CITY OF PORTLAND, OREGON
1851

2020 OPEN DATA HANDBOOK

OPEN DATA PROGRAM

SMART CITY PDX

[Preliminary Draft - 11-23-2020]

PORTLAND OPEN DATA HANDBOOK

EXECUTIVE SUMMARY

This Open Data Handbook shows the open data submission process. For workflow control, data submission creates a Data Manifest that includes different files including workflow control, candidate open dataset, metadata, and business rules to curate data.

In the last decade, the popularity of municipal open data programs has drastically increased. Common tenets of open data include an open-by-default orientation to a city's data, machine-readability, open formatting and the removal of licensing barriers to using and reusing open data by researchers, business, or the general public. Principles of open data don't stop there and also include aspects such as timeliness, permanence, use of bulk downloading and APIs, consideration of public feedback, and many other aspects.

The City of Portland, once a pioneer in open data, has fallen behind other city programs in many ways. The City's open data governance structure has many of the roles of other city programs, but lacks clearly assigned roles and responsibilities, and binding language and timelines. More challenging than this is the fractured state of the Open Data Program, which constitutes at least three different open data portals with differing data standards and portal features. The City's portals are not consistently updated and sometimes do not adhere to proper open data principles.

Data sharing and data integration are strategic issues in modern government to inform decision making by creating common knowledge. This is particularly important to support the general city objectives on racial equity, transparency, and community engagement, which are often peripheral.

This handbook represents a collective effort to consolidate open and accessible data in a single common place that may encourage trust between City Bureaus, other jurisdictions and the public.

CONTENT

INTRODUCTION

The open data handbook explains the process and steps the PDX Open Data program undertakes when a bureau submits an open dataset. More importantly, the handbook documents dataset rules, metadata requirements, and policies to make data consistent and standardized.

This applies to any dataset submitted for publication on the PDX open data portal. The purpose is three-fold, as follows:
1. Provides transparency and accountability for the City of Portland's bureaus data
2. Allows bureaus and the public to understand the overall data preparation, curation and publication processes.
3. Documents data policies, rules, requirements, and guidelines for open data consistency and standardization.

BACKGROUND

The City of Portland, according to the list provided by the Sunlight Foundation[1], was an early pioneer in municipal open data policy making by being the first city or state in 2009 to adopt an open data policy, and the second city or state to adopt such a policy overall.

Resolution No. 36735[2] addressed the basic tenets of open data affirming the principles of using technology to foster transparency, the need to share data freely, standardized data in machine readable formats, open standards, Open Source Software, and engagement with the local public and software community.

However, the ordinance was limited in its scope, largely delegating rulemaking to the Bureau of Technology Services with no deadline, oversight, or guidance on the matter. It directed the Bureau to:

a. Enter into agreements with our regional partners to publish and maintain public datasets that are open and freely available while respecting privacy and security concerns as identified by the City Attorney;
b. Develop a strategy to adopt prevailing open standards for data, documents, maps, and other formats of media;
c. Organize a regional contest to encourage the development of software applications to collect, organize, and share public data;

[1] Sunlight Foundation, "Policies by Date of Adoption," accessed September 24, 2020, https://opendatapolicyhub.sunlightfoundation.com/collection/by-date/.
[2] Portland City Council, "Resolution No. 36735," Portland, OR, (September 30, 2009), https://www.scribd.com/document/23617304/City-of-Portland-Resolution-No-36735-Regional-Technology-Community-Mobilization-and-Expansion-Resolution.

d. Establish best practices for analysis of business requirements in software review and selection processes, identify existing commercial software systems with licenses that are scheduled to expire in the near future, and encourage the consideration of Open Source Software in the review, replacement and continual improvement of business solutions;

e. Work with Travel Portland and regional partners to promote Portland as a host city for leading Open Source Software conferences and related technology events such as LinuxCon, Innotech, etc;[3]

The policy also directed "the City's Purchasing Agent to notify and distribute all formal technology related purchasing and contract opportunities for publication and distribution by the Software Association of Oregon, Oregon Entrepreneurs Network and the open source community in addition to those public notice requirements required under Portland City Code 5.33.300."[4]

While delegation of rulemaking to the bureaus or departments involved in their execution is a normal process of policymaking, that there were no deadlines or oversight attached to the above directives meant that much of the above was left unfulfilled over the following years, failing to coalesce into a comprehensive policy strategy.

This failure to coalesce is evident in the City's current approach, which is fractured and unorganized. The City of Portland's primary homepage pertaining to open data, "Maps, GIS & Open Data,"[5] currently provides links to the web viewer of Portland, a list of Metadata and description, the Portland GIS Open Data Site, a data portal called CivicApps.org, maps produced by Partner Bureaus, and resources by ESRI, the service platform used to host the GIS site. CivicApps.org was designed in 2009 and discontinued in 2013.

Then there is a another portal in use by the City of Portland's Police Bureau,[6] however there is no link to this portal from the primary portal web page. This portal consists of crime statistics, officer involved shootings, use of force reports, a traffic dashboard, dispatched calls, stolen vehicle statistics, stops data collection, traffic fatalities & serious injuries, and an open data feedback form.

On April 17, 2017, Portland City Council adopted a new open data policy, Ordinance No. 188356. This ordinance affirmed the earlier resolution's open data principles as well as the core policy challenges at hand:

[3] Ibid, 2.

[4] Ibid.

[5] Portland, Oregon, "Maps, GIS & Open Data," accessed September 25, 2020, https://www.portlandoregon.gov/28130.

[6] Portland Police Bureau, "Open Data," Portland, OR, accessed September 25, 2020, https://www.portlandoregon.gov/police/71673.

1. The City has no comprehensive, centralized list of existing datasets, and no process for prioritizing or reviewing data for release to the public.
2. No City policy requires City bureaus to collect, store, maintain, update, and release data to other agencies and the public on a regular basis. Multiple, redundant datasets exist across the City, leading to issues with data consistency, data quality, version control, interoperability and efficiency of access to information.
3. The number of public records requests to the City of Portland has increased substantially in recent years. A significant amount of staff time is devoted to addressing these requests.[7]

With council goals in mind, the ordinance established an official Open Data Policy, as well as an Open Data Program to implement said policy which was “to be committed to the publication, accessibility, and equitable and widespread sharing of data collected and generated by all City bureaus and by private sector companies, non-profit organizations, academic universities and other parties working on behalf of the City.”

