



HARVARD Kennedy School

ASH CENTER

for Democratic Governance
and Innovation

IMPROVING SERVICE AND COMMUNICATION WITH OPEN DATA: A HISTORY AND HOW-TO

EMILY SHAW, PHD

June 2015

IMPROVING SERVICE AND COMMUNICATION WITH OPEN DATA: A HISTORY AND HOW-TO

EMILY SHAW, PHD

Enacting policies to make more data publicly accessible also makes governments more accessible to the communities they serve. Open data fuels the kind of two-way communication that increases opportunities for the public to interact with government, while also allowing local governments to maximize the value of their data. This paper examines the ways that government can improve service and communication with open data, answering questions including:

- What is open data?
- How do governments open data?
- Why are state and local governments opening data?
- How can a government get started with open data?

PAPER SERIES

This paper is part of a series published by Data-Smart City Solutions, a project of the Ash Center at Harvard Kennedy School. The series explores data-related facets of civic engagement in today's cities.



ABOUT THE AUTHOR

Emily Shaw is the Deputy Policy Director at the Sunlight Foundation, where she helps the organization lead the charge to make useful public information available online.

WHAT IS OPEN DATA?

“Open data” is a term that can describe both content itself— data that is “open” to public view—and an overarching approach to publishing data in an open, transparent way. Open data is characterized by three core elements. To be open, data must first be available to everyone online, freely and without registration; second, it must be available without legal encumbrance on its use or reuse; and, third, it should be made available in forms and formats that make the data maximally useful for the broadest range of uses and users.

The term “open data” is conceptually rooted in the principles articulated in the 2005 Open Definition, a collaboratively developed statement of what it means for intellectual material to be “open.”¹ The Open Definition, now housed online at opendefinition.org, states that data is open when it is available to everyone, free for use and reuse, and when it has a minimal form of licensing which, at its most stringent, requires author attribution and the obligation to “share-alike,” or make subsequent derivative works similarly open. While open data need not necessarily come from a particular source, there is a particularly compelling public access claim on government as a source of open data, since government data was created and gathered under public authority. Because governments are most often the source of data that is opened, the term “open data” is often used interchangeably with the term “open government data.”

While open data is often assumed to be quantitative in nature, other perspectives are more expansive and consider data to be synonymous with information in general. This broader interpretation of “data” includes such text-heavy forms of information as legislation, which makes up the datasets used by the Sunlight Foundation’s Open States project (openstates.org). Textual data is “opened” when made legally accessible, made available to the public online, and structured in a way that makes it searchable and machine-readable, e.g., through the use of consistent metadata schema.

HOW DO GOVERNMENTS OPEN DATA?

Governments have taken two main approaches to opening up data: they either put data on a website with the intention of making it “open,” or they develop a legal structure to undergird a more complex, ongoing form of open data release. The first method—government simply putting data on a website in a way that is intended to facilitate public use and reuse—is still the more common of the two approaches. The second method, establishing an open data policy, is a newer approach. It has the

¹ The original “open definition” can be found at <https://web.archive.org/web/20060819043123/http://www.okfn.org/okd/definition.html>.

advantage of laying the groundwork for greater continuity of access to data, creating a better public understanding of government data holdings, and ensuring access to specific types of data enumerated by the policy.

The growth of governments into regular public data providers is a natural development of the earliest experiments with e-government. As government websites have gained capacity over the last two decades, they have become loci for a broad range of information and platforms for achieving a variety of goals. The role of government as an essential provider of high-quality public data has grown in tandem.

