[Do not use!]
Please use our
New Documentation

DataSF Guidebook:
Data Coordinators Edition

City and County of San Francisco

(Version 3.0)

This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Table of Contents

- Table of Contents
- 1. Background
  - Purpose of this Guidebook and Version Notes
  - Why release open data?
  - History of Open Data in San Francisco
  - DataSF: Our Open Data Portal
- 2. Roles and Responsibilities
- 3. How to Create Your Data Catalog (Inventory)
  - Step 1: Identify data sources
  - Steps 2 and 3 Guidance
    - Step 2: Brainstorm and identify potential datasets in each data source
    - Step 3: Complete dataset inventory template
  - Example: Rent Board
    - Step 1: Identify data sources
    - Step 2: Brainstorm and identify potential datasets in each data source
    - Step 3: Complete dataset inventory template
- 4. How to Prioritize Datasets for Publication
  - Step 1: Access your completed inventory
  - Step 2: Review the initial priority groups
  - Step 3: If needed, revise the priority using column C
    - Instructions for spreadsheet:
    - Optionally, add missing datasets
  - Step 4: Send us your updated priorities
  - Next steps
- 5. Create Publishing Plans
  - 1. Getting Started
    - Answer some quick questions
    - Sign up for a DataSF Plan Submission Portal account
  - 2. Crafting your narrative
    - Write Maintenance Narrative (skip if you still have datasets to publish)
    - Write Publishing Narrative
    - Write Planning Narrative
  - 3. Submitting your plan
    - Go to https://datasf.knackhq.com/plans
    - Add your plan
    - Edit your plan (optional)
- 6. Next Steps
- 7. Appendices
  - Appendix A. Data Catalog Templates
    - Data Catalog: Fields and Descriptions
  - Appendix B. Definitions

Appendix C. Resources & Credits

1. Background

Purpose of this Guidebook and Version Notes

The purpose of this guidebook is to provide guidance to Data Coordinators in the City and County of San Francisco. Data Coordinators should use this guidebook to help them in their new role. We'll update this guide as the role and responsibilities of the Data Coordinators evolve and as we learn more.

Date	Version	Description of changes
June 19, 2014	1.0	Creation of guidebook including: • Background and motivation for Open Data in San Francisco • Roles and responsibilities • Guidance on creating a data catalog (inventory)
February 4, 2015	1.1	Added link to detailed inventory guide
May 28, 2015	2.0	Added step 4: prioritization guidance to guidebook and links to other resources for data inventory
Aug 4, 2015	3.0	Added step 5 on creating publishing plans to guidebook
May 31, 2016	3.1	Small additions to incorporate data quality
Jan 10, 2018	3.2	Updated publishing plan guidance to reflect submission changes

Why release open data?

One of the first questions departments often ask when learning about open data is "Why should my department release data?" There are a number of reasons, both practical and philosophical, why releasing data can benefit your department and the people it serves.

Stimulate new ideas and services. By releasing open data, city departments may help to stimulate new and innovative ideas from our local technology community. There is great potential for open data to act as the fuel for new solutions and even new businesses that can address common problems or challenges facing those that live in, work in or travel to the City. View applications that have been built using the City's data.

Increase internal data sharing. Open data can also help with some of our internal challenges accessing data between departments. Right now, analysts often rely on personal relationships to access data from other departments. DataSF can provide a platform to share data internally. Combining information from different departments could provide valuable insights into how our city works and how departments may better serve those that live and work in San Francisco.

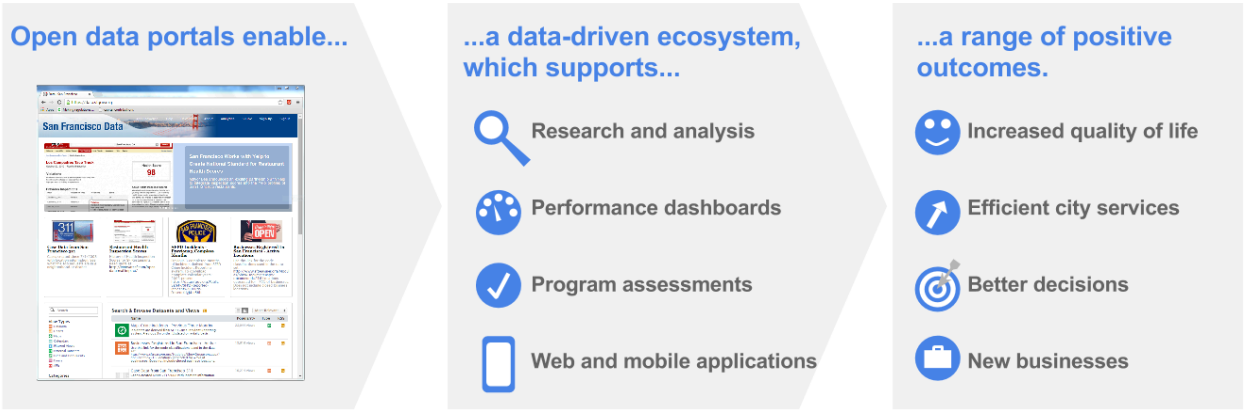
Simplify Sunshine Requests. Open data releases can be an effective way of responding to requests for data through the City's Sunshine Ordinance. One open data release may address multiple requests for information than can be repetitive and costly to respond to if addressed on an individual basis.