The ordinance further directed that the City will strive to make data open by default. Last, the ordinance also directed the Mayor’s Office to establish and appoint members to a Data Governance Team, “responsible for providing guidance to departments and City Council on the overall direction of the City’s Open Data Program, including recommending updates to the Open Data Policy and publishing an annual report on progress toward achieving strategic goals for the Open Data Program.”

Much more guidance was attached to the 2017 ordinance than in the 2009 resolution. Exhibit A of Ordinance No. 188356 (City of Portland, 2017) directed the Data Governance Team to address the following 13 program elements and processes:

a. Define and memorialize strategic goals for the City’s Open Data Program…
b. Develop a system of governance for the Open Data Program…
c. Establish a timeline for Open Data Program implementation and project milestones…
d. Appoint a Data Steward within each data-generating City bureau…
e. Develop standards to determine which datasets are appropriate for public disclosure…
f. Create a comprehensive inventory of data…
g. Explore opportunities for open, standard data formats…
h. Explore data analytics methodologies and systems…
i. Establish mechanisms for prioritizing the collection of data to release…
j. Ensure that data released as part of the Open Data Program is made freely available…

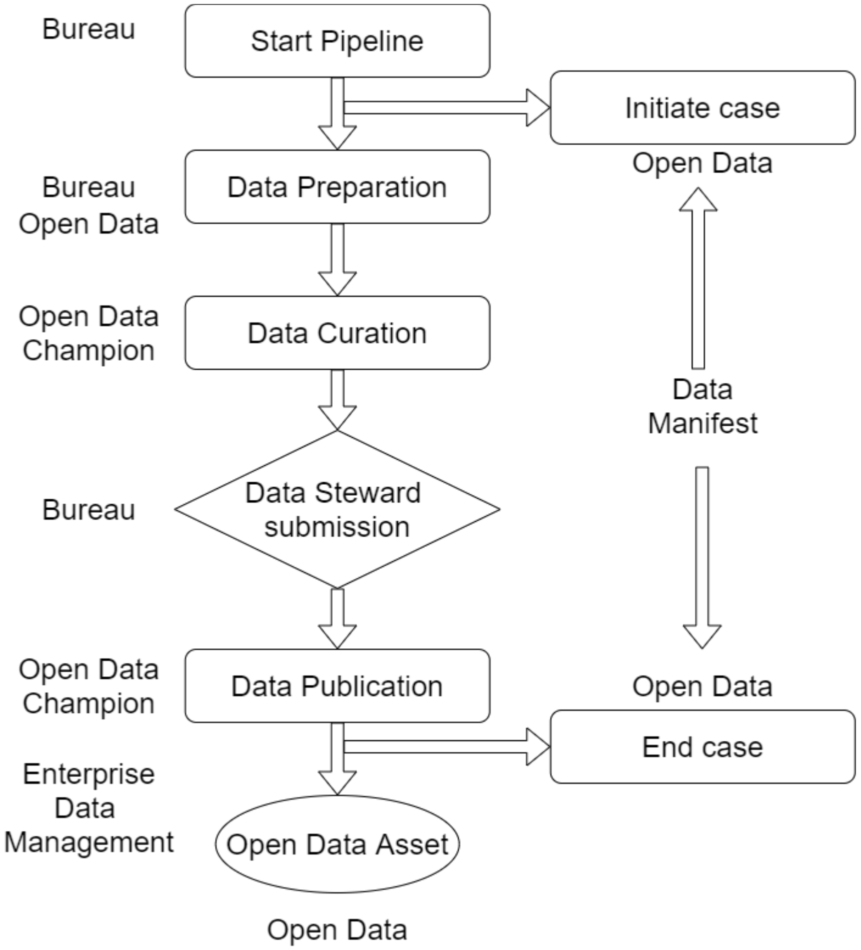
[7] Portland City Council, “Ordinance No. 188356,” Portland, OR, (April 17, 2017), https://www.portlandoregon.gov/cbo/article/636448.

k. Develop contract provisions that align with the goals of the Open Data Program…
l. Serve as a liaison with other state and local governments…
m. Identify methods for engaging external stakeholders…[8]

For this guide, the Smart City PDX researched exemplary and innovative open data programs in other jurisdictions (DC, New York City, Seattle, and San Francisco); as well as guidelines developed by the Federal government since 2015.

Discovering and using best practices, open data standards and strategies already developed by other jurisdictions would expedite Portland's implementation and management of its own open data program. Our smart city PDX and open data team are in deep appreciation for those who have walked the path before us.

DATA SUBMISSION PROCESS



[8] Ibid.

Portland Open Data (POD) submission process involves three steps as generally summarized below (a more detailed flow diagram is presented at the end of the Submission section):

Preparation (BUREAU/OFFICE): Initiate and document preparation, data profiling, metadata, licensing, business rules, internal QA/QC and compliance assessment.

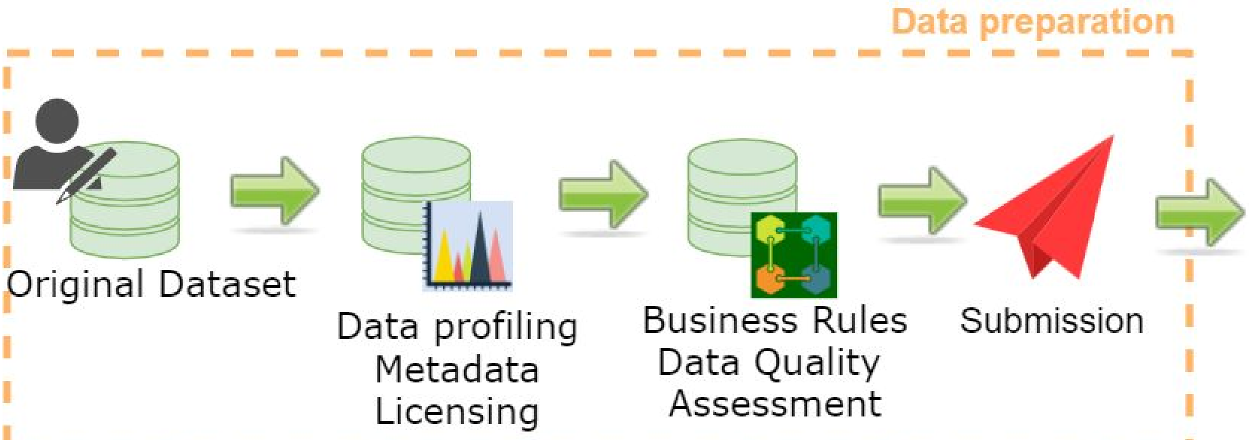
Curation (POD): Quality control, metadata testing, privacy assessment, social value assessment.

