

Version 1.0

# OPEN DATA PLAYBOOK





# CONTENTS

- 1 INTRODUCTION ..... 1**
- 1.1 THE CITY OF SEATTLE’S OPEN DATA PROGRAM..... 1
- 1.2 ROLES & RESPONSIBILITIES ..... 2
- 2 HOW TO CREATE A DATASET INVENTORY ..... 3**
- 2.1 WHAT IT IS..... 3
- 2.2 HOW TO CONTRIBUTE ..... 3
- 2.3 GUIDING QUESTIONS..... 3
- 2.4 TEMPLATE ..... 4
- 2.5 DEFINITIONS ..... 4
- 3 HOW TO PUBLISH A SINGLE DATASET ..... 8**
- 3.1 SELECT A DATASET ..... 8
- 3.2 CONVENE DATASET TEAM ..... 8
- 3.3 COMPLETE THE OPEN DATASET SUBMISSION FORM (METADATA)..... 9
- 3.4 SUBMIT TO THE OPEN DATA TEAM FOR REVIEW..... 10
- 3.5 DATASET TEAM: PUBLISH THE DATASET..... 10
- 3.6 REVIEW AND UPDATE ..... 11
- 4 HOW TO MAKE THE BEST USE OF OPEN DATA ..... 12**
- 4.1 ENGAGE STAKEHOLDERS..... 12
- 4.2 PARTICIPATE IN LARGER EFFORTS ..... 13
- 4.3 PARTICIPATE IN LOCAL EFFORTS ..... 14
- 4.4 JOIN THE GLOBAL COMMUNITY ..... 14
- 5 SUPPORTING DOCUMENTS ..... 16**



# 1 INTRODUCTION

Under the City of Seattle’s open data policy, each department is responsible for publishing its own data, with centralized support from the open data team in Seattle IT. As outlined in the policy, each department appoints an “open data champion” to lead their department’s work and participate in the Citywide network of champions. This playbook is intended primarily for open data champions in the City of Seattle. Whenever we use the word “you,” we are referring to the open data champions. Whenever we use the word “we,” we are referring to the open data team in Seattle IT.

## 1.1 THE CITY OF SEATTLE’S OPEN DATA PROGRAM

The City of Seattle’s open data program launched in early 2010 with the launch of our open data portal, [data.seattle.gov](http://data.seattle.gov). Since then, Seattle has become a leader in making government data available to the public. Our program is ranked sixth in the country by the Code for America Open Data Census, and in 2015, the Governing Institute named the program its Citizen Engagement Project of the Year.

We were the first in the nation to publish 911 dispatch data from our police and fire departments in near-real-time, which remain among the most popular datasets on our portal. Currently, there are more than 600 datasets and maps available. They include information on schools, public facilities, public art, and numerous other publicly regulated activities that affect life in Seattle. Approximately 14,000 people visit our portal each month.

In 2016, under Mayor Ed Murray, the City created an official Open Data Policy that is backed by an executive order from the Mayor. It directs all City departments to make their data “open by preference,” which both sets the expectation that public data will be public and makes clear that we have a responsibility to protect privacy. The new policy lays out clear goals that guide our work as a program.

Over the next few years, our program will focus on quality over quantity, prioritize datasets based on the public interest, improve governance of our open data through improved metadata and regular updates to datasets, and engage our users proactively throughout the process. We want to publish data that is meaningful and usable, maximizing the benefit to the public of the resources we spend to manage the program.

## 1.2 ROLES & RESPONSIBILITIES

### Seattle IT

Seattle's IT Department, led by the Chief Technology Officer, provides leadership for the overall open data program, liaising with the Mayor's Office and City Council as necessary, oversees the policy, and, via the privacy program, ensures that publication of open data is consistent with our privacy policy.

### Open Data Team

The open data team, part of Seattle IT's Digital Engagement division, is responsible for day-to-day management of the open data program, including training, process management, and quality control for all datasets published to the open data portal, as well as the open dataset inventory. The open data team is responsible for producing an annual report to the Mayor and the public, an annual plan that outlines the program's strategy for the coming year, and an annual risk assessment of both the program and all the data available on the portal.

