**TEMPE OPEN DATA**

**DATA COORDINATORS HANDBOOK**

**Document History**

|  |  |  |
| --- | --- | --- |
| **Version** | **Date** | **Revisions** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# **1. Introduction**

The City of Tempe established the Tempe Open Data Program to support Tempe’s commitment to expand both the data it makes available to the public, and tools for understanding that data, and to foster an open and collaborative environment. As part of this program, each department is responsible for inventorying and publishing its own data, with support from the Open Data Team in IT. As part of the Open Data Policy, each department appoints one or more Data Coordinators to lead their department’s work in the Open Data Program and to participate in collaboration efforts with Tempe’s Open Data Working Group.

## The purpose of this handbook is to provide guidance about the roles of the Data Coordinators and to give an overview of responsibilities and procedures for inventorying, cataloging, publishing and maintaining a department’s data. Throughout this handbook, “you” refers to the Data Coordinators and “we” refers to the city’s Open Data Team.

## This handbook will continue to be updated as the roles and responsibilities of Data Coordinators evolve over time and the Open Data Program continues to develop. For questions about the Open Data Program or feedback about the handbook, please contact [data@tempe.gov](mailto:data@tempe.gov).

## **1.1 What Is Open Data?**

Open data is data that is free of charge and accessible for anyone to use, reuse, and redistribute. It is available online, in an open format, and is available for all to access and download in full without fees or requirements for creating accounts to access the data. Attribution and sharing may be required, but open data is otherwise free of restrictions and consists of machine-readable data and datasets.

## **1.2 Why Should We Release Open Data?**

There are many reasons why releasing data is beneficial to your department and the constituents you serve (both internal to the City in other departments and to the public):

**Increase civic engagement** – Open data allows your department to work with residents and solicit their ideas, input and creative energy through data. Open data can also improve public understanding of City operations, key strategic priorities and information about their community. It will also help foster both two-way communication and opportunities for collaboration between your department and Tempe’s residents.

**Increase internal data sharing**  – Open data can help break down silos and overcome challenges in accessing data between departments. The Open Data Catalog enables intra-departmental data sharing and reduces the amount of time needed to gain access to data in other departments. Combining information from different departments can increase efficiency, reduce duplication of data, provide new insights and spur innovative approaches to help address our city’s most pressing issues.

**Increase economic development** – Releasing open data may help stimulate new and innovative ideas from individuals and companies that benefit from the knowledge generated by open data. Sharing data encourages the local technology and startup communities to develop new solutions and tools that address the challenges or problems facing those that live, work or travel in Tempe.

**Data as an asset –** Data is a valuable asset, and open data can serve as the platform to change how we use, share and consume data externally and internally. Open data is about enabling our use of data to support better decision-making and provision of services to the community.

**Simplify Public Records Requests –** Open data can be an effective way of responding to requests for data. Open data release may address multiple requests for the same data, which can be repetitive and costly when responded to on an individual basis.

**1.3 Open Data Portal**

The City of Tempe maintains an Open Data Catalog on the City’s Open Data portal at <data.tempe.gov>. The catalog allows users to find and access City data for use in applications or analysis and to gain insights into City operations, programs or services.

**2. Roles and Responsibilities**

Below is an overview of the primary roles and general responsibilities in support of the City’s Open Data Program.

**Open Data Governance Structure**

**Open Data Governance Committee**

The Open Data Governance Committee is comprised of members of the Technology & Innovation Steering Committee. This committee oversees the implementation of the Open Data Program, including making policy decisions, prioritizing data publishing, resolving systemic issues, resolving conflicts in categorizations between public and protected, and communicating success to the public and the city.

Tempe is currently developing its approach to coordinated oversight and stewardship of its data. This handbook will be updated once the model and associated processes are developed.

**Open Data Team (ODT)**

The Open Data Team is comprised of city employees who administer the Open Data Program and Open Data Portal. They provide technical, planning, review and coordination support to City departments publishing open data.

