public-private-sector-data-sharing-global-resilience (accessed May 13, 2016).
[24] See France-Telecom's release of "anonymized records of five million mobile phone users in Cote d'Ivoire…to the research community", referenced in: Andreas Pawelke and Anoush Rima Tatevossian, " Data Philanthropy: Where are we now?", United Nations Global Pulse, May 8, 2013, http://www.unglobalpulse.org/data-philanthropy-where-are-we-now (accessed May 13, 2016).
[25] Eric Holthaus, "Panasonic (Yes, Panasonic) May Have Created the World's Best Weather Model", Slate, New America, and Arizona State University, May 9, 2016, http://www.slate.com/articles/technology/future_tense/2016/05/panasonic_says_it_s_created_the_world_s_best_weather_model.html (accessed May 17, 2016).
[26] Joel Gurin, "Driving Innovation with Open Data", U.S. Chamber of Commerce Foundation, October 6, 2014, https://www.uschamberfoundation.org/driving-innovation-open-data (accessed May 17, 2016).

These data intermediaries may work with a broad range of data types or focus on one specific kind of data.[27] For example, one company collects information about the ownership and structure of corporate entities from around the world and aims to make that information "more accessible, more discoverable, and more usable, and thus give citizens, community groups, journalists, other companies, and society as a whole the ability to understand, monitor and regulate them." [28] Data intermediaries often operate on a hybrid "free/fee" business model. They may make their datasets available for free as a public service and to raise the company's visibility, while charging clients for consulting services to help them put the data to use.

Challenges and Prizes

Both the public and private sector have used challenges and prizes to boost engagement and drive new organizations to innovate with open government data. For example, as part of the Climate Data Initiative (CDI), some cloud hosting providers are running programs that will provide winning projects with access to cloud computing resources. Microsoft's Climate Data Award grants free cloud computing resources to climate data projects looking to "transform the way we inform and prepare communities, businesses, and citizens." [29] Also as part of the CDI, Microsoft partnered with the USDA on an Innovation Challenge for food resilience.[30] Meanwhile, Amazon is providing access to "super computing resources running on the Amazon Elastic Compute Cloud" as part of its Amazon Climate Research Grants Program.[31]

Challenges and prizes are often offered through cross-sector collaboration between private companies, NGOs, and government agencies. In the UK, for example, Nesta and the Open Data Institute have collaborated on an Open Data Challenge Series to help leverage open data by engaging startups, mature businesses, and individuals across a number of sectors.[32] Challenges can also be run by private organizations alone, and several companies have built entire businesses around challenges and competitions.[33]

Questions for Further Consideration

- How can private sector organizations be encouraged to work with government to increase the value of open data? Which of the examples detailed in this paper could be leveraged more broadly?
- Are there potential concerns (legal or otherwise) that haven't been encountered yet but may emerge in the future? How might they be addressed?

[27] See, for example Enigma.io, Graphiq, and OpenCorporates
[28] "Opencorporates Principles", OpenCorporates, https://opencorporates.com/info/principles (accessed May 17, 2016).
[29] "Climate Data Initiative", Microsoft Research, http://research.microsoft.com/en-us/projects/azure/cdi.aspx (accessed May 17, 2016).
[30] "Innovation Challenge for Food Resilience", Microsoft Research, http://research.microsoft.com/en-us/projects/azure/cdi2.aspx (accessed May 25, 2016).
[31] Jeff Barr, "New Amazon Climate Research Grants", Amazon Web Services Blog, July 29, 2014, http://aws.amazon.com/blogs/aws/amazon-climate-research-grants (accessed May 17, 2016).
[32] "Open Data Challenge Series", Nesta, http://www.nesta.org.uk/open-data-challenge-series (accessed May 17, 2016).
[33] See https://www.innocentive.com and https://www.kaggle.com/competitions

- What sectors and topics may be suitable for public-private partnership around data? What are the key characteristics of a topic that make it suitable?

Acknowledgments

The Center for Open Data Enterprise thanks Open Data Partner Microsoft and Open Data Supporter Booz Allen Hamilton for supporting the Center's work on the Open Data Roundtables and this Briefing Paper. The lead researcher and author for this paper is Matthew Rumsey, Research Fellow at the Center for Open Data Enterprise. The Center is an independent nonprofit 501(c)3 organization, based in Washington, DC, whose mission is to maximize the value of open data as a public resource.

Suggestions for Further Reading

Eric Holthaus, "Panasonic (Yes, Panasonic) May Have Created the World's Best Weather Model", Slate, New America, and Arizona State University, May 9, 2016, http://www.slate.com/articles/technology/future_tense/2016/05/panasonic_says_it_s_created_the_world_s_best_weather_model.htm

Martin Tisne, "The Missing Link: How to engage the private sector in OGP", Open Government Partnership Blog, March 19, 2014, http://www.opengovpartnership.org/blog/martin-tisne/2014/03/19/missing-link-how-engage-private-sector-ogp

Kristen Honey, Phaedra Chrousos, and Tom Black, "My Data: Empowering All Americans with Personal Data Access", The White House, March 15, 2016, https://www.whitehouse.gov/blog/2016/03/15/my-data-empowering-all-americans-personal-data-access

Stefaan Verhulst, "Mapping the next frontier of open data: Corporate data sharing", GovLab, September 16, 2014, http://thegovlab.org/mapping-the-next-frontier-of-open-data-corporate-data-sharing

"Open Enterprise: How three big businesses create value with open innovation", The Open Data Institute, 2016, http://theodi.org/open-enterprise-big-business

Amparo Ballivian and William Hoffman, "Public-Private Partnerships for Data: Issues Paper for Data Revolution Consultation", World Bank, Draft, January 2015, http://data.worldbank.org/sites/default/files/issue-paper-financing-the-data-revolution-ppps_0.pdf