Additionally, in collaboration with the City's Civic Technology Advocate, the program offers:

- Data Camp, an annual training for open data champions
- Breakfast of Champions, a monthly engagement for open data champions that blends program management, training, and exposure to civic technology opportunities
- Regular opportunities to engage with the civic technology community in Seattle and beyond

### Open Data Champions

Each department's open data champion is responsible for managing their department's participation in the program. They work with leadership to set priorities, oversee the publication and ongoing management of datasets for their department, and participate in ongoing activities with the network of champions. They also keep the open dataset inventory up-to-date and contribute to annual reports and plans.

### Data Owners

Data owners across City government contribute datasets to the inventory and to the portal.

### Department Leadership

Department directors and managers guide their department's participation in the open data program, setting open data-related performance goals for their teams, updating policies and procedures to reflect the open data policy, and making sure their staff have the time and resources to participate in the program. As needed, they work with open data champions to set priorities, engage stakeholders, and work through any sensitivities in their datasets prior to publication. They consult with open data champions at the start of new projects and software implementation to ensure that they facilitate compliance with the City's open data policy.

## 2 HOW TO CREATE A DATASET INVENTORY

### 2.1 WHAT IT IS

The open data program's dataset inventory serves as:

- a way to see, at a glance, what data the City has that might be appropriate for publication as open data under the new open data policy
- a tool for the open data team to keep track of all the potential open data in the City so we can manage the program and report up to the Mayor's Office and out to the public
- a tool for open data champions to keep track of their department's open data and make sure it is managed well on an ongoing basis

The dataset inventory is part of the City's comprehensive data inventory, which is managed by the Chief Privacy Officer. Unlike the comprehensive data inventory, which includes all data collected and stored by the City, the open data program's dataset inventory is focused on existing datasets that are already in use. This is a first step that may or may not lead to publication of the dataset, in full or in part.

### 2.2 HOW TO CONTRIBUTE

Each department is different in terms of how it is managed, how much data you own, what *kind* of data you own (ranging from reports created with other people's data, usually published in pdf, to huge database systems where datasets are not always easy to extract), and how you work together as a department. Depending on what works for you, you may want to start with any or all of these steps:

- meet with your department director and/or team leaders to share what open data is, what the Mayor's expectations are for the inventory, and how you might best go about soliciting input from data owners
- meet with your department's communications team, public information and/or public records officer, or persons responsible for performance management reporting
- meet directly with data owners to talk to them about what they own

At this point, do not exclude any datasets based on privacy, security, sensitivity or quality concerns. Every dataset will be reviewed and approved before it's published.

### 2.3 GUIDING QUESTIONS

- What data does your department use for internal performance and trend analysis?
- What data populates your monthly or quarterly reports?
- What information is published as a performance metric (e.g. on [performance.seattle.gov](http://performance.seattle.gov))?
- What information do you report to local, state, or federal agencies?
- What information do you share with other City departments?
- What information do you share with external partners?
- What information is repeatedly requested by the public, via the public disclosure process and/or open data requests?
- What kinds of open data are your peer agencies across the country publishing?  
*Hint: Check out [Chicago](#), [Philadelphia](#), and [San Francisco](#) to start*

## 2.4 TEMPLATE

The open data program has provided a template for the common inventory, currently available as an Excel spreadsheet. It includes fields in three categories:

### Basic information

Department	Work Group	Dataset Name	Data Description	Data Source	Data Source Type	Technical Details	Data Owner/Contact	Technical Contact
------------	------------	--------------	------------------	-------------	------------------	-------------------	--------------------	-------------------

### At-a-glance assessment

Assessment							Notes
Quality	Privacy/PII	Security	Sensitivity	Performance	Priority		

### Post-publication tracking

Date Published	Update Period	Date Last Updated	Date of Next Update
----------------	---------------	-------------------	---------------------

## 2.5 DEFINITIONS

The at-a-glance assessment is a tool for the inventory that is intended to assist with prioritization. It is not a substitute for the in-depth data quality and privacy assessments that are required prior to publication. In the template itself, you will find definitions for each assessment category as well as examples. Most fields are a drop-down.