**Open Data Working Group (ODW)**

The Open Data Working Group is comprised of individuals working on current open data projects in the City. Members include the Open Data Team and department data coordinators, as well as individuals who are involved only in work on the current dataset(s) (for example, programmers or subject matter experts).

**Open Data Key Positions**

**Open Data Manager**

The Open Data Manager, appointed by the Chief Technology Officer, is responsible for managing all aspects of the city’s Open Data Program. The program establishes and manages the Open Data Team.

**Department Leadership**

Department directors and managers guide their department’s participation in the Open Data Program, including

* Updating policies and procedures to reflect the Open Data Policy, designating data coordinators, and making sure staff have the time and resources to participate in the program.
* As needed, they work with data coordinators to set priorities, support work with stakeholders, and work through addressing any sensitive data in their datasets prior to publication.
* Consulting with data coordinators at the start of new projects and software implementation to ensure that they facilitate compliance with the City’s Open Data Policy.

**Department Data Coordinators**

Department Directors appoint Data Coordinators based on the department’s workgroups and data administration needs. Data Coordinators will work closely with departmental staff to inventory, prioritize, validate, publish, and maintain the department’s data.

Data Coordinators are members of the ODW and key to the success of the Open Data Program. Responsibilities include the following:

* **Developing Data Inventories**:
  + Catalog and maintain departmental data inventories
  + Update departmental data inventories as often as necessary, but at a minimum of twice yearly
* **Prioritize Data**:
  + Complete an initial prioritization worksheet for each dataset in the department’s inventory
  + Complete prioritization inventories for new datasets
* **Publishing and Maintaining Data**
  + Assist with access to, and final prioritization and validation of, datasets for publication as open data
  + Update (or oversee the update of) existing departmental open datasets according to the identified schedule or at least once per year;
  + Review data that is automatically loaded at least once per year or more frequently when appropriate;
  + Review metadata at least once a year or more frequently when determined appropriate.
* **Communicating and Collaborating**
  + Act as an open data advocate within their department in order to foster a culture that encourages open data and collaboration between departments
  + Participating in the Open Data Working Group and attending quarterly meetings.

**3. Open Data Management Process**

**3.1 Overview**

The Open Data Management Process focuses on inventorying, evaluating, prioritizing, publishing and maintaining city datasets. Each department will go through the following activities

**Data Identification and Inventory**

1. Identify department data sources
2. Identify datasets within each data source
3. Complete the dataset inventory for each dataset

**Evaluate and Review Data**

1. Evaluate and prioritize data for publication
2. Review security and privacy criteria

Inventory, Evaluate

and Review

The outcomes of the *Inventory, Evaluate and Review* activities are: 1) department data inventory, 2) prioritized list of datasets to be published, and 3) completed security and privacy evaluation

**Prepare for Publication**

1. Open Data Team Review
2. Prepare metadata and data dictionary
3. Enable automation or process for publication
4. Schedule data release
5. Publish data

**Maintain and Audit Data**

1. Maintain and audit published datasets

Prepare, Publish,

and Maintain

The outcomes of the *Prepare, Publish and Maintain* activities are 1) metadata and data dictionary 2) and automated publishing of data where possible 3) data published and maintained on the Open Data Portal.

A summary of both Inventory and Publishing phases, including roles and responsibilities, anticipated outcomes and associated templates can be found on the page that follows. More detailed step-by-step guidance can be found in the sections that follow.