Making data available and reusable has deep roots in existing practice. Most notably, the purpose of the United States Census has always been both to collect and to provide data to by making “published census statistics . . . available to anyone who needs them.”² While the Census used public and academic library repositories to share data through the early 2000s, other government entities had already begun putting their datasets online. The U.S. Government Printing Office put the daily activities of Congress online in 1994, and the U.S. Congress created the THOMAS website as a way of opening legislative data to public oversight in the following year. At the state level, Florida and Hawaii were early adopters of online publication for campaign finance disclosure, maintaining publicly accessible campaign finance websites by 1997. In general, making records available by computer tape, disk, or dial-up FTP provided an intermediate step between offering paper copies and the advent of government websites as a method of allowing the public free access to government data.³

While what government entities were doing during the 1990s and early 2000s got closer to the ideal of truly open data, they began describing their efforts to provide information to the public as “open data” only later in the process. Non-governmental groups like the Open House Project (theopenhouseproject.com) and the Open Government Working Group (opengovdata.org) advocated for an alignment of the established principles of open government with open, machine-readable, license-free, and online access to government data. In January 2009, the U.S. government made its most visible move yet toward the adoption of these principles. President Obama’s Memorandum on Transparency and Open Government identified online publication of government data as a method to improve governmental transparency and also implied that technology could be used to improve civic

² US Census, “Factfinder for the Nation: History and Organization,” <https://www.census.gov/history/pdf/cff4.pdf>, accessed May 31, 2015.

³ For an example of the widespread nature of this intermediate step to online data sharing, see the variation in public access to campaign finance disclosure in Elizabeth Hedlund and Lisa Rosenberg (1996). “Plugging In The Public: A model for campaign finance disclosure.” Center for Responsive Politics.

participation and collaboration. By later that year, the term “open data” was used by the White House and executive agencies in their open government plans.⁴

Simultaneously, local governments began proactively sharing large amounts of data and private companies began to provide platforms for state and local governments to offer the public this data in an open format. These sites have come to be known collectively as “open data portals.” In 2006, Washington, D.C. was among the first to create a local government website featuring a wide range of data with the express purpose of “streaming data that the agencies gather through normal operations” to the public.⁵ Soon after, Seattle-based company Socrata opened its doors, selling software tools for providing “open government data, readily accessible over the internet, in a form that maximizes comprehension, interactivity, participation, and sharing.”⁶ The earliest commercial provider of open data platforms, Socrata was selected to power a number of new federal, state and local open data sites by 2010.

Many U.S. state and local governments have continued to build on this trend of publishing raw data with the intention of making it open for public use and reuse. Some have engaged commercial or non-profit providers to provide a platform for them, while others have built their own data sites. Governments currently publish a wide range of datasets, from both internal and external sources. Because of the sheer number of state and local governmental open data sites and the wide variety of naming conventions used to publish them, it has proved challenging to aggregate a comprehensive list. Nonetheless, simple iterative web searches demonstrate that there are thousands of sites where U.S. federal agencies, states and localities share government data.

Opening data by simply putting it up on a webpage is a relatively easy governmental accomplishment, requiring only that a governmental unit decide that it would like to upload a spreadsheet to an existing webpage. Open data advocate Waldo Jaquith has pointed out that governments can create an open data site simply by running a search for CSV, XLS, and XML files that already exist on their network of public sites and posting them all together on the same page.⁷ The Open Knowledge

⁴ See, for example, the Open Energy Information plan at <http://www.whitehouse.gov/open/innovations/OpenEnergyInformation> or the Presidential Open Government Report at <http://www.whitehouse.gov/sites/default/files/microsites/ogi-progress-report-american-people.pdf>.

⁵ Robert Bobb, “Memorandum Re: Streaming of DC Stat Data to www.dc.gov,” http://www.scribd.com/fullscreen/26442622?access_key=key-20rfsh26eu0ob66xlbm, accessed May 31, 2015.

⁶ Socrata, “Opening Government, One Data Set at a Time,” <https://web.archive.org/web/20100208173200/http://www.socrata.com/about>, accessed May 31, 2015.

⁷ Waldo Jaquith, public presentation at Open Data NJ Summit, May 16, 2014, Montclair, NJ.