Here are the definitions as included in the template:


### Quality

- Good* This dataset is used routinely, is managed in a way that makes it legible to users besides the owner, and is reasonably accurate
- Acceptable* This dataset is used routinely but not necessarily legible to non-owners and/or has some gaps or discrepancies that limit its usability
- Poor* This dataset exists but is not regularly used, updated, or otherwise managed in a way that makes it valuable

### Privacy/PII

This is a flag only; datasets will be reviewed for privacy prior to publication. PII includes but is not limited to:

- Name and initials in any combination
- Identification number (e.g., Social Security #)
- Birthdate



Age  
Gender  
Home address  
Home telephone number  
Personal cellular, mobile or wireless number  
Personal e-mail address  
Drivers' license number  
Information on medical or health conditions  
Financial information (credit card/PCI, billing info, account info)  
Health information  
Marital status  
Nationality  
Sexual behavior or sexual preference  
Physical characteristics  
Racial or ethnic origin  
Religious, philosophical or political beliefs  
Trade union membership  
Biometric data  
Household information  
Consumer purchase or billing history  
Unique device identifiers (IP/ MAC addresses)  
Location (e.g., GPS) info (including that provided by mobile devices)

## **Security**

### *Public*

Public information can be or currently is released to the public. It does not need protection from unauthorized disclosure, but does need integrity and availability protection controls. This would include general public information, published reference documents (within copyright restrictions), open source materials, approved promotional information and press releases.

Examples:

- Information provided on City web sites
- Information for public distribution (e.g. budget documents after public release)
- GIS maps
- Meeting agendas and minutes

### *Sensitive*

Sensitive information may not be specifically protected from disclosure by law and is for official use only. Sensitive information is generally not released to the public unless specifically requested. Although most all of this information is subject to disclosure laws because of the City's status as a public entity, it still requires careful management and protection to ensure the integrity and obligations of the City's business operations and compliance requirements. It also includes data associated with internal email systems and City User account activity information.



Examples:

- Work phone numbers
- Organizational charts
- Interdepartmental documents
- Policies, procedures, and standards

*Confidential* Confidential information is information that is specifically protected in all or in part from disclosure under the State of Washington Public Disclosure Laws ([See Link to State Disclosure Law in Appendix B](#)).

Examples:

- Personally Identifiable Information (PII)
- Information concerning employee personnel records
- Information regarding IT infrastructure and security of computer and telecommunications systems, information security plans
- Information related to law enforcement (e.g. witness protection information)
- Information related to minors (e.g. adoption and foster records)

### *Special Handling*

Some confidential information is specifically protected from disclosure by law and subject to strict handling requirements dictated by statutes, regulations, or legal agreements. Serious consequences could arise from unauthorized disclosure, such as threats to critical infrastructure, increased systems vulnerability and health and safety, or legal sanctions. Departments handling this category of information must demonstrate compliance with applicable statutes, regulatory requirements and legal agreements.

Examples:

- Payment Card Industry (PCI) information
- North American Electric Reliability Corporation (NERC)
  - Performance and standards metrics
  - Risk management and capability information
- Criminal Justice Information Services Division (CJIS)
  - NCIC data
  - Open investigations
- HIPAA personal health records information, such as:
  - Health insurance
  - Patient treatment information
- Family Education Rights and Privacy Act (FERPA) information, such as:
  - Tax records of parents and students
  - Grades



  
**Sensitivity**

*High* This information is considered sensitive by stakeholders, leadership, or the Mayor's Office, such that it would likely require outreach to those groups prior to publication (e.g. labor, vendors, impacted populations)

*Medium* This information has an impact on stakeholders that should be taken into account, but is unlikely to be disruptive to ongoing processes

*Low* This information is already public in some form or does not contain data that would be surprising to stakeholders