The total time it will take to move through the data inventory to publication process will vary depending on several factors, including the following

**Factors that impact the time needed to complete the *Data Identification and Inventory* phase include**

* the size of the department
* the number of data sources/datasets; and
* how familiar the data coordinators are with the department’s existing data

**Factors that impact the time needed to complete the *Evaluate and Review Data* phase include**

* the amount of data clean up that needs to occur;
* agreement within the department on the prioritization of the dataset(s); and
* whether or not data needs legal review

Inventory, Evaluate

and Review

**Factors that impact the time needed to complete the P*repare for Publication* phase include**

* whether the data is loaded manually or through an API or programming;
* the amount of metadata that needs to be researched or created; and
* whether or not supplemental material or visualizations need to be completed

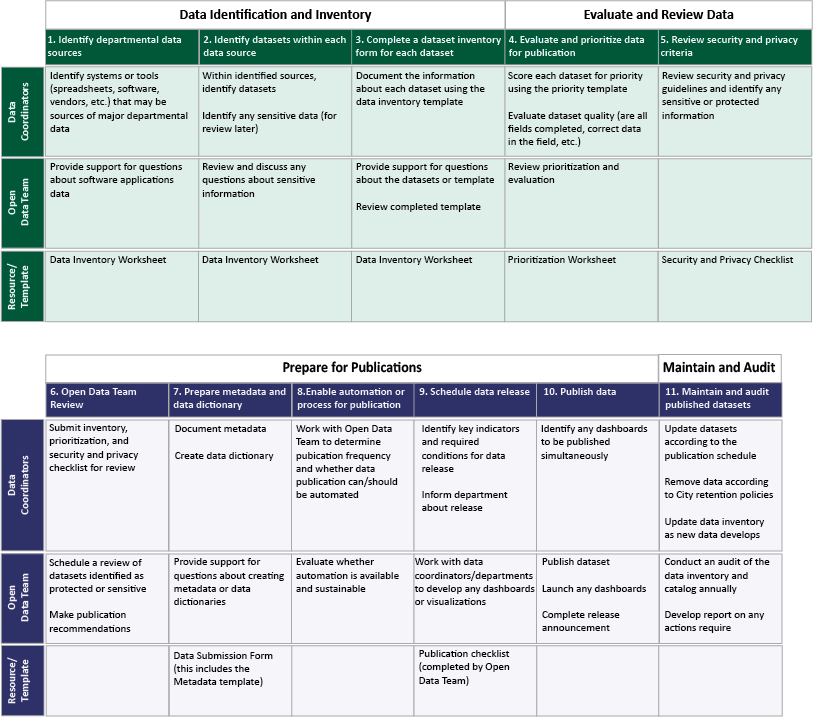
**Factors that impact the time needed for *Maintain and Audit Data* activities**

* the number of datasets;
* how often the data are updated; and
* whether changes are made to the data structure

Prepare, Publish,

and Maintain

Overall, it is expected that the process will take from 3-6 months for publication of a dataset.

****

**3.2. Data Identification and Inventory Activities**

The goal of these activities is to develop a comprehensive inventory of data sources and data in your department. Identifying a data source or dataset does not mean that it will end up published on the Open Data Portal or released to the public.

1. **Identify department data sources**

A data source is anything that contains your data. Your data may be housed in a variety of places including enterprise or departmental applications and databases; spreadsheets and databases on shared drives and folders; with third-party vendors that collect data on your department’s behalf as part of a program or service; or externally hosted systems, applications or services. Step 1 is about identifying the major data sources in your department.

Questions to help identify and discover data sources:

* What enterprise applications does your department use?
* What databases does your department use?
* What department specific applications capture information or are used in everyday processes?
* What programs or projects does your department administer? Where is that data stored?
* What spreadsheets on shared or local drive does your department use to store data?
* 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 source.
* Capture any technical details and points of contact.
* Use the included ***Data Source tab*** in the ***Data Inventory & Prioritization Template*** for this process

1. **Identify datasets within each data source**

Each data source you identified in Step 1 will have one or more datasets associated with it. Some data sources may be simple and straightforward and only have a single dataset (ex. a single Excel spreadsheet) while others may be very complex (ex. a relational database or a hosted application). Use the questions below to help you identify datasets for each data source:

* What data is used for your monthly, quarterly or annual reports?
* What data do City Council members, the City Manager’s Office, Boards or Commissions request?
* What data is used to support performance based budgeting requests?
* What departmental data is already published to Tempe’s Open Data Catalog or elsewhere on the city’s website?
* What data does your department use for internal performance or trend analysis?
* What data is reported to federal, state or local agencies?
* What data has been requested as a public information request?
* What data are other departments asking for?
* What kinds of data are similar departments across the country publishing as open data?