Foundation suggests beginning with an even easier method of entry, by opening up “just one dataset.”⁸

However, to ensure lasting access to government data, governments must take the additional step of legally codifying public access to data. Since the advent of open data publication, governments have known that they need to establish regular and sustainable practices of data collection, review, and release in order to ensure the quality of the information they publish. Open data policies create clear lines of responsibility for the management and oversight of data publication, create official spaces for public participation around data selection and publication, and ensure sustained commitment from government—all of which increase the value of the data to its potential users.

Policies to achieve these goals have been emerging for over a decade, but, as was the case with the initial development of open data portals, the development of these policies began without the “open data” label. Washington, D.C.’s 2006 administrative memorandum, for example, does not mention “open data” but contains many of the elements of what we would now call an open data policy: mandating an “open” rationale for streaming data online, describing a timeline for data release, mandating specific agency responsibilities and describing the need to maintain data quality through a review process. After the 2009 White House Transparency and Open Government Memorandum, meanwhile, state and local governments began to develop policies that used the language of “open data” more explicitly. The Portland, Oregon City Council, for example, passed a resolution “to mobilize and expand the regional technology community . . . by promoting open and transparent government, open data, and partnership opportunities”⁹ while the mayor of San Francisco, California issued an Open Data Executive Directive to “enhance open government, transparency, and accountability by improving access to City data that adheres to privacy and security policies.”¹⁰

In addition to providing good examples of early municipal open data policies, the Portland and San Francisco examples also demonstrate the two chief paths which emerged for passing a state or local open data policy: open data executive orders and open data legislation. Since 2009, a number of states and cities have chosen to follow both paths. As of this writing, the Sunlight Foundation has documented over fifty formal state and local open data policies: around two-thirds of these policies were established through legislative means (by law, resolution or ordinance) and around one-third

⁸ Open Knowledge, “How to Open Data,” <https://okfn.org/opendata/how-to-open-data/>, accessed May 31, 2015.

⁹ “City of Portland: Resolution No. 36735,” <http://www.portlandonline.com/shared/cfm/image.cfm?id=275696>, accessed May 31, 2015.

¹⁰ Gavin Newsom, “Executive Directive 09-06: Open Data,” <http://sfmayor.org/ftp/archive/209.126.225.7/executive-directive-09-06-open-data/index.html>, accessed May 31, 2015.

were created through executive means (by memo or directive). In at least one case, San Francisco, an open data policy originally established by the executive was later superseded by a legislative policy.

The pace of local development of open data policies has increased in the years following these earliest examples. Open data policies have been enacted in all of the most populous American cities—Los Angeles, New York City, and Chicago—in a number of mid-sized cities, and in places as small as Williamsville, New York, a village of 5,277 residents. As of 2015, eight states—Texas, Illinois, Utah, New York, Hawaii, New Hampshire, Connecticut, and Maryland—have passed open data policies, while other states, like Ohio, Michigan, and California, are presently debating open data measures.

Because it involves political processes, achieving an open data policy is more complicated than simply putting government data on a website. However, despite that difficulty, the value of a formal open data policy lies in the fact that it offers the public far a greater guarantee of access to specific datasets, in specific formats and at a specific level of quality. Without a formal policy, the public may lose access to posted data if the government website is revised or there is a change in departmental staff. Individuals who posted those datasets may choose to keep them current, or they might not. They may choose to make the information available in a format which makes it easy to use and reuse, or they might not. Moreover, open data policies usually describe a specific rationale for making government data available to the public, and this formal statement provides people who interact with government data an opportunity to make a case for access to new or existing data.

The legislative and executive methods of providing open access to government data are certainly not mutually exclusive. In many cases, governments begin by putting open data online in an ad hoc way and then expand on this initial activity by developing a formal policy about their practice. Individuals who are interested in obtaining greater access to government data may gain the necessary backing for a broader-scale policy by building momentum gradually, through both political support for open data and by demonstrating the value of existing government open data sets.