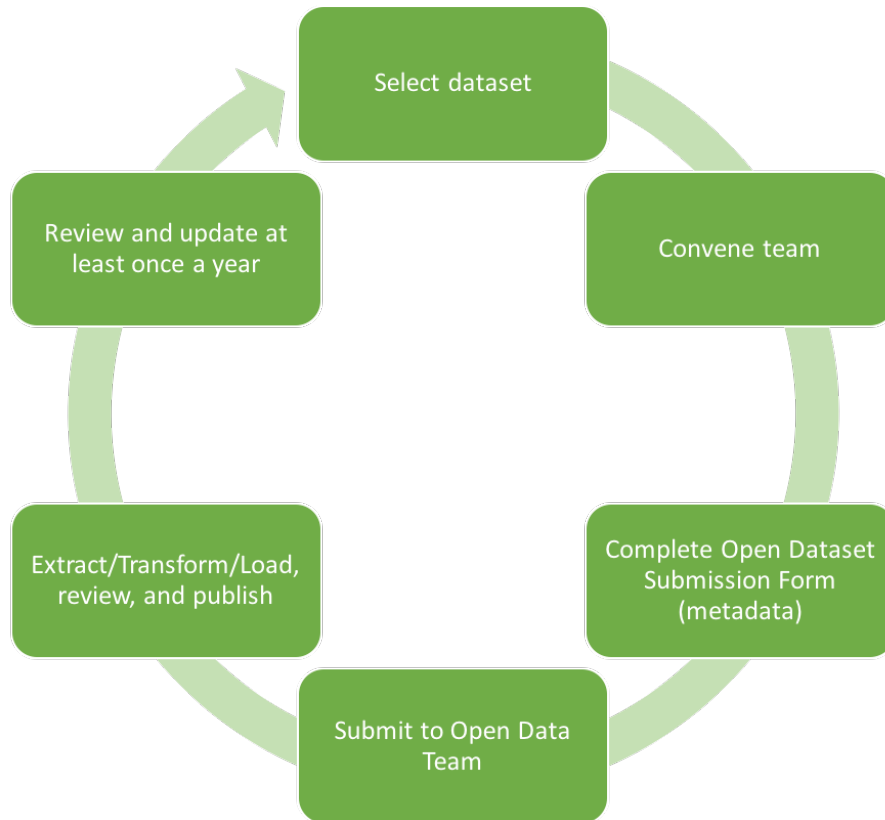
**Priority**

*High* There is demand for this dataset by the department and/or the public, or it is relevant to high-profile work or objectives

*Medium* The dataset should be published but is not an immediate priority

*Low* There is no apparent demand yet for this dataset

## 3 HOW TO PUBLISH A SINGLE DATASET



### 3.1 SELECT A DATASET

While the City’s open data policy makes our data “open by preference,” in reality, publishing each dataset requires work on both a one-time and ongoing basis – to standardize the data, make it machine-readable, write good metadata, and take steps to protect privacy and mitigate security risks. Each department will need to prioritize which datasets to publish based on a variety of factors, including input from the public.

### 3.2 CONVENE DATASET TEAM

To prepare the dataset for publication, you will need a team. Here’s who you’ll want on it:

- **Open data champion**  
The person responsible for oversight of all of the department’s open data
- **Dataset owner(s)**  
The person or people who manage the data and/or work related to it, who understand what it is used for, who can provide the context/metadata that ensures the public understands what it means, and who can lead stakeholder engagement

- **Technical resource**

The person who has the technical skill to extract datasets from databases and outer sources, prepare the data for publication and ultimately publish the data to data.seattle.gov

These do not need to be separate people. For example, depending on the dataset, the business owner may have the technical skills or also serve as Open data champion. You will enter the information about who is on the team, and which member is responsible for which task, into the Open Dataset Submission Form.