You may need to involve other subject matter experts in your department in order to answer these questions and identify datasets in your data sources.

**NOTE**: Identify or flag any datasets that you think may contain sensitive, protected or confidential information and include them in your inventory. Any flagged datasets will be reviewed in detail as part of the Prioritization process.

1. **Complete a dataset inventory form for each dataset**

For each dataset you identify in Step 2, document what is known as “metadata” or information detailing what the dataset includes.

Use the ***Data Inventory tab*** in the ***Data Inventory & Prioritization Template*** to document metadata for each dataset.

**3.3 Evaluate and Review Activities**

1. **Evaluate and prioritize data for publication**

*Evaluate*

Once your data inventory process is complete, the Open Data Manager or Open Data Team will review your department’s inventory and work with you to evaluate your department’s datasets.

Data quality impacts the usefulness of a released dataset, and so each dataset should be evaluated for quality as part of the evaluation process. Answering the following questions will help complete the Data Quality portion of the prioritization as well (outlined in Table 1 below)

* Are there missing values in the dataset that affect its usefulness?
* Is the dataset updated regularly and consistently?
* Is the data contained in the dataset accurate?
* Is the data in a machine-readable format?

*Prioritize*

After a dataset is identified and the inventory evaluated, the Data Coordinator, with assistance from the Department and Open Data Team, will evaluate each dataset for prioritization across five categories.

1. Strategic Importance
2. Collaborative Partnerships
3. External Interest
4. Operational Usefulness
5. Data Quality

Each category is explained in Table 1.

**Table 1. Prioritization Categories and Criteria**

|  |  |  |
| --- | --- | --- |
| **Prioritization Categories** | **Criteria** | **Prioritization Value** |
| Strategic Importance | * Is the dataset primarily about one of the Council’s priorities? * Is the data currently used to evaluate the performance of a department in delivering its strategic mission? * Is the dataset primarily about a project or program brought about by high-profile legislation | **0-5** |
| Collaborative Partnerships | * Is the dataset primarily focused on a national grant or award (e.g. a TIGER transportation project, U.S. Department of Energy, or a state-supported program)? * Does the dataset support a national or federal initiative (e.g. White House Police Data Initiative)? * Is the dataset part of any collaboration with the private sector (e.g. WAZE transportation data)? | **0-5** |
| External Interest | * Has an open records request been filed to receive access to the dataset? * Has the dataset been requested through the Open Data Portal? * Have residents requested the dataset during external events (e.g. hackathons, community engagement events)? * Does the dataset deal with a topic that is commonly valuable to external users, even if it hasn’t been requested? * Does the dataset have the capacity to fuel external innovation (e.g., enable the development of new tools and applications)? | **0-5** |
| Operational Usefulness | * Is the data used for decision making within a department? * Would opening the data increase internal government efficiency or effectiveness? * Did data users name the dataset as one that they use frequently? * Do usage statistics suggest that internal IPs are frequently calling upon this data to conduct their business? | **0-5** |
| Data Quality | * Are there missing values in the dataset that affect its usefulness? * Is the dataset updated regularly and consistently? * Is the data contained in the dataset accurate? * Is the data in a machine-readable format? | **0-5** |

