Data Governance Plan Template - Open Data Components

<<City>>, <<State>>

<<Month>> <<Year>>

### Tracking Progress (Metrics & Performance Measures)

#### Model Language

It is important for <<ORGANIZATION>> to periodically measure the effectiveness and efficiency of its data governance and open data program. By tracking the following measures, <<ORGANIZATION>> will identify when changes to the program’s operation can be necessary. At present, <<ORGANIZATION>> performance measures related to <<DESCRIPTION OF MEASURES>>.

|  |  |  |
| --- | --- | --- |
| **Measure** | **Target/Trend** | **Frequency** |
|  |  |  |
|  |  |  |

#### Questions to Address

This section pertains to cities and organizations with open data programs. Successful open data programs will develop and track performance measures that provide it a clear understanding of how well its program is doing. Use this portion of the open data governance document should answer the following questions:

* How do you know if your open data program is a success?
* What performance measures will you monitor?
* How often will you review these performance measures?

You may consider using the following measures to track the progress of your open data program.

Data metrics are directly related to the data assets that you have and publish, encompassing the structure and quality of your data. The following metrics describe a city’s plan for and progress on inventorying, prioritizing, publishing, and updating datasets.

|  |  |  |
| --- | --- | --- |
| **Measure(s)** | **Desired Trend / Target** | **Frequency** |
| Percent/# of departments that have completed the dataset inventory | 100% | Monthly, Quarterly |
| Percent/# of departments that have completed publishing plans | 100% | Monthly, Quarterly |
| Percent of inventoried datasets that are published | Increase over time | Weekly, Monthly |
| Percent of inventoried datasets published by dataset classification | 100% for public data | Weekly, Monthly |
| Percent of eligible datasets that are natively hosted | Increase over time / 90% | Weekly, Monthly |
| Percent of datasets updated on time | 100% | Daily |
| Percent of eligible datasets without a publishing lag | 100% of eligible | Daily |
| Median publishing lag in days for datasets with a publishing lag | Less than 30 | Weekly, Monthly |
| Percent of datasets with required metadata | 100% | Weekly, Monthly |

Use metrics describe how your open data is handled internally in your city, as well as externally within the community. Your city can conduct surveys to measure users’ perceived value of open data, as well as measure how often people are using certain datasets. This information can provide you with potential success stories and insight into what data the public is interested in.

|  |  |  |  |
| --- | --- | --- | --- |
| **Measure(s)** | **Desired Trend / Target** | **Frequency Collected** | **Internal / External** |
| **Internal changes in behavior / circumstances** | | | |
| Respondents indicating that open data has made their analytical work:   * easier * faster * more accurate | Increase over time | Annual | Internal |
| Percent of respondents indicating that they use the data portal to access data internally and from other departments | Increase over time | Annual | Internal |
| Respondents indicating that the following are medium or major barriers   * Knowledge/awareness of datasets (inter and intra departmental) * Getting data out of information systems (inter and intra departmental) * Not knowing who to contact * No standard process * Data not available in digital form | Reduction over time | Annual | Internal |
| Respondents indicating that the following would be helpful or very helpful   * Better access to data produced or collected by other departments | Reduction over time | Annual | Internal |
| **Products made with open data** | | | |
| Number of products made with open data (by type) | Increase over time | Ongoing | Both |
| Respondents indicating that open data was essential or important for creating their product | >50% | Ongoing | Both |
| **Indicators of use of open data** | | | |
| Creation of derived views (not by core team) | Increase over time | At least daily | Both |
| Percent of users that are return users | Maintain | At least daily | Both |
| Number of dataset downloads\* | Increase over time | At least daily | Both |
| Number of Data Portal accounts\* | Increase over time | At least daily | Both |
| Number of API hits (all sources)\* | Increase over time | At least daily | Both |
| Number of API token holders\* | Increase over time | At least daily | Both |

The following are some potential metrics to measure civic engagement, government effectiveness, and economic development. In addition, your city can conduct surveys to measure users’ perceived value of open data, as well as measure how often people are using certain datasets.