The past decade has seen the broad diffusion of the idea that providing open access to data is in itself an appropriate and laudable government function. From the earliest efforts to make data available online to the growing proliferation of formal open data policies, governments are increasingly willing to make their data available for public use and reuse. It is unclear whether the increasing pace of open data policy enactment will continue—or whether we will ever actually see a majority of

government entities adopting open data as a formal policy—but if the norm of governments proactively providing their data for public use continues to spread, we may achieve a similar outcome even without the adoption of official policies.

WHY ARE STATE AND LOCAL GOVERNMENTS OPENING THEIR DATA?

While there are precedents for open data publication, we are in the midst of a rapid period of adoption of open data policies. Governments rarely demonstrate such rapid change without strong motivation. In our conversations with government leaders about their interest in open data, several key reasons for releasing data publically have emerged.

Needing to do more with less

Many governments are increasingly aware of the disconnect between what they are capable of doing with their own data and what the private sector is accomplishing with similar datasets. This gap is an unfortunate but predictable consequence of the current state of state and local government funding and employment, which even in 2015 remains below pre-recession levels. Budgetary difficulties in adding new technological capacity means that, in many cases, governments must try to work solely with existing resources and staff.

In this context, open data provides a useful method to broaden the range of people who can derive benefit from government data, thereby increasing governmental capacity at low cost. Open data connects government staff with external data users in ways that provide more oversight, more idea-generation, and more analysis than government actors may be able to accomplish on their own. This is a particularly attractive possibility given the increasing number of data-using groups, such as Code for America brigades, data-focused journalism centers, and for-profit data analysts. While volunteer “civic hacker” groups are one well-known target of government outreach around open data, university researchers, software entrepreneurs, and activist groups are all likely to be able to perform useful data work that helps crystallize public problems, identify gaps, and improve services within a jurisdiction.

Improving internal data use

The fact that open data broadens the range of data users directly benefits internal governmental data users as well, and these benefits become particularly clear when it comes to interagency data sharing. Governmental data users often complain about the complications of “siloed” data and about the time and political cost of obtaining data held by separate departments. Once data is made publicly open and proactively posted online, data users within government gain full and regular access to it. This improved data sharing across departments helps enhance workflow in several beneficial ways. Data shared through a common open data portal makes data sharing seamless, eliminating the work involved in requesting and fulfilling internal requests for data. Data available on a common portal gives internal researchers and evaluators more possible measures for evaluating programmatic options. Agencies that have common or overlapping data needs can obtain efficiencies by using the data shared on a common portal to reduce duplicative data-collection work, conduct research projects by using data already collected and made available by another agency, and see ways to align measures for greater cross-governmental consistency. With greater experience in data sharing, agencies are likely to have additional conversations about the data they share, sparking opportunities for new interagency collaborations.

Regular practices that accompany open data initiatives also improve the quality and efficacy of internal data use. Open data programs often feature formal or informal inventorying processes to identify the datasets that different departments hold. Data inventories list datasets and their basic qualities. They often identify whether certain datasets are suitable candidates for open data release. Through the interdepartmental sharing of information about what data exists within a government body, internal data users gain the opportunity to learn comprehensively about data collection occurring in other departments that might be relevant to their work. Even when some datasets may be too sensitive for open publication, the inventory process can provide valuable information about the existence of these datasets to internal stakeholders, who can then create Memoranda of Understanding (MoUs) or other legal mechanisms to achieve improved internal data sharing.

Open data publication requires departments to explore their existing data collection and publication practices and consider improvements. With each dataset gaining additional potential users and reviewers, data quality across the public datasets is likely to improve in a number of ways. Open data initiatives allow internal users to become more aware of data holdings, which can lead to improvements in data quality. Internal evaluations can help departments identify areas where additional data collection could help improve the delivery of government services.