Using the ***Prioritization tab*** in the ***Data Inventory & Prioritization Template,*** assign each dataset a 0-5 ranking (0 is no value and 5 is the highest value) for each category. A composite score will be automatically calculated in the spreadsheet.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Example Prioritization – Public Works** | | | | | | | |
| **Department** | **Dataset** | **Strategic** | **Partnerships** | **External** | **Operational** | **Quality** | **Total** |
| Public Works | Pavement Quality | 5 | 0 | 1 | 5 | 3 | 14 |
| Public Works | Water quality | 5 | 4 | 5 | 5 | 5 | 24 |
| Public Works | Graffiti clean up | 5 | 3 | 4 | 5 | 4 | 21 |

After each dataset is prioritized, the inventory of datasets will be ranked. Then in coordination with other departments’ inventories, datasets will be scheduled for release according to priority and quality.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Example Prioritization – All Departments** | | | | | | | |
| **Department** | **Dataset** | **Strategic** | **Partnerships** | **External** | **Operational** | **Quality** | **Total** |
| Mayor | 311 service requests | 4 | 4 | 5 | 4 | 5 | 22 |
| Public Works | Graffiti clean up | 5 | 3 | 4 | 5 | 4 | 21 |
| Community Development | Building permits | 5 | 0 | 4 | 5 | 2 | 16 |
| Internal Services | Employee pay & benefits | 0 | 0 | 5 | 4 | 5 | 14 |

1. **Review security and privacy criteria**

Each dataset must also be evaluated for security and privacy issues. The inventory of datasets will be ranked monthly and then reviewed and scheduled for release according to priority, security and privacy rankings.

* Does the dataset contain personally identifiable information (PII)?
* Does the dataset contain protected or sensitive information (PSI)?

If a dataset is identified as having PII or PSI, it may require a legal review for further assessment (this is discussed further in Section 3.4 Step 6).

Please see the **Sensitive Regulated Data Policy** for full City of Tempe privacy guidelines. The policy can be found on the ***Open Data SharePoint site***.

Table 2 provides definitions of what is classified as secure or private data.

**Table 2. Security and Privacy Criteria**

|  |  |
| --- | --- |
| **Security and Privacy Criteria** | |
| **Privacy** | |
| Personally Identifiable Information | Personally identifiable information (PII) is any information that can be used to identify, contact, or locate an individual, either alone or combined with other easily accessible sources. It includes information that is linked or linkable to an individual, such as medical, educational, financial and employment information such as the following   * Social Security number (There are additional restrictions on where Social Security numbers can be stored and shared.) * National ID number, passport number or Visa permit number * Driver's license number * Bank and credit/debit card numbers * Tax information (e.g., W-2, W-4, 1099) * Disability information * Ethnicity * Gender * The location of an individual at a particular time * Web sites visited * Materials downloaded * Any other information reflecting preferences and behaviors of an individual |
| **Security -** Secure information can be broken down into two categories: protected and sensitive | |
| Protected | Protected information means any dataset or portion thereof to which the City may deny access pursuant to A.R.S. 39-121 or any other applicable law, rule, regulation, court order or as otherwise required pursuant to established City.  Protected information includes data that the City is prohibited from disclosing by operation of law. Some protected information is specifically protected from disclosure by law and subject to strict handling requirements dictated by statutes, regulations or legal agreements including the federal Health Insurance Portability and Accountability Act (HIPAA), the federal Driver’s Privacy Protection Act, the federal Fair Credit Reporting Act and the federal Family Education Rights and Privacy Act (FERPA). |
| Sensitive | Sensitive information means any data which, if published by the City online, could raise privacy, confidentiality, proprietary or security concerns or have the potential to jeopardize public health, safety or welfare to an extent that is greater than the potential public benefit of publishing that data.  Datasets should be assessed for issues of confidentiality, intellectual property rights, financial and security risk to the City, and the protection of the privacy of individuals. |

Complete the ***Security and Privacy Worksheet*** for each dataset. Based on the results, go to the ***Security and Privacy tab*** in the ***Data Inventory & Prioritization Template*** and provide the results for each dataset.