### 3.3 COMPLETE THE OPEN DATASET SUBMISSION FORM (METADATA)

To submit a dataset for publication, you will fill out the open dataset submission form, which also acts as the metadata for each dataset. This information will be accessible to the public alongside the dataset itself. This is where you will want to put any information that does not belong inside the dataset but is important for understanding what it is and how to use it. Relevant links may be included, so long as you take care to keep them up to date.

The submission form has two parts: general dataset information and a sample record. Here is what you will be asked to provide:

- A title, utilizing the City’s naming standards
- Description & business purpose (any additional information about the dataset that will be useful)
- Keywords describing the data that help people searching for it
- URL’s that link to the department and other websites related to the dataset
- Primary contact information
- Source and source type
- How the dataset is extracted and prepared for publication
- How the dataset is published to data.seattle.gov
- How often the dataset is published to data.seattle.gov

For each column in the dataset, you will document the:

- Column (field name)
- Data type (text, date/time, geocode, etc.)
- Sample value
- Column metadata description (what the data in the column represents or means)

You will also include a one-record example of dataset with all fields populated. Here is a great example from the Sold Fleet Equipment dataset:

Field Name	Data Type	Value	Metadata Description
EQUIP_ID	Text	1404	Vehicle equipment identification number
YEAR	Text	2000	Vehicle manufacture year
MAKE	Text	PLYMOUTH	Vehicle make
MODEL	Text	NEON	Vehicle model
DESCRIPTION	Text	SUBCOMPACT SEDAN	Type of vehicle

DEPT	Text	SPD	City department
SALE_PRICE	Text	3700	Vehicle sale price
SOLD_BY	Text	BIDADOO ONLINE AUCTION	Sales agent

You can learn more about metadata for open data through this guide, published by the Center for Government Excellence at Johns Hopkins University: <https://centerforgov.gitbooks.io/open-data-metadata-guide/content/>

### 3.4 SUBMIT TO THE OPEN DATA TEAM FOR REVIEW

The open data team will review each dataset for privacy, security, and quality prior to publication. If there is a significant risk of privacy harm, they will escalate the review to the Chief Privacy Officer. If they notice security issues, they will work with you to reduce the risk of harms prior to publication. They will flag any quality concerns so they can either be addressed prior to publication or described in the metadata, so that users are aware of the dataset’s limitations.

### 3.5 DATASET TEAM: PUBLISH THE DATASET

There are three methods for publishing datasets to data.seattle.gov:

**Manual.** This is generally used for datasets that are relatively small and uncomplicated, and where the refresh rate is quarterly or longer. With some training, this method is easily mastered by non-technical data providers.

**Automated.** This is recommended when datasets will be updated more frequently than quarterly. The tool we use for this, Socrata’s DataSync, can also be mastered by non-technical data providers.

**Programmatic (API to API).** This is ideal for real- or near-real-time transfers. Application development skills are required.

For any of the methods above, the following steps are required.

#### EXTRACT

Export data from the source system. Every system is different so the method for doing this will vary. If possible, you should export only the data fields (or columns) that correspond with the schema you defined in the Dataset Submission Form (see Section 3.2 above).

#### TRANSFORM

If necessary, modify the exported data so that the data structure matches the data schema defined in the Dataset Submission Form.

#### LOAD

Using one of the publication methods described above, upload the data into data.seattle.gov.



### **PUBLISH PRIVATELY**

Initially, the dataset will be marked "private" and is limited to you, selected testers, and the open data team. The open data team will review the privately published dataset and work with data owners and appropriate stakeholders to make any needed changes. The open data manager will determine when it is ready for final publication.

### **OPEN TO THE PUBLIC**

Once the data is ready for publication, the open data manager will authorize it to be released to the public.

## **3.6 REVIEW AND UPDATE**

Once the dataset has been published, the open data champion is responsible for keeping it up-to-date, handling questions about the dataset from the public, and working with the open data team to resolve any issues that emerge. At least once a quarter, you should review all your department's published datasets to ensure they are being updated on the basis promised in the metadata. This is true for datasets updated automatically as well as those that require a manual update. The dataset inventory serves as a tool for you to keep track of what has been published so it is easier to verify that the data are up-to-date.