In addition to fostering internal examination, making more data publicly available will inevitably open the door for public scrutiny that will ultimately benefit both government entities and the community. Public review can be used by government to gauge data quality, help agency staff address necessary improvements to data collection and organization, and decide whether additional datasets should be created. Opening data also empowers both the public and policymakers to determine whether data is being appropriately managed for the benefit of the public.

Where critical information about data fields or collection processes is missing, making it difficult for anyone but the dataset's creator to use it, the process of readying that data for publication to an open data portal will force data creators to improve their provision of that metadata, and is likely to generate regular improvements in their information management practices. Where datasets feature missing values, the process of preparing data for broader publication is likely to compel data managers to address and improve data quality. As data is prepared for public release, problems in consistency and interoperability come to the fore, and data managers have an opportunity to change practices across departments to make their data collection processes more broadly effective and useful.

Improving two-way communication between governments and the public

In addition to solving internal data-sharing issues, open data is a method by which governments seek to improve their communication with community residents. Two-way communication with the public is key to achieving municipal goals. Governments have a great deal of information they would like residents to have, from details about how services are provided to explanations of how they are meeting existing challenges. In many cases, governments are also interested in receiving public feedback about existing programs and issues, as well as timely information about new problems. While social media, online public meetings, and emailed press communiques all demonstrate valuable uses of technology to increase public awareness and invite public commentary, open data provides a new way to think about enhancing two-way communication between governments and residents.

The decision by a government to release raw data through an open data program creates a different kind of information flow than can be achieved through traditional public relations practices.

Open, data-powered communication functions by inviting data-users to derive both

anticipated and unanticipated insights from government data. In a significant way, open data allows governments and members of the public to co-create public information, providing new ways of sending information out, analyzing it, and getting responses back.

Some of the key communications functions enabled by open data are:

- **Collaboratively identify problems:** Although governments and community organizations may be aware that certain aspects of public regulation or service delivery are not functioning optimally, having precise information about the nature of problems makes addressing them easier. Providing open data about public service provision allows interested parties to analyze relevant information based on their concerns and learn more about the nature of the issue that concerns them. Having precise information then allows these analysts to participate in leading a public conversation about what kind of change communities need. Some recent examples of this kind of public, data-focused problem analysis include:

- Code for DC’s analysis of the distribution of supplementary education funding for “at risk” students, relative to the number of eligible students per school: atriskfunds.ourdcschools.org
- Baltimore Neighborhood Indicators Alliance’s map of densities of children showing elevated levels of lead in blood tests: bniajfi.org/wp-content/uploads/2015/03/Leadtest13map.jpg
- The Anti-Eviction Mapping Project’s map showing the frequency of evictions in San Francisco which are justified by reference to California’s Ellis Act, a law permitting landlords to evict tenants if the landlord is “going out of business”: www.antievictionmappingproject.net/ellis.html

- **Make use of external expertise:** Open data supports the quick development of informal or ad hoc public-private partnerships that can be used to let the community

KEY COMMUNICATIONS FUNCTIONS ENABLED BY OPEN DATA

- ✓ Collaboratively identify problems
- ✓ Make use of external expertise
- ✓ Improve access to public services and information
- ✓ Improve methods of collecting public input

answer specific public questions. Because data-analytic capacity is often a carefully-metered resource within governments, having additional, more flexible external capacity makes it possible to answer those questions that might be less urgent—but which are nonetheless sometimes very important. Some examples of the kind of external analytic work that can be done by organizations or individuals using open data include:

- Datakind worked with open data from the city of New York to evaluate the costs and benefits of a more aggressive tree-pruning schedule. They determined that pruning trees more frequently would reduce emergency cleanup costs resulting from downed electrical wires and other consequences of less-maintained tree stock by 22%.¹¹
 - Data Science for Social Good, a project of the University of Chicago, analyzed open data from Chicago to learn that speedier replacement of broken streetlights had a positive impact on crime frequency, all other variables held equal.¹²
 - I Quant NY, created by a Visiting Professor at the Pratt Institute in Brooklyn, New York, uses open data to answer a wide variety of questions about public services and regulations in New York. The questions that are explored on the site range from the impact of gentrification on the availability of laundromats to New York City residents.
- **Improve access to public services and information:** With the advent of APIs (Application Programming Interfaces), all well-structured data sources can be automatically “plugged in” to a variety of online platforms. This is useful because very often, people do not use available government information simply because they don’t know it exists. While governments ultimately have limited capacity for public outreach, providing information openly allows non-governmental actors to convey that information in ways they know to be useful. A few examples of how open data works to improve public access to government service include:
 - Yelp’s integration of local health departments’ restaurant inspection data into their online restaurant search-and-review platform. The growing adoption of a standard restaurant inspection data publication scheme makes it increasingly easy for governments to provide greater public access to information about the safety of local restaurants. Yelp’s real-time integration of that data into its other services allows it to be used in public decision-making.
 - California’s open provision of Healthcare Associated Infections (HAI) rate data, by hospital, has made it possible to map the incidence of hospital-acquired

¹¹ “Out on a Limb – For Data.” DataKind. <http://www.datakind.org/projects/out-on-a-limbfor-data/>

¹² Seeskin, Zach. “CDOT: Using Data to Shine a Light on the Impact of City Services.” Data Science For Social Good. <http://dssg.io/2014/02/06/cdot-streetlights-crime.html>

infections to promote comparison by patients in advance of choosing a hospital. The integration of this information into public decision-making practices may be even farther along in New York, where HAI data has been included in reviews of provider-associated hospitals on DocSpot.com, a medical provider rating-and-search engine.

- New methods of increasing uptake of social services take advantage of the popularity of SMS messaging services as a method of reaching potential clients. mRelief, a program which allows municipal residents to quickly check whether they qualify for different social services, translates municipal open data on eligibility rules into an SMS-based questionnaire that potential clients can use to screen themselves before committing their time—and the time of municipal workers—to in-person visits to social service offices.
- **Improve methods of collecting public input:** Just as open data allows third-party websites to share government information with the public, open data also makes it possible for developers to create new ways of receiving public input. By making government information easy to integrate into sites that also collect public feedback on government processes, open data can increase the range of opportunities the public has to provide useful information to government.
 - South Bend Cityvoice (www.southbendvoices.com) is a project of the local Code for America brigade intended to improve opportunities for local residents to inquire into the status of vacant or abandoned properties. It uses open parcel data to create an interactive map that community members can use to identify problem properties, see other inquiries, and access the municipal response to citizen complaints.
 - Code for America fellows have used data on public records requests to create a system called RecordTrac (records.oaklandnet.com) for the city of Oakland. RecordTrac makes it easier for residents to see past public records requests and to submit new ones. Public records requests are an excellent signal of citizen interest and can be used to identify datasets that could be made proactively available through a municipal open data program.
 - The OpenGov Foundation created The Madison Project, a platform that allows citizens to monitor and comment on legislation as it comes before a legislative body. Currently implemented in Washington, D.C. (<http://dc.mymadison.io>), the platform uses open legislative data to allow the public to track and respond to legislative change as it occurs.

While all of the functions of open data powered two-way communication with the public can be accomplished by other means, these functions—public collaboration around problem identification, improving access to raw information, and collecting public input—are core to the expectations of the open data model.

HOW CAN A GOVERNMENT GET STARTED WITH OPEN DATA?

Democracy is improved when people have data to help them understand how leaders and their policies are delivering results. Because data-users have different interests, it is optimal for governments to involve as many departments and release as much data as possible without violating privacy restrictions and security concerns. In other words, to begin the process of opening up government data, governments should consider making their datasets open by default, essentially making transparency “opt-out.” While most government disclosure laws (i.e., FOIA) entail reactive disclosure, the goal of open data is to foster proactive disclosure: releasing information before it has been requested. “Defaulting to open” means aiming to become proactive in making the information governments have gathered available for use and reuse.