**3.4 Preparing for Publication**

1. **Open Data Team Review**

Once the inventorying and prioritization processes are complete, we will schedule a review of the dataset(s).

***Review Protected and Private Classification***

For any datasets identified as having information identified as private or secure, we will work with you to identify the attributes that should not be released as open data (i.e. they were classified as PII, Protected or Sensitive).

Data that will violate security, privacy or legal concerns will not be released as open data.

If the dataset is determined to have privacy and/or security issues that require further scrutiny, we will halt the review process and the Open Data Manager will forward the dataset to the either the open data legal representative in the City Attorney’s Office for an in-depth privacy review and/or to the Open Data Steering Committee for a security review.

The three possible outcomes from a privacy and/or security review are:

1. Recommend for release without modification - The dataset is sent back to the Open Data Team for final approval for release.
2. Recommend for release with modifications (removal of fields that contain personally identifiable information, compromise security, etc.) – The dataset is sent back to the Open Data Team, who will work with the Open Data Coordinator to revise the dataset and resubmit for release.
3. Recommend against release – The dataset is sent back to the Open Data Team who will reclassify the dataset as not suitable for release as open data.
4. **Prepare metadata and data dictionary**

In order to submit a dataset for publication, you will need to complete the ***Data Submission Form***, which includes metadata and data dictionary information. 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 (for example, what processing has been done from the raw data to the publish data or what information a column contains). 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 (metadata) 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). This can be a narrative that provides context for the data or any information that you want users to know about the dataset.
* Keywords describing the data that help people searching for it  (check existing keywords to avoid duplication)
* 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.tempe.gov
* How often the dataset is published to data.tempe.gov

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

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

You will also include a one-record example of the dataset with all fields populated. Here is an example of a sample dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| **Field Name** | **Data Type** | **Value** | **Metadata Description** |
| INC\_NUMBER | Text | 2/15/2017-PC-6160 | Primary key (unique ID) for that record |
| SQ\_FOOT | SHORT | 1 | Estimated square footage of graffiti cleaned |
| LOCATION | Text | ROW | Description of graffiti location |
| PROP\_TYPE | Text | Utilities | Description of property where graffiti is located |
| NOTIFICATION | Text | Proactive Graffiti | Was the graffiti cleaned proactively or reported |
| … | … | … | … |

You can learn more about metadata for open data through the guide provided at <https://www.gitbook.com/book/centerforgov/open-data-metadata-guide/details>. The PDF is also provided on the ***Open Data SharePoint Site***.

1. **Enable automation or process for publication**

Once a dataset is ready for release as open data, we will coordinate with the Open Data Working Group to assess whether the publication of the dataset can and/or should be automated. Release frequency, data source, ease of access, staff resources and other factors will be used to weigh the benefits and feasibility of automating the publication of the dataset.

If it is determined that the publication should be automated, we will work with the data coordinators, internal staff and/or other resources as needed to automate the publication of the dataset.

If not, you will be responsible for loading the dataset to the Open Data Portal according to the release frequency as described in the data inventory for that dataset.

There are three methods for publishing datasets to data.tempe.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 possible for datasets that will be updated more frequently than quarterly. The Open Data Team will assist data providers with automated processing.

**Programmatic (API to API)**. This is ideal for real- or near-real-time transfers. The Open Data Team will manage processes that require application development skills.

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 (table/data format) you defined in the ***Data Submission Form***.

**Transform.** If necessary, modify the exported data so that the data structure matches the data schema (table/data format) defined in the ***Data Submission Form***.

**Load.** Using one of the publication methods described above, upload the data into <data.tempe.gov>.

1. **Schedule data release**

Next, the dataset will be scheduled for publication to the Open Data Portal. Several factors can affect when a dataset might be published, including automation work and staff availability. Some datasets may be planned for strategic release, which may include bundling complementary datasets together or with visualizations, storytelling components or other explanatory resources. The release of datasets may also be timed for release in support of an event, program or initiative.