|  |  |  |
| --- | --- | --- |
| **Measure(s)** | **Desired Trend / Target** | **Frequency** |
| **Civic Engagement** | | Weekly, Monthly, Quarterly |
| Number of city responses to open data suggestions | Increase over time |
| Open data presence on social media | Increase over time |
| Number of projects made with open data (by type) | Increase over time |
| Number of participants at a community event related to data on open data portal (e.g., hackathon, strategic plan visioning, data release party, etc.) | Increase over time |
| **City Government Effectiveness** | |
| Number of ad hoc data requests | Decrease over time |
| Number of FOIA requests | Decrease over time |
| Respondents indicating satisfaction with city services | Increase over time |
| Number of internal or intra-departmental data requests | Decrease over time |
| Number of staff hours responding to data requests | Decrease over time |
| **Economic Development** | |
| Respondents indicated that open data was essential or important for creating their product | Increase over time |
| Economic value generated by open data in public/private sector | Increase over time |

### 

#### 

#### 

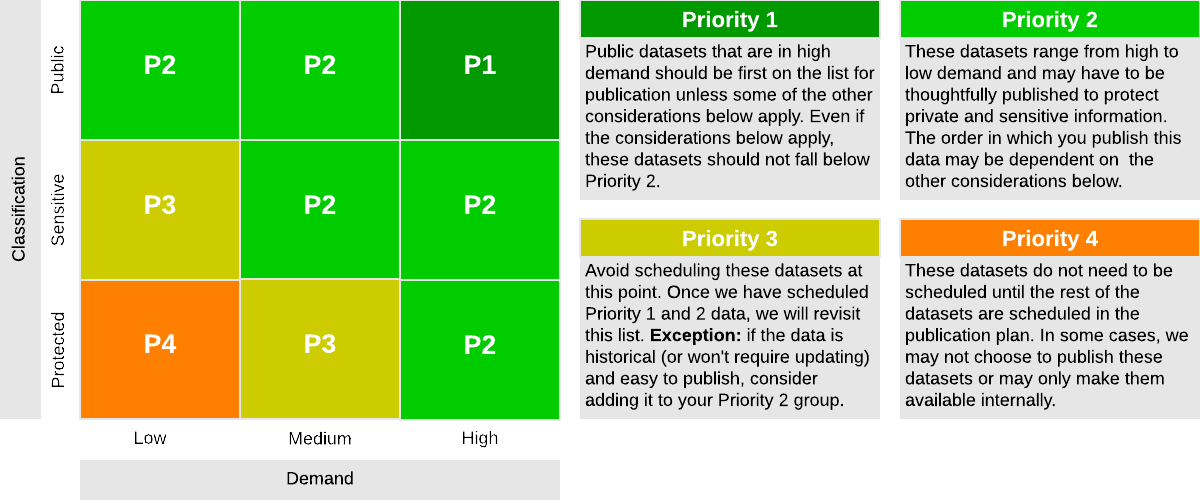
### Prioritizing Data for Release

#### Model Language

Following the inventory process,<<ORGANIZATION>>’s open data team will prioritize datasets for release to the public. These prioritizations will be based on the three dimensions:

* **Demand/Value**. Is there clear demand for the data from the public or from other departments? Is there an easily identifiable use case for the dataset? Does it clearly align with the the <<ORGANIZATION>>’s strategic priorities? Has data been to the subject of public information requests in the past?
* **Ease of Publication**. Can the data easily be uploaded to the portal? Is the data in a single, flat file, or is it fractured across multiple tables in a relational database? How much time will data coordinators need to add metadata to the dataset? Will a change in business processes be necessary to consistently refresh the dataset? Does the data contain information that could be considered protected, private, or sensitive? What level of effort will be necessary to anonymize these fields?
* **Classification based on sensitivity of data.** Does the dataset contain any personally identifiable information (PII)? Could it still be released once PII is redacted? Should the dataset not be released at all?