Meeting and taking stock

A first step to pursuing an open data program is to identify the potential champions for an open data initiative within government. A second step is getting them together. Meeting with internal advocates for an open data program is an essential first step to figuring out how open data might be integrated into the government’s existing practices. In preparation for a group meeting, it may be useful to review the range of existing resources available for getting started with open data. Some of the resources available to help advocates become familiar with open data include:

- The Sunlight Foundation’s Open Data Policy Guidelines: sunlightfoundation.com/opendataguidelines
- Code for America’s Open Data Playbook: www.codeforamerica.org/governments/principles/open-data/
- The Open Knowledge Open Data Handbook: opendatahandbook.org/guide/en/
- Socrata, Inc.’s Open Data Field Guide: www.socrata.com/open-data-field-guide/
- Mark Headd’s Open Data Guide: opendata.guide
- Josh Tauberer’s Open Government Data: The Book: opengovdata.io

A major goal of early meetings should be to identify the city's goals for an open data program. Is the program fundamentally aiming to improve information flows within government? To improve the transparency of government operations to the public? Will it aim to create public-private partnerships to extend government's analytic and programming capacity? Does it aim to generate new methods of reaching citizens directly? Clarifying a set of specific goals for an open data program will help the program develop in line with an articulated set of outcomes and also help advocates clearly communicate about the program in public.

Once internal advocates have begun to articulate a vision for a municipal open data program, it is valuable to identify what the government is already doing to open its data. Perhaps there are a number of datasets which are already published to the government's websites, or maybe there are regular data-sharing relationships in place with certain organizations that could fairly easily be more broadly opened up for wider sharing. Searching the data that is already available on the government's websites is an excellent place to begin. A second point of collection lies in exploring which data has been mentioned most recently in public legislative or executive processes. If government officials are communicating about data with the public, it's a good sign of an interesting dataset—and members of the public would most likely be interested in exploring the full dataset that is under discussion as well.

Engaging the Public

Next, once internal conversations have begun, it is useful to open the discussion to public actors as soon as it is viable. An important key to success with an open data program lies in two-way communication with the public prior to and during the data release process. Open data depends on both internal and external partners, so engaging those external partners as early as possible ensures that the program develops in a way that maximizes their potential involvement.

It is useful for public data users to have an opportunity to weigh in on what information they want the most. For example, when New York City opened its data, citizens were asked to engage in the process through the city's open data portal, and encouraged to submit feedback about the data online, including suggestions for new data collection. Governments can also run focus or advisory groups to determine what information users want and discover more about what they'd like to do with it. As an additional benefit, public oversight of the open data development process engages external stakeholders in the program's success, increasing the chances that they will in turn become ambassadors for the program.

Developing partnerships with open data groups, businesses, and other community partners can help streamline the data release process and support better decision making. Outside experts can help to identify and create connections between community members and the government to examine priorities for data release. During Cook County's open data initiative, for example, the government partnered with the Smart Chicago Collaborative to use the organization's expertise in developing open data sets and applications to benefit the public, as well as to engage the open data community with the county's newly available data. Partners may also help the open data process along by identifying existing data from outside groups such as nonprofits, think tanks, and academic institutions that can be used in place of or alongside government data to increase government data's utility. Finally, well-established community partners might work with the government to help mitigate any trust issues that exist between the public and government authorities or agencies.

Publishing data

Data publication is an ongoing process. Collecting together the datasets that are already published in one spot represents an excellent starting point, if the government does not yet have a formal portal.