Improve Data Quality. The Open Data process allows/encourages users to provide feedback on

accuracy, consistency and other quality measures, important feedback that can help departments get better results from their own internal data uses.

Reduce unwanted web traffic. Publishing open data can also help reduce unwanted web traffic on department websites, which is often the result of "data scraping" by individuals seeking to obtain data in bulk from the City. This puts unnecessary stress on the city's technology infrastructure and unneeded burden on city IT staff.

Changing how we use data. Ultimately, open data can serve as a platform to change how we use, share and consume our data externally and internally; transform data into services, and foster continuous improvement in decision-making and the business of government. Ultimately, open data is about enabling use of data to help support a range of positive outcomes:



History of Open Data in San Francisco

Date	Event
August 2009	Launch of DataSF, our open data portal
October 21, 2009	Mayor Newsom releases Executive Directive 09-06 (The Open Data Directive)
June 21, 2010	Responsibility for Administration of the DataSF site is transferred to 311 Customer Service Center
October 28, 2010	Creation of Administrative Code Chapter 22D: Open Data Policy[1]
March 28, 2013	Ordinance No. 69-13, amending Administrative Code Chapter 22D, establishing position and duties of the Chief Data Officer and Departmental Data Coordinators
November 21, 2013	Ordinance No. 285-13, amending Administrative Code Chapter 22D, establishing timelines and procedures for releasing open data

DataSF: Our Open Data Portal

DataSF (http://data.sfgov.org) is the City and County of San Francisco's open data portal. DataSF is hosted by a vendor, Socrata. DataSF allows users to find, visualize and use our data, whether

[1] Available under the Administrative Code, linked to at http://www.amlegal.com/library/ca/sfrancisco.shtml.

developing novel applications or combining the data across multiple agencies to support new services.

2. Roles and Responsibilities

Below we included an overview of the roles and general responsibilities in support of open data. The Open Data policy in Chapter 22D (available here under the Administrative Code section) provides greater detail on both the Data Coordinators and the Chief Data Officer. We expect to modify or change these as we learn more and as the roles become more refined.

Role	**General Responsibilities**
Data Coordinators	Data Coordinators are designated for each department as the main point of contact and accountability for open data in their department. General responsibilities include: • Inventorying department data sets • Establishing a plan and timeline for publishing them • Serving as a key point of accountability for timelines and questions about data sets • Implementing privacy, data licensing, metadata and other standards and practices • Providing quarterly reports on progress in implementing the open data plan

Chief Data Officer	The Chief Data Officer is designated by the Mayor and is accountable for the city's overall implementation of the open data policy. General responsibilities include:
	• Creating processes, rules and standards to implement the open data policy, including but not limited to:
	  ○ Prioritizing data sets for publication
	  ○ Determining what datasets are appropriate for public disclosure
	  ○ Creating data licensing and metadata standards and guidelines
	  ○ Providing guidance and assistance to City departments in releasing open data
	  ○ Providing guidance and assistance to City departments in assessing and, where appropriate, improving the accuracy, completeness, interoperability and other quality dimensions of data
	  ○ Facilitating creation of department implementation plans and reporting
	• Maintaining the open data website
	• Presenting an annual citywide implementation plan for open data
	• Assisting departments with analysis of city datasets

Data Stewards	Data Stewards are individuals in charge of individual databases, datasets, or information systems. In general, a data steward has business knowledge of the data and can answer questions about the data itself. General

	responsibilities likely include: • Managing the dataset or source and authorizing changes to it • Managing access to and use of the data, including documentation • Managing accuracy, quality and completeness of the data
Data Custodian	Data Custodians are individuals that assist with the technical implementation of individual databases, datasets, or information systems. Not all systems or data sources will have a data custodian. General responsibilities likely include: • Implementing technical changes requested by the data steward • Administration and maintenance for the database or system

3. How to Create Your Data Catalog (Inventory)

The first major task for the Data Coordinators is to create a data catalog (or inventory) of your department's data. Below we provided guidance on how to do this. But we recognize that this will be in part a learning process. As a result, we will make changes to this guidance based on your feedback. After the initial catalog creation, this process may continue as new datasets are created, discovered or requested by other parties.

Follow the 3 major steps below to conduct your data inventory:

1. Identify data sources
2. Brainstorm and identify datasets in each data source
3. Complete dataset inventory template (for each dataset)

We also included an example from the Rent Board that you can use for guidance. Appendix A. includes templates to support this process.