## 4 HOW TO MAKE THE BEST USE OF OPEN DATA

### 4.1 ENGAGE STAKEHOLDERS

Open data has many stakeholders: those who are impacted by the data directly, those who are impacted by the data indirectly, those who use of the data, and those who consume the data. Anyone who is affected by or interested in what's happening in our city is a stakeholder in open data.

We recommend engaging stakeholders at all stages in the process – as you determine demand for datasets from the public, as you prepare to release each dataset, and on an ongoing basis to ensure the dataset is as useful as it can be. This also allows you to see how people react to your proposed data before it is published, identifying areas of potential confusion or unnecessary limitations on its usefulness. It can also reveal opportunities for us to work together with outside groups to make the most of the data for the public good.

For example, a dataset that reveals information about Police Department activity is likely to have a broad group of stakeholders, from victims of crime and their families to officers and their labor union to the Chief of Police and the Mayor, as well as the general public. If we simply publish that data without first engaging with those whose lives it touches, we do them a great disservice. However, if we think about these many stakeholders from the beginning, that dataset can be the basis for an open and transparent conversation between the Police Department and the community.

Let's define each of these groups.

**Direct Stakeholders:** People and institutions who are represented directly in the data.

For example:

- Police officers identified in the officer-involved shootings dataset
- Non-profit providers of free meals listed in the meal programs dataset
- Owners of the buildings listed in the unreinforced masonry dataset

**Indirect stakeholders:** People and institutions who may be impacted by the release of the data or analysis conducted on it. For example:

- Residents of zip codes listed as having high rates of households below the poverty level
- Property owners in neighborhoods where crime rates are higher than average
- Students at schools that are underperforming

**Users:** People and institutions who will use the raw data. For example:

- Journalists who will look for possible stories in the data (e.g. [“How racially diverse is your school?”](#), [The Seattle Times](#))
- App developers who want to build tools on top of the data (e.g. [ArtQuest Seattle](#))
- Companies that ingest the data (e.g. [Zillow](#))

- City departments who use it to communicate with stakeholders (e.g. [Spacefinder](#) from the Office of Arts & Culture)

**Consumers:** Anyone who reads or uses the information. The list is endless, but this can include:

- People in the City administration who use it to inform their work and/or performance, such as the data owner, the department's leadership, the Mayor, and policy advisors
- People with a political interest in the work it represents, such as City Council members and their staff, community organizations, and activists
- People managing similar programs or who have similar goals, such as other jurisdictions, the county, the state, the federal government, philanthropists, and non-profits
- People who pay attention to what is happening in the city, including researchers, media, local businesses, residents, and people who work here, as well as those generally interested in urban planning

As you can imagine, there are many people and institutions who are impacted by the data in different ways. The more you can map them out as you think about your datasets, the easier it will be to make decisions about what to include (or what not to), what format to provide the data in, and what activities you might undertake around its release to ensure the data is well understood.

We recommend looking at all of this through an equity lens. The open data program provides outreach opportunities and can help to connect you with users of data as well as to support your efforts to engage those impacted by the data in the process, from publication to promotion. However, we rely on your subject matter expertise to consider the possible impacts and start brainstorming about who needs to be engaged.

Here's a great [Slideshare](#) from Laurenellen McCann of the New America Foundation that offers more insights.

## 4.2 PARTICIPATE IN LARGER EFFORTS

The City of Seattle has participated in national efforts to release more and better data, such as the What Works Cities program and White House Police Data Initiative. We also collaborate with regional jurisdictions, such as the county and state, and share best practices with other cities in our area. These connections allow us to share and adopt the best approaches, creating a community of practice that makes all this work easier in the long run.

There are many efforts underway to create and implement open data standards, allowing for comparability of data and interoperability of applications built for cities. Standards significantly improve the potential for useful analysis and drive down the costs of development for civic technology. For example, the General Transit Feed Specification, or GTFS, powers the OneBusAway app and allows for it to be used in other cities that publish transit information in GTFS.