Publication (POD): Data federalization and integrity services, data validation, standardization, and linking, data analytics, reporting, data storage and data release and publication.

This process starts with the data owner at the city agency submitting a service ticket. The ticket will ask for basic information on the dataset and inform the data curation team to start the submission process.

Then, a bureau data champion will support the data submission by consulting through the preparation stage and performing data curation and assessments required before submitting the candidate open dataset to the Corporate Open Data Repository.

Data Preparation



The service ticket will then get assigned to a data champion within the bureau or the open data city network. Once assigned, the service ticket can be closed.

That champion will work with the data owner or representative and other resources to start constructing a Data Manifest (DM). A DM consists mainly of the following components:

Component	What is it?	Why is it important?
Data control form	A form that keeps history and control of operations done over the candidate dataset. It includes: 1. The case ticket ID 2. Bureau authorization forms 3. Data assessment reports Each dataset gets compared to the overall data quality and privacy rules and relevant requirements in the data submission form.This includes data integrity, quality, privacy assessments.	Data control lets the data champion know the status of the process. Having a ticket promotes accountability and the opportunity to develop metrics as a potential performance measure. Since rules and specifications will be applied to the data, the report will generally measure the level of effort in the quality assurance/quality control (QA/QC), and data transformation and clean-up parts of the curation process. It is also good information to provide to the agency regarding the quality of their data.
Data Submission Form	The form records information and specifications for the submission. It is broken into: 1. Data 2. Metadata 3. Transformation 4. miscellaneous information. It interacts with the enterprise data inventory (EDI). While the EDI records general information about the dataset itself, the submission form gets into specific technical details needed to process the dataset. (See Appendix B for the data dictionary)	This part allows data champions to perform the data curation process. By having the needed information, staff can independently and efficiently curate the data.
Data Descriptor	The data descriptor compiles three different layers describing the dataset structure and history. It is compiled by: ● **Data Dictionary.** Each dataset should include a data dictionary as a separate document. The data dictionary lists the table structure, with each column defined in easy to understand terms. This includes providing the values and	These documents help data ingestion and provide transparency on how data is generated and managed. The data dictionary permits the database administrator to define and create the required tables for the data. The data mapping makes creating the ETL process a lot easier and documented. The ETL process can

	descriptors for any domains. ● **Data Mapping Document.** A data mapping document is a special type of data dictionary that shows how data from the source maps to data in the destination database. This is used to define the extract, transform, and load (ETL) workflow that brings the data from the source database over to the destination. ● **Data documentation.** Additional documentation may include a survey instrument, data collection tool, study or report specific to the data, explanatory documentation for complex datasets, data lineage and other relevant documents.	be one of the most cumbersome parts of the curation process. Data documentation can be useful to end users to, for example, explain how the data might be utilized, as well as aid with interpretation and additional understanding of complex data.

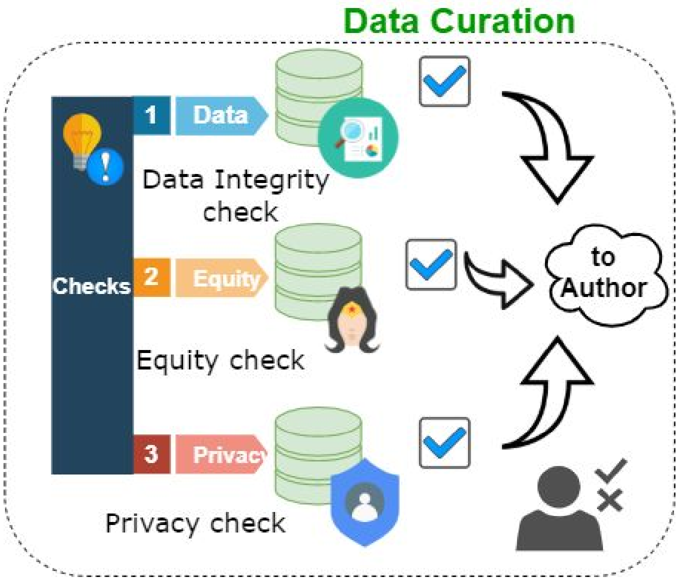
These files will be encapsulated in the Data Manifest in a tree structure as follows:

```
-/control
        /datastrategy.json
        /control.json
 /data
        /data.json
        /metadata.json
        /transformation.json (optional)
        /miscellaneous.json (optional)
 /descriptor
        /dictionary.json
        /mapping.json (optional)
        /documentation.json (optional)
```

Data Curation

Data curation involves reviewing and finalizing the Data Manifest content: information, specifications, and requirements. The documentation of the curation and validation process is documented in the datastrategy.json file.

It is divided in two processes: Data integrity check and privacy check. While the intended process includes an extra verification for cases where data is of high social impact.



Data Integrity involves:

(a) First, if needed, the service ticket is examined for content and authorization control. Next, the data submission form is thoroughly reviewed for content. Then the data dictionary is evaluated against the source data, including domain values.

(b) Next, the data mapping document is reviewed to ensure proper transformations are applied. Then, the metadata is reviewed with the data owner or agency representative for completeness.

(c) Finally, the data summarization report is examined for potential transformations that ensure data meets business rules.

Privacy check follows a privacy impact assessment and verification that all fields are not describing any private, personal, sensitive or confidential information. This process involves:

(a) Revision of the fields and descriptors in the dataset

(b) Check any personal identifiable, restricted or confidential information

(c) Check information that can contribute to re-identification of data and set a privacy risk ranking

Equity check is an assessment done when a high privacy risk has been found. This can include risk of data re-identification, dataset represents a high impact to a community, liability or any other legal impact, direct community request, high value data, or impacts on civil rights or civil liberties.