Also, visit these resources for additional guidance:

- Data Coordinators Portal
- Which includes the "DataSF Guidebook: Detailed Inventory Guide Steps 2 & 3"

Step 1: Identify data sources

Your data may be housed in a variety of places - from inside information systems or databases to shared drives and folders. This can also include 3rd party vendors and data hosted on vendor systems. Step 1 is about identifying the major data sources in your department.

Questions to help identify and discover data sources:

- What information systems does your department use?
- What databases does your department use?
- What applications capture information or are used in your business processes?
- Are some data resources kept in spreadsheets (on shared or individual drives)?
- What information are we already publishing and where did that information come from?

For each of the data sources:

- Provide a name and brief description of the data sources
- Capture any technical details and point of contacts

Appendix A. includes a templates for this process.

Steps 2 and 3 Guidance

Below we describe at a high level the next 2 steps of the inventory. Visit the Detailed Inventory Guide for Steps 2 and 3 for more detail.

Step 2: Brainstorm and identify potential datasets in each data source

Some of your information sources may be fairly straightforward (e.g. a single sheet in a spreadsheet). In these cases, you have already identified the dataset.

In addition, you may already have a list of datasets you are publishing or plan to publish.

But others, like relational databases, may be very complex. Identifying subsets of the database that could serve as datasets, probably requires some brainstorming. You may want to include your PIO, data stewards and lead analysts in this process.

To help brainstorm, use the questions below:

- What data populates your monthly or quarterly reports?
- What departmental data is currently publicly available on DataSF or elsewhere online?
- What data does your department use for internal performance and trend analysis?
- What information is published as a performance metric?
- What data is reported to federal, state or local agencies?
- Talk with your Public Information Officer (PIO) - what data has been requested under Sunshine?
- What data do other departments ask for?
- What kinds of open data are similar agencies across the country publishing?

Caution: Don't exclude any datasets based on privacy or confidentiality concerns! Our goal is to have a holistic picture of our data. Based on this big picture, we can then decide what we should publish. Step 3 provides a means to capture privacy and confidentiality concerns.

Step 3: Complete dataset inventory template

For each dataset you identify in Step 2, complete the inventory template. Include:

- New datasets (identified via brainstorming)
- Existing datasets, including already published datasets

Appendix A. includes the templates and description of template fields.

Example: Rent Board

Below we provide an example of conducting an inventory for the Rent Board.

Step 1: Identify data sources

#	Name	Description	Technical details	Key POCs
1	“Filemaker”	The vast majority of the Rent Board’s data is kept in a single relational database, referred to as “Filemaker”. The database consists of 5 main types of information: 1. Petitions 2. Appeals 3. Eviction Notices 4. Address database 5. Capital improvements	Filemaker x.y	Maintained by x Business POC is y

Step 2: Brainstorm and identify potential datasets in each data source

Given the 5 types of information in Filemaker, the following datasets could be extracted from the database:

- Petition records. Petitions can be filed by landlords or tenants. This database includes information such as:
    - Property details
    - Type of petition (the types vary based on whether or not it is a landlord or tenant)
    - Various details related to processing and hearings
- Appeal records. Both landlords and tenants can appeal decisions from hearings and mediations. This database includes information such as:
    - Property details
    - Type of appeal
    - Various details related to processing and hearings
- Eviction notices. Landlords are required to file copies of eviction notices with the Rent Board. This database includes information such as:
    - Property details
    - Type of eviction
- Capital improvement records. Landlords must include costs of capital improvements when they want to pass the costs onto tenants. This database provides a history of capital improvements included in petitions, including:
    - Property details
    - Cost for improvements and rent increases

The address database is probably not a useful dataset to share as it is provided by other departments that do the actual maintenance on the dataset.

Step 3: Complete dataset inventory template

See a completed inventory for one dataset from the Rent Board (link to google spreadsheet).

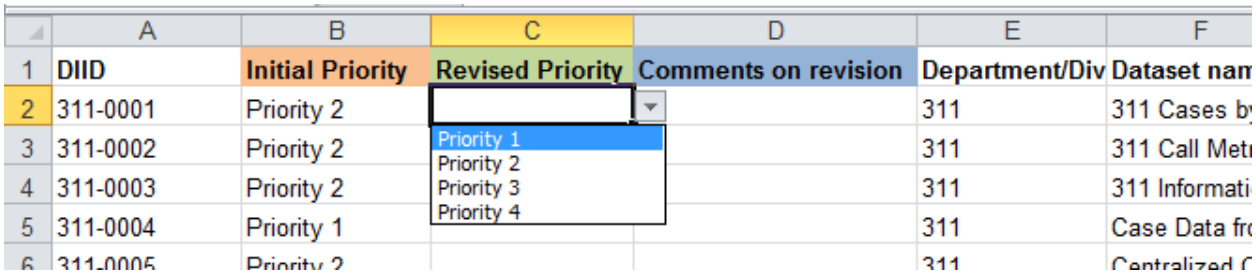
4. How to Prioritize Datasets for Publication

Now that you've completed your dataset inventory, you are ready to prioritize your datasets for publication and then create publication plans (coming soon). The good news is - you've already done the basic prioritization work already! Follow the steps below to complete the prioritization.