*NOTE: The City of San Francisco developed the most comprehensive way of prioritizing datasets based on the demand for the particular dataset and the classification of the data. You can view what the city created below. Accompanying the city’s methodology for releasing data typically is visualization that describes how data is prioritized for release.*



#### Questions to Address

Organizations prioritize data for release in a variety of ways. Some organizations focus on releasing data aligned with city-wide or mayoral priorities, while others focus on the needs of internal staff or residents. In this section, you will answer:

* How will you prioritize data for release to your open data portal?
* What methodology will you use
* How will the objectives for your open data program be translated in your prioritization process?

### Publishing Data

#### Model Language

Before the dataset is uploaded to the portal and released to the public, two reviews take place:

* A quality assessment (QA) to find and correct any errors in the data itself or in the metadata. The QA is conducted by either the program manager or the data coordinator.
* A review of sensitivity issues associated with potentially protected data in the dataset. The <<DEPARTMENT AND/OR INDIVIDUAL NAME>> conducts the first stage of the sensitivity review, identifying potential concerns as either minor, significant, or overwhelming.
  + *Minor concerns*, such as a column of names that can be easily removed, are communicated to the program manager or data coordinator for mitigation. The dataset is then re-reviewed by the <<DEPARTMENT AND/OR INDIVIDUAL NAME>>.
  + *Overwhelming concerns* immediately disqualify a dataset from being posted to open data. Examples include datasets that are protected by law or that would pose a security risk. Datasets identified as overwhelmingly concerning will not be published.
  + *Significant concerns* represent issues that are not easily resolved but do not immediately disqualify the dataset, such as the possibility that anonymized individuals in a dataset could be easily re-identified when the dataset is combined with another dataset on the portal. Significant sensitivity concerns require values-based judgement calls, which is the purpose of data governance board. Once <<DEPARTMENT AND/OR INDIVIDUAL NAME>> identifies such concerns, the dataset is forwarded to the <<DEPARTMENT AND/OR INDIVIDUAL NAME>> and <<DEPARTMENT AND/OR INDIVIDUAL NAME>>, respectively, for review.

Datasets that pass both the QA and the sensitivity review are made public on the portal.

Datasets will be uploaded to the Portal by either the Open Data Program Manager, or the Data Coordinator depending on the level of difficulty and technical expertise required to upload the dataset. The Open Data Program Manager and the Data Coordinator will work together to identify the best method, and will reevaluate as necessary to best meet the needs of the <<ORGANIZATION>>. The Data Governance Team will work with the Information Technology department to determine whether a direct connection can be made from the open data portal to the database holding the dataset, or whether the dataset should be uploaded directly to the portal as a spreadsheet.

#### Questions to Address

This section describes how data is published in an organization’s open data portal. This section should answer the following questions:

* Who is responsible for uploading the data on the organization’s open data portal? What are their responsibilities?
* What must be done to the data before it is published?

#### 

### 

### Data Quality Checks

#### Model Language

Each dataset is unique and requires a different process to review and assess quality (QA). However, certain commonalities apply across all datasets. <<ORGANIZATION>>’s open data program employs the following quality standards during the dataset QA process:

* The dataset is the most complete, accurate, and current version appropriate for public release.
* The data have been spot checked for common errors such as missing and misplaced values.
* Any missing data points are accounted for
* Columns are formatted appropriately.
* Metadata is complete, concise, and free of jargon.
* Metadata explain the process used to create the data and summarize any changes.
* Metadata clearly explain any limitations or omissions for each dataset.
* Metadata clearly identify an update frequency and plan.

The <<ORGANIZATION>>’s open data program is not, however, a data editing program. No dataset is perfect. Data coordinators will not be asked to substantially change or improve datasets pulled from city systems before uploading them to the portal.

#### Questions to Address

This section describes the quality standards for the open data program. The city should lists the minimum standards a dataset should meet before it is published on the organization’s open data portal.

### Additional Components

[Data Governance Template](https://docs.google.com/document/d/1cvkFwpvpBg9y4srrxpijNXrwzvsXTqtOgdF9CdRKb_E/edit#)

[Open Data Components](#_bbtrqo2e9cp9)

[Data Retention Template](https://docs.google.com/document/d/1Jy9xPDtbpz-k_6X-5q6zDmeqYX0KtbGFGP5eB0Y1Dnk/edit)  
Privacy & Security - Pending