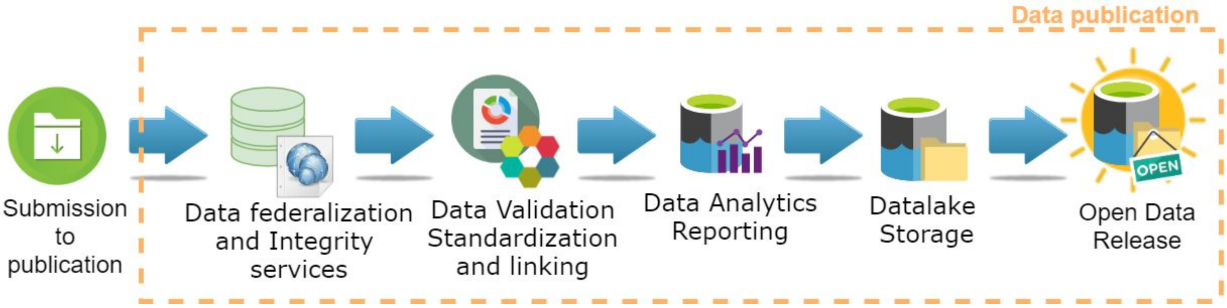
With the submission pack, staff can work with the data owner to review the quality of the data, obtain approvals if needed, finalize submission requirements, and finish documentation. Conversely, the consultation can determine that more work is needed on the agency's part before it can go to the curation phase.

Data Publication

Data publication relates to the ETL (Extract, Transform, Load) process that uploads datasets to Portland's enterprise repository. Data publication is a semi-automatic process where a data champion uploads a candidate open dataset that gets transformed from the data manifest to a geodatabase file that can be uploaded to the City of Portland Corporate GIS Data Repository.

The Data loader as a default, determines which datasets have been updated since last being loaded and updates those datasets in the Output Geodatabase. This process will be done by a data champion after the Data Manifest is ready and submitted.

Files data.json and metadata.json are transformed into a geodatabase file by mapping metadata fields into a ISO19115 compatible file as described in Appendix C.


Data publication
Submission to publication
Data federalization and Integrity services
Data Validation Standardization and linking
Data Analytics Reporting
Datalake Storage
OPEN
Open Data Release

Data Maintenance

Published data may have updates while it is available in the open data portal. These updates need to include proper documentation of changes and what kind of periodicity. A new submission needs to start when the schema of the table changes; for instance, new columns are added or the type of the fields have changed.

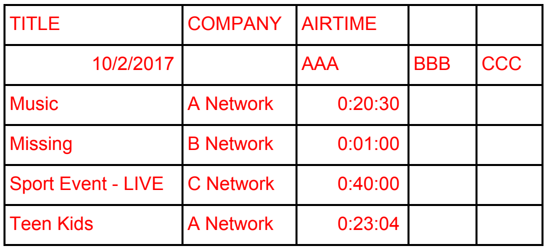
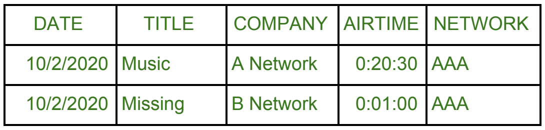
Published data will be stored as long as the defined retention time demands it. Then, data will be removed from the City repository. Any published dataset that may have a historic or use at the city, may be transferred to the City Archives.

DATA BUSINESS RULES

City bureaus and offices should follow the requirements of “tidy data” whenever possible and to the best of their abilities in publishing open tabular data. All agencies publishing tabular (table formatted), machine-readable data need to comply with the following data business rules.

These business rules are intended to increase the value of their data and make it easier for data to be analyzed, transferred, and used by others. The application of these business rules makes it easier to identify quality issues and creates a standard format for data on the open data portal, interoperability with other government jurisdictions, including state and federal agencies.

Data business rules will assure that published datasets follow quality and risk assessment standards that evaluate privacy and other value impacts for the City as enterprise and the generation of public good in the community. There are some examples where attaching this information brings great benefits to decision making and transparency[9].

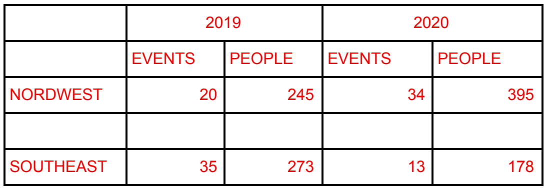
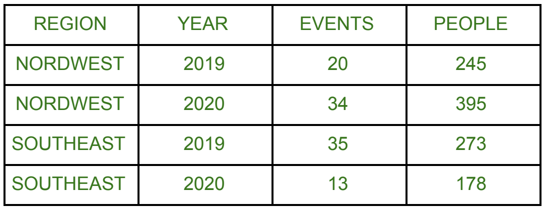
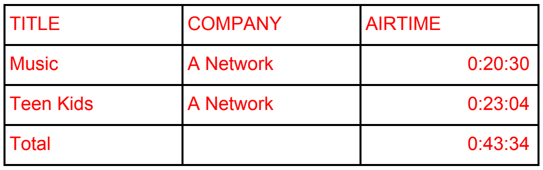
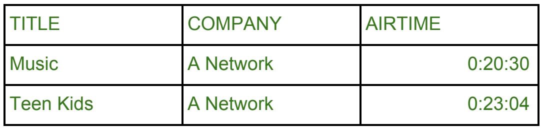
Data format and structure standards. This standards are placed to be compatible with the State of Oregon formats		
Data integrity rules	canonical machine-readable schema for describing action items for data integrity validation	
Rule	**Description**	**Examples**
1.1 Data Format		
Data completeness	Every qualitative and quantitative value must belong to a field (column) and a record (row)	 

TITLE	COMPANY	AIRTIME		
10/2/2017		AAA	BBB	CCC
Music	A Network	0:20:30		
Missing	B Network	0:01:00		
Sport Event - LIVE	C Network	0:40:00		
Teen Kids	A Network	0:23:04		