Step 1: Access your completed inventory

Your prioritization spreadsheet should have been emailed to you. If you can't find it, please email jason.lally@sfgov.org and he'll get you your document.

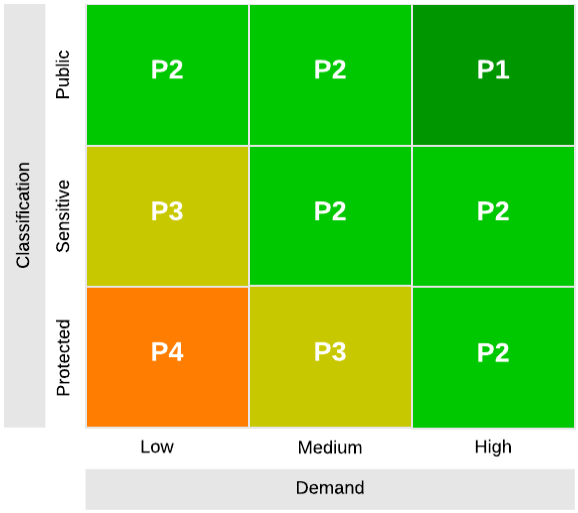
Once you open your spreadsheet, notice that we've added three columns - B, C, & D - in the front (all the rest of your inventory is also included for reference):



Additionally, there is a field **Posted on DataSF Now?** in column H. Some departments indicated this, others did not. Please review to make sure it is correct. This field helps us assure you get credit for already published datasets.

Step 2: Review the initial priority groups

We created the Initial Priority column (B) based on two questions you answered during the dataset inventory: value (Low, Medium, High) and classification (protected, sensitive, public). The chart below describes the 4 priority groups:



Priority 1

Public datasets that are in high demand should be first on the list for publication unless some of the other considerations below apply. Even if the considerations below apply, these datasets should not fall below Priority 2.

Priority 2

These datasets range from high to low demand and may have to be thoughtfully published to protect private and sensitive information. The order in which you publish this data may be dependent on the other considerations below.

Priority 3

Avoid scheduling these datasets at this point. Once we have scheduled Priority 1 and 2 data, we will revisit this list. **Exception:** if the data is historical (or won't require updating) and easy to publish, consider adding it to your Priority 2 group.

Priority 4

These datasets do not need to be scheduled until the rest of the datasets are scheduled in the publication plan. In some cases, we may not choose to publish these datasets or may only make them available internally.

Step 3: If needed, revise the priority using column C

We recognize that other factors may affect your ability to prioritize the publication of a dataset. The table below lists some considerations that may influence your decision to move an individual dataset up or down the priority list. Feel free to consult with the Data Steward and technical point of contact if you need help with this step. Please make sure before you do this, you've checked whether the dataset is currently being actively published on the open data portal and indicating that in column H. Also note that if Initial Priority (B) is Unassigned, that means that either a Data Classification or Demand value was missing in your original submission. That's okay, as you can set the priority now based on that assessment.

Instructions for spreadsheet:

- To change priority group:
  - Use the drop down menu in column C to update the priority group for the dataset
  - Provide a brief note on why you think it should be lower/higher
- To keep the initial priority group: do nothing!

Factors that increase priority	Factors that decrease priority
• Existing reporting requirements that are inefficient and resource intensive • A clearly defined use case emerges for publishing the data on DataSF • It's easy to publish and doesn't require updating • The data will inform a high-profile issue or concern but people may not be aware of it • Existing reporting already automates the data extractions making it easy to add to DataSF • During the publication of a related dataset, it becomes trivial to publish it as well	• An upcoming migration to a new backend system (this would create additional work by automating the publication twice) • You have major data quality concerns such that the data is not usable • The data is not available in a structured manner (e.g. it's not in a database or well-designed spreadsheet)

Optionally, add missing datasets

If you notice that you're missing a dataset, you can add that to the bottom of your list. Do not assign an ID number as we'll do that when we incorporate into the master inventories. Please make sure to include the other fields and a prioritization for the new dataset.

Step 4: Send us your updated priorities

After you reviewed your priority levels, send to jason.lally@sfgov.org

Next steps

Over the next few months, we'll be asking you to propose broad (not specific) time frames for publishing your datasets. You'll use this prioritization to help allocate your resources and set goals (e.g. our department will publish x priority 1 datasets over the next x months).

And don't worry - if you find you need to revise the priority level in the future, we'll create a method and time frame for doing so. We understand that technology and plans change!

5. Create Publishing Plans

The nature of publishing datasets can vary by department, so we're providing some simple guidance to help focus and standardize department publishing plans and give you a simple tool for tracking and communicating progress.