Here are some well-known Open Data standards that we will strive to use in the future, along with examples of applications that use them, could use them, or could expand significantly if our City's data were available in this standard:

- [Open 311](#) (all non-emergency issue reporting)  
*Sample use cases: [FixMyStreet UK](#), [SeeClickFix](#), [Chicago Works](#)*
- [Open Referral](#) (community resources and social services)  
*Sample use cases: [Purple Binder](#) and [mRelief](#), Chicago*
  - [Open Eligibility](#)
  - [Human Services Data Specification](#)
- [Open Trails](#) (public trails and related geographic data)  
*Sample use cases: [Trails Editor](#), [Totago](#)*
- [Housefacts](#) (residential buildings)  
*Sample use case: [Trulia](#)*
- [Building & Land Development Specification \(BLDS\)](#) (commercial buildings and permits)  
*Sample use case: [Seattle in Progress](#)*
- [Open Contracting](#) (public contracting)  
*Sample use case: [dgMarket](#)*
- [General Transit Feed Specification \(GTFS\)](#) (transit)  
*Sample use cases: [OneBusAway](#), [TransitApp](#)*
- [State Decoded](#) (laws, codes, and statutes)  
*Sample use case: [LawHelp.org](#)*
- [Federal Spending Transparency](#) (budget information)  
*Sample use case: [USAspending.gov](#), [Open Budget](#)*

### 4.3 PARTICIPATE IN LOCAL EFFORTS


Seattle IT's new civic technology program works closely with the open data program to facilitate better communication between the City and the technology community. As a participant in the Network of Champions, you will have regular opportunities to learn about what's happening in the open data movement, to collaborate with civic technologists, and to participate in local events, from hackathons to ongoing volunteer projects. This information is shared internally via a listserv and a Sharepoint calendar for easy access and is mentioned at the monthly Breakfast of Champions meetings.

### 4.4 JOIN THE GLOBAL COMMUNITY

Open data is a global movement that is helping governments improve their operations, better communicate with the public, power analysis that makes cities "smarter," and stimulate the development of civic technology tools that improve people's quality of life. By becoming an Open data champion, you also become part of a growing network of public servants working to build 21<sup>st</sup>-century government. We think this is exciting work, and we hope you will, too.

If you want to explore open data in an open-ended way on your own, we suggest checking out the following online resources:



- 
- The U.S. federal government’s open data site [data.gov](https://data.gov)
  - Code for America, a national nonprofit that partners with government and builds open-source tools [codeforamerica.org](https://codeforamerica.org)
  - What Works Cities, a national networked initiative from Bloomberg Philanthropies [whatworkscities.bloomberg.org](https://whatworkscities.bloomberg.org)
  - The Open Data Handbook, a project of the Open Knowledge Foundation that includes resources and case studies [opendatahandbook.org](https://opendatahandbook.org)
  - Open Gov Foundation, a project of Open Gov [opengovfoundation.org](https://opengovfoundation.org)
  - The “I Quant NY” blog by data scientist Ben Wellington, plus his amusing & insightful TEDx talk (yes, we are encouraging you to use YouTube on City time!) [IQuantNY](https://IQuantNY.com) and [TEDx video](https://www.youtube.com/watch?v=...)

Social media is another great way to get engaged. Use the hashtags #opendata, #civictech, and even #civitech, and follow accounts like Code for America founder Jen Pahkla [@pahlkadot](https://twitter.com/pahlkadot), Sunlight Foundation analyst Alex Howard [@digiphile](https://twitter.com/digiphile), Code for America Brigade program manager Chris Whitaker [@civicwhitaker](https://twitter.com/civicwhitaker), GovTech magazine [@govtechnews](https://twitter.com/govtechnews), and the City of Seattle’s civic technology program [@civitechsea](https://twitter.com/civitechsea).



## 5 SUPPORTING DOCUMENTS

- [Open Dataset Inventory Template](#)
- [Open Dataset Submission Form](#)
- [Privacy One-Pager](#)