DATE	TITLE	COMPANY	AIRTIME	NETWORK
10/2/2020	Music	A Network	0:20:30	AAA
10/2/2020	Missing	B Network	0:01:00	AAA

[9] Datasheets for Datasets. Gebru et all. March 2020. https://arxiv.org/pdf/1803.09010.pdf

<table>
<tr><td></td><td></td><td>
<table>
<tr><td>10/2/2020</td><td>Sport Event - LIVE</td><td>C Network</td><td>0:40:00</td><td>AAA</td></tr>
<tr><td>10/2/2020</td><td>Teen Kids</td><td>A Network</td><td>0:23:04</td><td>BBB</td></tr>
</table>
</td></tr>
<tr><td>Data orientation</td><td>Horizontal data orientation should be restructured to vertical whenever possible. Vertical datasets are more easily consumed by applications and databases.</td><td>
<table>
<tr><td>TITLE</td><td>Music</td><td>Teen Kids</td><td>Missing</td></tr>
<tr><td>COMPANY</td><td>A Network</td><td>A Network</td><td>B Network</td></tr>
<tr><td>AIRTIME</td><td>0:20:30</td><td>0:23:04</td><td>0:01:00</td></tr>
</table>
<table>
<tr><td>TITLE</td><td>COMPANY</td><td>AIRTIME</td></tr>
<tr><td>Music</td><td>A Network</td><td>0:20:30</td></tr>
<tr><td>Teen Kids</td><td>A Network</td><td>0:23:04</td></tr>
<tr><td>Missing</td><td>B Network</td><td>0:01:00</td></tr>
</table>
</td></tr>
<tr><td>Header row</td><td>Data should contain one and only one header row. Multi-row headers are not acceptable. i.e. contains values for a single underlying attribute (height, duration, milepost, city, average cost). Fields can generally be classified as:
o Dimensions - qualitative values (e.g. a person’s race, a business’ industry) and date values
o Measurements - quantitative values measured at a point in time (e.g., a person’s age, a business’ sales, etc.). Measurements are generally based on counts or calculations.</td><td>
<table>
<tr><td>1st QUARTER REPORT</td><td></td><td></td><td></td><td></td></tr>
<tr><td>FY 2020</td><td></td><td></td><td></td><td></td></tr>
<tr><td>PROJECT</td><td></td><td></td><td></td><td></td></tr>
<tr><td>TITLE</td><td>MANAGER</td><td>BUDGET</td><td></td><td></td></tr>
</table>
<table>
<tr><td>PROJECT_TITLE</td><td>MANAGER</td><td>BUDGET</td><td>FISCAL YEAR</td><td>FYQUATER</td></tr>
</table>
</td></tr>
<tr><td>Row content</td><td>Each row is one observation, meaning each record contains all values for the same underlying unit (a family, a participant, a person, a business, a county, etc.) at the same point in time.</td><td></td></tr>
<tr><td>Empty cells</td><td>To clarify cells with no value, the following values should be assigned. If the blank field represents zero, then the field should be zero. "No value" in text formatted cells are always NULL. For numeric format, "no value" should be null except if zero is warranted.</td><td>
<table>
<tr><td>NAME</td><td>CATEGORY</td><td>PROJECT</td><td>TYPE</td><td>FY</td><td>NUM</td></tr>
<tr><td>1776</td><td>Meals</td><td>Basketball</td><td>TV</td><td>2016</td><td>100</td></tr>
<tr><td>1327</td><td>N/A</td><td>Baseball</td><td>TV</td><td>2016</td><td></td></tr>
<tr><td>7736</td><td>Lodging</td><td>Football</td><td>TV</td><td>2016</td><td>50</td></tr>
</table>
<table>
<tr><td>NAME</td><td>CATEGORY</td><td>PROJECT</td><td>TYPE</td><td>FY</td><td>NUM</td></tr>
<tr><td>1776</td><td>Meals</td><td>Basketball</td><td>TV</td><td>2016</td><td>100</td></tr>
<tr><td>1327</td><td></td><td>Baseball</td><td>TV</td><td>2016</td><td>0</td></tr>
<tr><td>7736</td><td>Lodging</td><td>Football</td><td>TV</td><td>2016</td><td>50</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td>Cell integrity</td><td>Data should only have records associated with the same type of underlying unit</td><td>i.e., underlying units represented in records should not be a mix of zip codes and census tract.</td></tr>
<tr><td>Grouped Data</td><td>Data tidiness[10]. Data should have only one header row (column listing), and should not have “spacer rows,” multi-row headers. Do not group cells using headers between rows.</td><td>
<table>
<tr><th></th><th>2019</th><th></th><th>2020</th><th></th></tr>
<tr><td></td><td>EVENTS</td><td>PEOPLE</td><td>EVENTS</td><td>PEOPLE</td></tr>
<tr><td>NORDWEST</td><td>20</td><td>245</td><td>34</td><td>395</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>SOUTHEAST</td><td>35</td><td>273</td><td>13</td><td>178</td></tr>
</table>

<table>
<tr><th>REGION</th><th>YEAR</th><th>EVENTS</th><th>PEOPLE</th></tr>
<tr><td>NORDWEST</td><td>2019</td><td>20</td><td>245</td></tr>
<tr><td>NORDWEST</td><td>2020</td><td>34</td><td>395</td></tr>
<tr><td>SOUTHEAST</td><td>2019</td><td>35</td><td>273</td></tr>
<tr><td>SOUTHEAST</td><td>2020</td><td>13</td><td>178</td></tr>
</table></td></tr>
<tr><td>Summarized Data</td><td>Avoid including roll-ups, subtotal and total of values in cells as part of the column. Typically, applications can compute these values and have totals of subtotals skews results.</td><td>
<table>
<tr><th>TITLE</th><th>COMPANY</th><th>AIRTIME</th></tr>
<tr><td>Music</td><td>A Network</td><td>0:20:30</td></tr>
<tr><td>Teen Kids</td><td>A Network</td><td>0:23:04</td></tr>
<tr><td>Total</td><td></td><td>0:43:34</td></tr>
</table>