The process is pretty straightforward and we've provided guidance below. The table gives you a quick glance of how much time we expect each step to take. Most of the time will be spent in writing your narrative paragraphs. We suggest sharing this task with others who can help you relate the open data plan to the strategic goals and priorities of the Department.

Task	**Brief description**	**Estimated time**
Get started	Figure out what you need to prep and write, sign up for a user account so you can submit your complete narrative(s) later.	10 minutes
Write your narrative	Write your narrative (2 to 6 paragraphs). We suggest involving your colleagues as necessary and having your draft go through internal review.	30 to 120 minutes (with some time for review cycles)
Submit your plan	Submit your narratives online at the DataSF Plan Submission Portal	10 minutes

1. Get Started

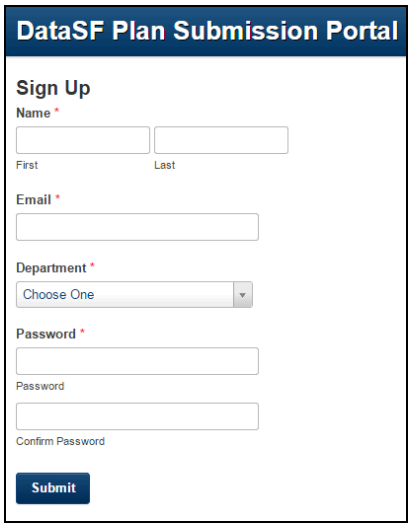
To prepare for your plan, you will be producing a brief narrative with this frame in mind:

> "Over the next year, what will my department commit to regarding open data?"

***Note:** plans are now written annually each calendar year, please consider your narratives and targets over a year long period.

Before starting, scan this guidance to get a sense of what you'll need to do and set aside the appropriate amount of time. This should not take a large amount of time.

Sign up for a DataSF Plan Submission Portal account



Later on you'll be submitting your plan narrative online. You'll need a user account that will be approved by DataSF staff.

Please take 5 minutes to register online so you can submit your plan later on. *Please note, this account has no relationship to your open data portal account. Unfortunately at this time, we cannot integrate the logins, but we're looking into options.*

Go to: https://datasf.knack.com/plans and click on **Sign Up** at the bottom of the page. Please use your work email address and choose your department from the dropdown.

If you've already done this, you can simply log in with the account you've created. If you cannot remember your password, you can reset it at https://datasf.knack.com/plans by clicking (forgot?).

2. Write your narrative

The submission will require you to identify which phase you are in with respect to open data. The phases are:

1. **Planning.** If you do not anticipate publishing any datasets in the calendar year plan cycle, then you should still be engaging in planning activities. Your narrative should reflect what you'll be doing to get ready for publishing in the future.
2. **Publishing.** If you are publishing data in the calendar year, your narrative should focus on the nature and general categories of data you plan on targeting this year.
3. **Maintenance.** If you have completed publishing all data in your inventory, your narrative should simply acknowledge that.

You will write a narrative that reflects which phase you are in above. **Please use the guidance below as you write.**

If you or your department has already written plans in the past, they'll be available at https://datasf.org/plans. You can also copy prior plans within the plan submission portal as a starting point within the publishing portal if you desire (See Copying Prior Plans)

The narratives should be concise and informative. Keeping the frame in mind - what will the department do in the next year - you will write just a handful of paragraphs.

During drafting, you can write the narrative wherever you like, but you will be logging in later and copying and pasting your work into a form submission.

Please **keep the formatting simple and focus on content** while drafting. The submissions will be standardized, and you will be able to ***apply limited formats on submission***.

Also, when writing your overall narrative, keep in mind things like:

- **Departmental mission.** Thread in how your publishing plan is related to the overall mission.
- **Departmental service areas.** What are the key activities of your department, how does your data publishing plan relate to some or all of these areas.
- **Your audience**. Assume a naive audience and write as if someone is reading your plan independently of the broader open data program. We will stitch this together and provide broader context, but it's good for these narratives to stand on their own as well.

Planning Narrative Guidance

Planning phase is designed to make sure you are queuing up important planning activities sooner rather than later for datasets down the road. You should **identify the datasets you'll be assessing and investigating for publishing later.**

It's important to keep an eye toward the future so you keep momentum on publishing. For example, there are departments with tricky datasets that are sensitive or protected. What should you be doing now to begin having the necessary conversations to publish variations of those datasets? Or maybe there's a complicated system migration that you're keeping an eye on. At its simplest you should use the planning narrative to point out what's "on deck" or next up.

You will write a paragraph or two that do the following:

- **Create a basic "frame" for your planning activity.** For example, by program area, by priority category, by data classification, by division, by major system, or a mix. This is important internally as well as you can align your main publishing activity strategically within your department. This should also include the overall rationale.
- **Call attention to at least one focus area.** It may be too early to call out a dataset, but you should be able to zero in on an area where the department will be focusing attention. For example, *"the department is undergoing a major system upgrade on _____ that should be wrapping up at the end of the year, we will monitor that and be planning for publishing datasets out of the new system when it is operational."*

Publishing Narrative Guidance

If you are in publishing phase, then you should be **identifying datasets that can be published over the next year**. You do not have to name all of the datasets in your narrative, but you must **identify a target number of datasets** that you will submit with your narrative.