If the government wants to develop a formal open data portal, this can be done either in-house or by contracting with an external vendor. All open data portals, no matter how simple, should aim to be externally findable; for this reason, they must permit indexing and searching by search engines and other third parties. It is useful if the portal contains a full, human-readable list of all of the datasets that are on the portal. It's also useful to have publicly-viewable analytics for the portal so that everyone can understand which data is of greatest interest to users at any one moment. One of the following lists of open data vendors and software might be helpful in making a choice about portal providers and solutions:

- Sunlight Foundation, "What portal should we use?":
sunlightfoundation.com/policy/opendatafaq/#portal
- Code for America, "Select an open data platform":
www.codeforamerica.org/governments/principles/open-data/#intro
- Open Data Stack Exchange, "Collect a list of open data systems":
opendata.stackexchange.com/questions/1728/collect-a-list-of-open-data-systems

Once an open data initiative has begun, there are a number of best practices governments can build

on to improve their open data publication. Governments should reach out to organizations with open data expertise to get answers to specific implementation questions. Some of the activities a government might choose to pursue to improve open data release include:

- Indexing existing data held by governmental units
- Referencing existing records access policies in a policy for proactive publication
- Determining processes for protecting individual privacy and safeguarding sensitive information in data release
- Developing a process for prioritizing the processing and release of open data
- Choosing a consistent metadata schema for government data
- Ensuring that opened data is machine-readable and available for bulk download
- Ensuring that published data has licensing terms that permit unencumbered public use and reuse
- Exploring implementing existing data standards to increase interoperability

Opening the door for two-way communication is just as important after the release of data.

Government officials should collect public feedback about the quality, ease of use, and completeness of the data. At this stage, they should continue to seek input about what new data citizens want to see. An open dialogue between cities and the public has the additional benefit of improving community access to and uptake of public services, and it increases civic engagement overall. Public suggestions for improving open data practices or policies can be solicited through online tools on city websites. Some cities, such as Albuquerque, Louisville, and Phoenix, allow the public to submit feedback directly through their open data portals. Another practical way to keep up the open data momentum is to highlight past successful results of open data projects. It is common practice for cities with open data portals to achieve this by featuring applications created with open data on city websites or data portals. It is also a good idea to notify potential users when new data is added or when existing datasets are updated. In this process, cities could consider targeted outreach, for example, reaching out to hospitals when new healthcare data is available. The Open Knowledge Foundation additionally suggests contacting organizations that work with open data using existing mailing lists or social networking groups, or directly contacting users who are known to have an interest in the data.

To engage members of the public new to working with government data, leaders can use a wide range of methods. More traditional methods for outreach, such as press releases or other public announcements, might be a good starting point. Governments can also choose to get more creative in their outreach. Contests, for example, are an innovative way to advertise open data. In 2008, Washington, D.C. hosted one of the first competitions to increase citizen engagement with

open data. The City's Chief Technology Officer challenged developers to use newly-available data to build applications that would make it useful for citizens, businesses, visitors, and city agencies. In one month's time, 47 applications, with an estimated value added to the local economy upwards of \$2.6 million dollars, were created, paving the way for similar challenges in other cities.¹³ In 2014, the city of Los Angeles, California launched an annual Mayor's Challenge, in which city employees compete to win a Civic Innovation Award for using data from the city's open data portal to develop innovative solutions to the city's greatest challenges.

Connecting with others

One of the key features of open data programs is their potential to facilitate connections, based on data sharing, that extend beyond the offices where that data was collected. This characteristic of data-based connecting goes to the heart of not just the activity of sharing data, but also to the ethic of the community of open data users and advocates. As a government builds an open data program, it should regularly seek to tap into the networks of open data practice that have developed across the country and the world. With many governments and organizations thinking creatively about maximizing the value of data release, the conversations between partners in the private, public and nonprofit sectors are lively and productive. Very quickly, governments new to open data find that they have their own discoveries and examples to share, and that they are suddenly playing an important role as leaders for the next generation of open data publishers. The end result is improved access to information, opportunities for improved decision-making and service delivery, and improved communication between communities and their governments about the issues that matter to them most.

¹³ Open Knowledge, "So I've Opened Up Some Data - Now What?" <http://opendatahandbook.org/guide/en/following-up/>, accessed May 31, 2015.