<table>
<tr><th>TITLE</th><th>COMPANY</th><th>AIRTIME</th></tr>
<tr><td>Music</td><td>A Network</td><td>0:20:30</td></tr>
<tr><td>Teen Kids</td><td>A Network</td><td>0:23:04</td></tr>
</table></td></tr>
<tr><td><b>1.2 Data Field</b></td><td></td><td></td></tr>
<tr><td>Column Names</td><td>System column names must be all upper case and limited to 30 characters and must start with an alphabetic character. Use only alphanumeric characters and period (.), dash(-) or underscore (_). Avoid use of abbreviations. Instead, use the title case for field names and be sure that the names match that in the Data Dictionary. Aliases reflect real-world context, use simple language, names limited to 30 characters, initcaps words and spaces to separate words.</td><td></td></tr>
</table>

[10] Tidy Data. Hadley Wickham. 2014. https://www.jstatsoft.org/article/view/v059i10

record identifiers	Where possible, datasets should contain a primary key, row identifier, or unique identifier for each row contained within the data. Unique Row IDs allow for automated updating of datasets and make it easier to amend records in the future and avoid data duplication.	Codes work best for unique identifiers. Where there are multiple rows for an individual observation (e.g. new rows added for different points in time), concatenating a unique ID and a record date can be used to create a unique identifier (e.g. 15703-20201003)
record date	Where a record date exists, records within a dataset should have a field that contains the record date. The record date represents the date the measurements were taken or recorded. In instances where datasets only include the most current information for an underlying unit, the record date should represent the date the values within the record were last modified.	The date of the inspection, the date of the permit, sales date, count date
Log date	Log date refers to the time and date that the information was uploaded. It differs from 'record date' as this represents date and time that information was collected, not when information was uploaded. It is named as ‘generated”.	
Leading or Trailing Spaces	Text fields must be trimmed of leading or trailing space(s).	
Codes	Agencies should use industry and government specific codes and standards when possible. Codes such as expense codes, object codes, geographic features, or classifications facilitate relating datasets to one another and in performing analyses with external data or data from other regions.	● Object codes or accounting codes used in the State accounting system. ● Federal Geographic Data Committee endorsed standards, such as the Geographic Names Information System ● The North American Industry Classification System (NAICS) code used by Federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. ● Standard Occupational Classification (SOC) system code is a federal statistical standard used by federal agencies to classify workers into occupational categories for the purpose of collecting, calculating, or disseminating data. All workers are classified into detailed occupations according to their occupational definition. Detailed occupations are combined to form broad occupations, minor groups, and major groups.

<table>
<tr>
<td></td>
<td></td>
<td><table>
<tr><td>Cost object - Bureau Code</td><td>Description Bureau or Business Area</td></tr>
<tr><td>AT</td><td>Office of the City Attorney</td></tr>
<tr><td>AU</td><td>Office of the City Auditor</td></tr>
<tr><td>BO</td><td>City Budget Office</td></tr>
</table></td>
</tr>
<tr>
<td>Column Order</td>
<td>Column order does not impact reporting or analysis, but it supports human readability of open data by allowing users to visually scan column headings. Recommended Column order is as follows (left to right):
● The record identifier/unique identifier
● The record date (e.g. sales date, permit date, inspection date, etc.)
● The underlying unit associated with the record (i.e., field containing the business name, permittee, etc.)
● Fields with dimensions
● Fields with measurements
Related fields should be placed next to one another for readability and clarity.</td>
<td>Examples would be formatting addresses so that Street Name, City, State, Zip Code are contiguous, or placing a NAICS code next to the corresponding industry value in another column.</td>
</tr>
<tr>
<td><b>1.3 Formatting values</b></td>
<td colspan="2">Formatting suggestions and recommendations are included in the table and should be used by city bureaus and offices when structuring and publishing a dataset.</td>
</tr>
<tr>
<td>Data File Format</td>
<td>Separate data fields with a comma and enclose values in double quotes. Keep each record on a separate line. Do not follow the last records in a file with a carriage return. In the first line of the file, include a header with a list of the column names in the file. The header list is separated in the same way as the rest of the file.</td>
<td>“NAME”,”ID”,”PHONE”
“John Doe”,”7”,”202-555-5555”
“Jane Doe”,”8”,”555-555-5555”</td>
</tr>
<tr>
<td>Text Field Format</td>
<td>UTF-8 encoded alphanumeric text. Text values should be all upper case, lower case, or initial caps. Special characters and text formatting won't be preserved. Text should be provided as plain text and not include html tags or formatting.</td>
<td><table>
<tr><td>SCHOOL</td></tr>
<tr><td>A Middle School</td></tr>
<tr><td>B HIGH SCHOOL</td></tr>
<tr><td>a high school</td></tr>
</table>
<table>
<tr><td>SCHOOL</td></tr>
<tr><td>A Middle School</td></tr>
<tr><td>B High School</td></tr>
<tr><td>A High School</td></tr>
</table>
<table>
<tr><td>SCHOOL</td></tr>
<tr><td>A MIDDLE SCHOOL</td></tr>
<tr><td>B HIGH SCHOOL</td></tr>
<tr><td>A HIGH SCHOOL</td></tr>
</table></td>
</tr>
</table>