You will write a paragraph or two that do the following:

- **Create a basic "frame" for your publishing activity.** For example, by program area, by priority category, by division, by major system, or a mix. This is important internally as well as you can align your main publishing activity strategically within your department.
- **Call out at least one specific example dataset**. While calling out all datasets may be impractical for operational reasons, you should be able to call out at least one or two example datasets that you are targeting (more are welcome). You should be sure that these are datasets you will be able to complete in the six month period.
- **Say something about accomplishments to date.** If you have already published data, say something about what you've accomplished broadly. For example, publishing data on

_____ has led to _____ or helped us _____.

Maintenance Narrative Guidance

If you are in maintenance, then you just have to write up one simple paragraph. **Being in maintenance means you've completed the publishing of your inventoried datasets**, but that doesn't mean the department's work is done (just maybe a little bit easier).

In maintenance, there isn't as much to actively do, however, you may want to be creative or take the opportunity to highlight new initiatives or celebrate successes. Here are several ongoing activities that departments can do during maintenance phase:

- Continuing to monitor incoming feedback and questions on datasets
- Improving published datasets based on feedback
- Monitoring for missing or new datasets to add to the dataset inventory
- Monitoring data quality and planning for improvement

Here's a simple example. We recommend using this as inspiration, but please feel free to break the mold. We encourage you to both reflect on past accomplishments and future aspirations.

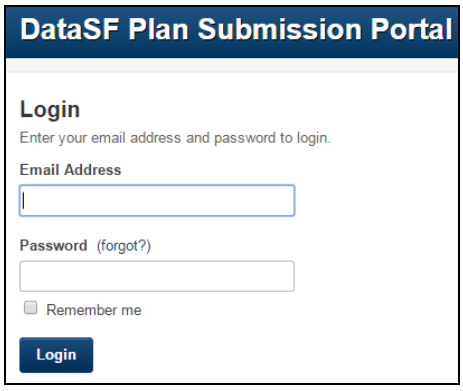
> "The Department has published all of its inventoried data as of _____. We will continue to monitor our dataset use and any feedback so we can proactively improve them. We have seen great interest in our dataset on _____. Since publishing that dataset, there has been _____. We look forward to _____."

3. Submit your plan

Submitting your plan is quick and can be done online.

Go to https://datasf.knack.com/plans

1. Sign in using the user account you created during Getting Started.

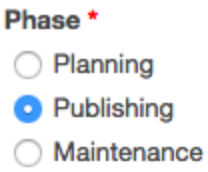


Add your plan

1. You will see a screen with an empty table and an "Add Plan" button at the top
2. Click on "Add Plan" (you can also start by copying prior plans)



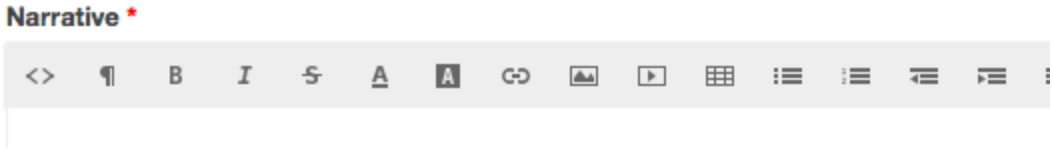
3. Fill out each option in the form (reference images inline)
    a. *Phase* (checkboxes, required, choose 1)
        i. Planning
        ii. Publishing
        iii. Maintenance



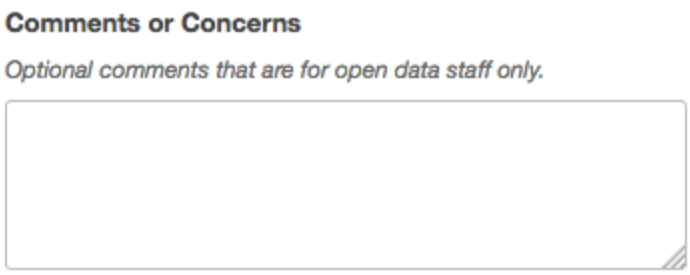
    b. *Targeted Number of Datasets* (number, required if in Publishing Phase)
        i. While you do not have to name all targeted datasets, we ask for a target number



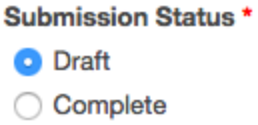
    c. *Narrative* (rich text, required)
        i. Copy your narrative paragraphs here



    d. *Comments or Concerns* (text, optional)
        i. If you have any concerns you want to highlight, these are for internal planning and feedback