The Data Coordinator, Communications and City Manager’s Office staff will all be informed of the proposed publication schedule at least 10 days prior to release. It is the your responsibility to keep your respective department informed of impending dataset releases.

At this stage, the Open Data Working Group may also need to work with other city staff to coordinate and manage any complementary visualizations, storytelling or other explanatory resources, as well as work with communications staff to schedule public and/or internal communications including news releases, social media posts, etc.

1. **Publish data**

Once the dataset is loaded, it will be available to review before being released to the public.

**Publish Privately**

Initially, the dataset will be marked "private" and is limited to you, selected testers, and the Open Data Team. We 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.

Publication of the dataset must conform to the principles of open data whenever possible. We will review the planned release against these principles

* **Accessible** – Data shall be made available to the broadest range of users for the widest range of purposes. Data shall be made available in bulk format.
* **Complete** – Data should be complete, not partial or a subset of a larger dataset (holding to any privacy or security restrictions).
* **Primary** –Data shall be collected at the source, with the highest possible level of granularity (holding to any privacy or security restrictions), not in aggregate or modified formats.
* **Timely** – Data shall be made available as soon as feasible in order to preserve the data's value.
* **License-free** – Data shall not be subject to any copyright, patent, trademark, regulation or trade secret.
* **Machine Readable**– Data shall be structured so that it can be read and interpreted by a computer program without the need for manual human intervention.
* **Non-discriminatory** – Data shall be available to anyone and does not require registration.
* **Non-proprietary** – Data shall be available in an open file format that does not require purchase of a specific piece of software or operating system in order to access the file

In accordance with the publication schedule, the dataset will be published to the Open Data Portal along with any other complementary resources.

Scheduled communications will also coincide with the dataset publication. Examples of scheduled communications could include a combination of:

* A post to the city’s Twitter, Facebook and/or other social media accounts;
* A news release from the Communications department;
* Direct communication to civic groups and organizations;
* Direct communication to local, state or federal agencies;
* The publication of articles, case studies or storytelling on the city’s website or other outlets such as local, state or national media outlets
* An announcement on the Open Data Portal

All datasets will be published under the Creative Commons CC Zero License and shall include a link to the license under which the dataset is published (<https://creativecommons.org/publicdomain/zero/1.0/legalcode>).

* 1. **Maintain and Audit Data**

1. Maintain and audit published datasets

After the dataset has been published, ongoing data coordinator activities include ensuring that the dataset is updated according to the publication frequency, handling questions about the dataset from the public, working with us to resolve any issues, and ensuring that any data is removed in accordance with the City’s retention policy.

At least once a quarter, you should review all of their department’s published datasets to ensure they are being updated on the basis included 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 to keep track of what has been published so it is easier to verify that the data are up-to-date.

If any changes are being considered to the structure of the data tables (for example, removing columns of information), change management guidelines are provided in the ***Open Data*** ***Change Management Guidelines*** on the ***Open Data SharePoint site***.

At least once per year we will conduct an audit of the data inventory and Open Data Catalog. The purpose of the audit is to ensure that datasets are being published per their publication frequency as identified in the data inventory; that datasets are properly tagged and categorized; that dataset metadata is complete and accurate; and that the data inventory and Open Data Catalog perform as expected.

The Open Data Manager will be responsible for producing a report that captures any outcomes from the audit to be delivered to the Open Data Working Group. We will be responsible for taking action to correct any issues with Open Data Portal, while the data coordinators will be responsible for correcting any issues with their respective datasets or inventory items.

**4. References**

This handbook was based on elements of open data handbooks from several cities including Boulder, Colorado, Seattle, Washington, San Francisco, California and San Jose, California. The willingness of these cities to share their materials with us is a strong example of the collaborations that come from open data initiatives.