<table>
<tr><th></th><th></th><th></th></tr>
<tr>
<td>Numeric Field Values</td>
<td>Do not mix text in a field that is intended to contain numeric or date data. Any numerical values, including decimals, negatives, or other values without special symbols (%, $, °, etc.). Do not include commas in large number formats.</td>
<td>
<table>
<tr><th>PROJECT</th><th>PROGRESS</th></tr>
<tr><td>A Construction</td><td>80</td></tr>
<tr><td>B Highway</td><td>50</td></tr>
<tr><td>C Building</td><td>25 %</td></tr>
</table>
<table>
<tr><th>PROJECT</th><th>PROGRESS</th></tr>
<tr><td>A Construction</td><td>80</td></tr>
<tr><td>B Highway</td><td>50</td></tr>
<tr><td>C Building</td><td>25</td></tr>
</table>
</td>
</tr>
<tr>
<td>Monetary Fields</td>
<td>Numeric data that represents money should be provided with either no decimal places or two decimal places.</td>
<td>
<table>
<tr><th>PROJECT</th><th>COST</th></tr>
<tr><td>A Construction</td><td>$500,500.22</td></tr>
<tr><td>B Highway</td><td>-55,250,000.00</td></tr>
<tr><td>C Playground</td><td>50K</td></tr>
</table>
<table>
<tr><th>PROJECT</th><th>COST</th></tr>
<tr><td>A Construction</td><td>500500.22</td></tr>
<tr><td>B Highway</td><td>55250000</td></tr>
<tr><td>C Playground</td><td>50000</td></tr>
</table>
</td>
</tr>
<tr>
<td>Negative Values</td>
<td>Negative values should be preceded with a minus-sign (-), not placed within parentheses or another notation.</td>
<td>
<table>
<tr><th>NAME</th><th>KEY_ INDEX</th></tr>
<tr><td>A AGENCY</td><td>Negative 50</td></tr>
<tr><td>B AGENCY</td><td>(10)</td></tr>
</table>
<table>
<tr><th>NAME</th><th>KEY_INDEX</th></tr>
<tr><td>A AGENCY</td><td>-50</td></tr>
<tr><td>B AGENCY</td><td>-10</td></tr>
</table>
</td>
</tr>
<tr>
<td>Percentage Values</td>
<td>For currency or percentage columns, upload the values as numeric fields with no special characters and apply the “currency” or “percent”</td>
<td>
<table>
<tr><th>PROJECT</th><th>COST</th><th>PROGRESS</th></tr>
<tr><td>Project A</td><td>$40,000</td><td>60%</td></tr>
<tr><td>Project B</td><td>$23,000</td><td>80%</td></tr>
</table>
<table>
<tr><th>PROGRESS</th><th>COST</th><th>PROGRESS</th></tr>
<tr><td>Project A</td><td>40000</td><td>60</td></tr>
<tr><td>Project B</td><td>23000</td><td>80</td></tr>
</table>
</td>
</tr>
<tr>
<td>Codes with Leading Zeroes</td>
<td>Identification numbers and numeric codes (FIPS codes, NAICS codes, SIC codes, etc.) where leading zeroes are part of the values. Columns must be assigned as text format preventing the loss of leading zeroes.</td>
<td>
<table>
<tr><th>SIC</th><th>INDUSTRY</th></tr>
<tr><td>191</td><td>General Farms, Primarily Crop</td></tr>
<tr><td>1521</td><td>General Contractors-Single-Family Houses</td></tr>
</table>
</td>
</tr>
</table>

<table>
<tr>
<td>Date and Time Fields</td>
<td>Time should be stored as military (i.e. 24-hour time). If common day time is used, it is stored as same format, so the data can be read as AM or PM. Time is presented as PT. Dates are automatically parsed by default in the PST timezone. Timezones can be adjusted in the dataset, as can display options for dates. Supported ISO 8601 date formats as well as dates in the following format:
MMM d, yyyy (e.g. "Jan 4, 1982")
MMM d, yy (e.g. "Jan 4, 82")
MMMM d, yyyy (e.g. "January 4, 1982")
MMMM d, yy (e.g. "January 4, 82")
M-d-yyyy (e.g. "1-4-1982")
M/d/yyyy (e.g. "1/4/1982")
M.d.yyyy (e.g. "1.4.1982")
M-d-yy (e.g. "1-4-82")
M/d/yy (e.g. "1/4/82")
M.d.yy (e.g. "1.4.82")</td>
<td>Recommended date formatting options for readability and consistency are:
yyyy-MM-dd (e.g. 2020-01-21)
yyyy-MM-ddTHH:mm:ss (e.g., 2019-01-22T00:00:00)
yyyy-MM-dd HH:mm:ss (e.g., 2018-01-22 00:00:00)
<table>
<tr><th>UPDATED_DATE</th><th>UPDATED_TIME</th></tr>
<tr><td>2020-09-15</td><td>10:15:20</td></tr>
<tr><td>2020-09-15</td><td>16:25:45</td></tr>
</table>
<table>
<tr><th>UPDATED_DATE</th><th>UPDATED_TIME</th></tr>
<tr><td>2020-09-15</td><td>10:15 AM</td></tr>
<tr><td>2020-09-15</td><td>4:25 PM</td></tr>
</table>
</td>
</tr>
<tr>
<td>Zip Codes</td>
<td>Five-digit or nine-digit Zip Codes are acceptable. Consistency within a dataset is critical. Nine-digit Zip Codes can be provided as hyphenated values (i.e.12345-9876). Do not mix both formats within the same column. Field definitions must be text.</td>
<td>
<table>
<tr><th>NAME</th><th>ZIPCODE</th></tr>
<tr><td>A Building</td><td>97201</td></tr>
<tr><td>B Building</td><td>97204</td></tr>
</table>
<table>
<tr><th>NAME</th><th>ZIPCODE</th></tr>
<tr><td>A Building</td><td>97201-5350</td></tr>
<tr><td>B Building</td><td>97204-1900</td></tr>
</table>
</td>
</tr>
<tr>
<td>Phone Numbers</td>
<td>Phone numbers must include area code. Area codes are mandatory. The format is XXX-XXX-XXXX.</td>
<td>
<table>
<tr><th>NAME</th><th>PHONE</th></tr>
<tr><td>A Building</td><td>555-555-5555</td></tr>
<tr><td>B Building</td><td>5555555555</td></tr>
<tr><td>C Store</td><td>(555) 555-5555</td></tr>
</table>
<table>
<tr><th>NAME</th><th>PHONE</th></tr>
<tr><td>A Building</td><td>555-555-5555</td></tr>
<tr><td>B Building</td><td>555-555-5555</td></tr>
<tr><td>C Store</td><td>555-555-5555</td></tr>
</table>
</td>
</tr>
<tr>
<td>Name Field</td>
<td>The primary name of a feature shall be stored in a column named as NAME.</td>
<td></td>
</tr>
</table>