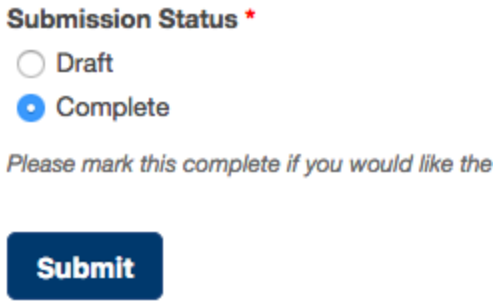


e. *Submission Status* (radio button, default: Draft)
    i. If you aren't ready to submit, you can keep your submission in Draft and come back later (see: editing an existing plan below)



Submitting your plan for review

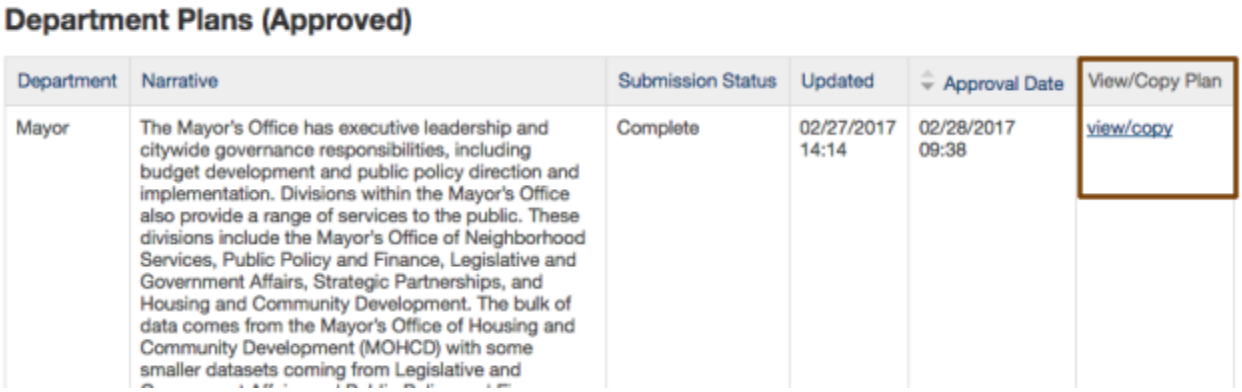
Choose **Complete** on the form and click **Submit** to notify open data staff that your plan is ready for review. If there are no suggested changes, the plan will be approved. Either way, **you'll hear from us with any next steps.**



Copying prior plans (optional)

If you'd like to start a submission from a prior narrative, you can:

1. Find a prior plan record in the table of approved plans
2. Click view/copy
3. Scroll to the bottom
4. Click Submit under Copy Plan
5. The plan

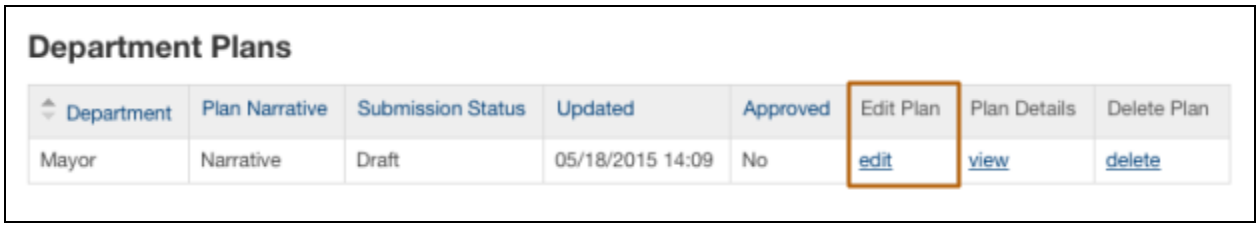


Department Plans (Approved)

Department	Narrative	Submission Status	Updated	Approval Date	View/Copy Plan
Mayor	The Mayor's Office has executive leadership and citywide governance responsibilities, including budget development and public policy direction and implementation. Divisions within the Mayor's Office also provide a range of services to the public. These divisions include the Mayor's Office of Neighborhood Services, Public Policy and Finance, Legislative and Government Affairs, Strategic Partnerships, and Housing and Community Development. The bulk of data comes from the Mayor's Office of Housing and Community Development (MOHCD) with some smaller datasets coming from Legislative and	Complete	02/27/2017 14:14	02/28/2017 09:38	view/copy

Editing your plan (optional)

If you need to change your plan. You can login and click Edit Plan next to your submission in the table. You'll be presented with the same form as above, but now with your entries in place. Edit and click submit.



Department Plans

Department	Plan Narrative	Submission Status	Updated	Approved	Edit Plan	Plan Details	Delete Plan
Mayor	Narrative	Draft	05/18/2015 14:09	No	edit	view	delete

6. Next Steps

1. **Start publishing!** We'll be providing more guidance on how we'll approach the publication process and what resources and support will be provided in publishing your department's datasets, including required metadata elements, data reviews, etc.
2. **Adding value.** We'll be identifying ways that you can increase the value of your published data for your departments - whether it is creating visuals or maps that you can build and include on your website, improving data quality (e.g., accuracy, interoperability or completeness) or identifying opportunities for cross department collaborations.

7. Appendices

Appendix A. Data Catalog Templates

The data catalog includes drop down menus and other data tips to make it as easy as possible to complete your inventory. Template Options:

- Google: If you have a google account and want to use google, use the Google template (copy the template into your own account - do not edit the document)
- Excel: Download excel template

Data Catalog: Fields and Descriptions

The data catalog templates include the following fields. The spreadsheet version includes data validation, including controlled vocabulary and other data checks.

Field	**Description**
Department/Division	What department collects and manages the data as an official record?
Dataset name	Brief descriptive name for the dataset
Brief description of data	Include a brief description of the dataset. What is the purpose? What is it used for? Include key data fields if possible.
Data Steward (Business knowledge) - Name	Who manages the data and is responsible for making changes to the data? Who understands what the dataset includes and can answer questions about it?
Data Steward (Business knowledge) - Email	Please enter the email of the Data Steward.
Data Custodian (Technical knowledge) - Name	If applicable, include a contact who manages the technical execution of the database (e.g. database management, access and extraction).
Data Custodian (Technical knowledge) - Email	Please enter the email of the Data Custodian.
Data source	What information system or database contains the data? Or what shared server or shared drive contains the data?
Start date	How far back does the data go? Use format (MM/DD/YYYY)
End date	When does the data end? Use format (MM/DD/YYYY). If the data is still being updated, use "Current".
Geographic coverage	Does this data cover the whole city or a subset? Or does it have broader coverage than the city?
Geographic granularity	What is the lowest level of geography in the data? For example, if the data is collected by address, it would be Street Address.
Frequency of data change	At what rate does the information in the dataset change?
Number of records	How many records or entries does the dataset include?
Format	What format is the data in? e.g. excel, sql, oracle database, pdf, word, etc.

Existing ETLs	Are there existing database connections or extractions?
Existing publication	Is this data already published (made publicly available) in some form or another, e.g. a report.
Link to existing publication	If you answered yes to the previous question, please include a link to the report or document.
Priority/value	What is your sense of the relative value in publishing this data? • High - Existing and ongoing requests for this data; this data addresses pressing information needs or pain points (within or without the city); or we have heard compelling examples of how this data could be used • Medium - This data may be useful for other departments or for people external to the city; we occasionally receive requests for this information; or we have heard some examples for how this data could be used • Low - This data has unclear value for either the public or other city departments; we have never received requests for this data; or we have never heard a use case for this data
Priority/value comments	Any details or comments to add to your response on priority?
Technical challenges	What technical challenges, if any, do you anticipate in publishing this data?
Data Classification	How would you classify this data? • Public - this data could be publicly disseminated without any concerns • Sensitive - in its raw form, this data poses security concerns, could be misused to target individuals or poses other concerns • Protected - this data is protected by law or regulation and can only be shared or accessed internally and per organizational procedures; OR this information includes individually identified information
Protected - Details	If you marked “Protected” for Data Classification, please indicate what law(s)/regulation(s) protect this data.
Sensitive - Details	If you marked “Sensitive” for Data Classification, please describe your concerns
Data Quality	Do you have concerns about the quality of this dataset?
Data Quality - Details	Please describe your concerns with the quality of this dataset.

Appendix B. Definitions

Term	**Definition**
Dataset	Contents of a single database table, worksheet or defined view; data is provided as a single combination of unique rows (or records) and corresponding columns (or fields) describing that row Example - Database: A database may contain several data tables - each data table constitutes a dataset. However, you could also create new datasets by combining data from different tables into a new table.

Data Schema or Standard	Specification that defines the structure of the data (i.e. required data elements and types and supporting definitions)
Data source	Technology or system that stores data, including databases, named spreadsheets, information systems, business applications, etc.
ETL	Extract, Transform, Load - three database functions that are combined into one tool to pull data out of one database and place it into another. • Extract - process of reading data from a database. • Transform - process of converting the extracted data from its previous form into the form it needs to be in so that it can be placed into another database. (rules, lookup tables, combining with other data) • Load - process of writing the data into the target database.
Geospatial data	Data related to the position of things in the real world, including boundaries or locations
Metadata	Descriptive information about a dataset
Tabular	Data that is presented in columns or tables
Taxonomy or Category	Methodology by which items or datasets are classified or grouped under a similar theme or topic

Appendix C. Resources & Credits

This guidebook was created using input from a number of resources, including:

Title	**Attribution**	**License**
City of Philadelphia Open Data Guidebook	City of Philadelphia, Office of Innovation and Technology	Creative Commons Attribution-ShareAlike 4.0 International license
New York State Open Data Handbook	New York State Open Data Initiative	
Open Data Handbook	Open Knowledge Foundation	
Sunlight Foundation Open Data Guidelines	Sunlight Foundation	