<table>
<tr><td>Unique Identifier</td><td>For Open Date and GIS publication, OBJECTID (auto generated sequential number) will be added. OBJECTID will not be used as a unique code. Another column, whether contained in the data or assigned, must be used as the unique identifier. In addition, there will be a GIS_ID in GIS layers which is coded as "table name"_<num> where the number is randomized. This will be set in the <b>geodatabase</b> as the primary key.</td><td><table>
<tr><th>EDI_ID</th><th>GIS_ID</th><th>Object ID</th><th>SERVICE_ID</th></tr>
<tr><td>1</td><td>EDI_1</td><td>1</td><td>SERVICEREQUEST_1</td></tr>
<tr><td>2</td><td>EDI_2</td><td>2</td><td>SERVICEREQUEST_2</td></tr>
<tr><td>3</td><td>EDI_3</td><td>3</td><td>SERVICEREQUEST_3</td></tr>
</table></td></tr>
<tr><td>Web links</td><td>Although discouraged due to heavy maintenance, if a web url is needed it will be stored in a field called URL. Only one URL can be entered into a cell and in the following formats <a href="https://www.portland.gov/">City of Portland website</a> https://www.portland.gov/</td><td><table>
<tr><th>NAME</th><th>URL</th></tr>
<tr><td>Portland</td><td>https://www.portland.gov</td></tr>
<tr><td>Smart City PDX</td><td>https://www.smartcitypdx.com</td></tr>
</table></td></tr>
<tr><td>Email</td><td>An email address must be made up of a local-part, an @ symbol, then domain.</td><td><table>
<tr><th>EMAIL</th></tr>
<tr><td>Tayen Liu@portland.gov</td></tr>
<tr><td>Tayen Liu at portland.gov</td></tr>
<tr><td>Tayen Liu@ portland.gov</td></tr>
</table>
<table>
<tr><th>EMAIL</th></tr>
<tr><td>TayenLiu@portland.gov</td></tr>
<tr><td>Tayen.Liu@portland.gov</td></tr>
</table></td></tr>
<tr><td>Checkboxes</td><td>Checkbox and binary values are acceptable formats for a dataset.
Valid false values: {0, f, false, n, no, off}
Valid true values: {1, t, true, y, yes, on}</td><td><table>
<tr><th>COURSE</th><th>COMPLETED</th></tr>
<tr><td>Mathematics</td><td>✓</td></tr>
<tr><td>English</td><td>*</td></tr>
<tr><td>Physics</td><td>NOK</td></tr>
</table>
<table>
<tr><th>COURSE</th><th>COMPLETED</th></tr>
<tr><td>Mathematics</td><td>TRUE</td></tr>
<tr><td>English</td><td>TRUE</td></tr>
<tr><td>Physics</td><td>FALSE</td></tr>
</table></td></tr>
<tr><td><b>1.4 Addressing</b></td><td></td><td></td></tr>
<tr><td>Address Data</td><td>Address data will always be run through the Portland Master Address Repository (MAR). This gives the data a common format. In addition, the following fields will be kept: MAR_ID, XCOORD, YCOORD, LATITUDE, and LONGITUDE. The address will be stored in the format contained in FULLADDRESS.</td><td><table>
<tr><th>ADDRESS</th><th>MAR_ID</th><th>XCOORD</th><th>YCOORD</th><th>LATITUDE</th><th>LONGITUDE</th></tr>
<tr><td>1221 SW 4th Ave</td><td>308596</td><td>94540.5</td><td>63339.5</td><td>45.51135</td><td>-122.68679</td></tr>
</table></td></tr>
<tr><td>Address</td><td>If the address is broken out into separate fields, the street number, name, type, and quadrant must be combined in a single</td><td><table>
<tr><th>NAME</th><th>STREENUMBER</th><th>STREET</th><th>QUADRANT</th></tr>
<tr><td>City Hall</td><td>1221</td><td>4th Ave</td><td>SW</td></tr>
</table></td></tr>
</table>

<table>
<tr>
<td></td>
<td>field called ADDRESS and then geocoded against the MAR.</td>
<td>
<table>
<tr><td>BPS</td><td>1900</td><td>4th Ave</td><td>SW</td></tr>
</table>
<table>
<tr><th>NAME</th><th>ADDRESS</th></tr>
<tr><td>City Hall</td><td>1221 SW 4th Ave</td></tr>
<tr><td>BPS</td><td>1900 SW 4th Ave</td></tr>
</table>
</td>
</tr>
<tr>
<td>City, State, and Zip</td>
<td>City State and zip codes should be separated out of the address and stored in separate columns named as CITY, STATE, and ZIPCODE.</td>
<td>
<table>
<tr><th>NAME</th><th>ADDRESS</th></tr>
<tr><td>City Hall</td><td>1221 SW 4th Ave, Portland, OR 97204</td></tr>
<tr><td>BPS</td><td>1900 SW 4th Ave, Portland, OR 97201</td></tr>
</table>
<table>
<tr><th>NAME</th><th>ADDRESS</th><th>CITY</th><th>STATE</th><th>ZIPCODE</th></tr>
<tr><td>City Hall</td><td>1221 SW 4th Ave</td><td>Portland</td><td>OR</td><td>97204</td></tr>
<tr><td>BPS</td><td>1900 SW 4th Ave</td><td>Portland</td><td>OR</td><td>97201</td></tr>
</table>
</td>
</tr>
<tr>
<td>Latitude and Longitude</td>
<td>Geolocation based on Latitude and Longitude. Latitude and longitude, if available, should be provided in two separate fields. Values should be in decimal degrees. Latitude is bounded by 90 and -90, and longitude is bounded by 180 and -180.
For point columns the format should be: POINT (long lat)</td>
<td>
<table>
<tr><th>ADDRESS</th><th>LATITUDE</th><th>LONGITUDE</th></tr>
<tr><td>1221 SW 4th Ave</td><td>45.51135</td><td>-122.68679</td></tr>
<tr><td>1900 SW 4th Ave</td><td>45.50929</td><td>-122.68091</td></tr>
</table>
</td>
</tr>
<tr>
<td><b>1.4 Miscellaneous</b></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domains</td>
<td>Domains must have a dimension table explaining the coded domain values. Furthermore, when export tables that have domains, the description will be included.</td>
<td></td>
</tr>
<tr>
<td>Geometry</td>
<td>Geometry in spatial layers will not contain corrupt geometry. This includes null, self-intersecting, short-segment, incorrect ring order, incorrect segment orientation, unclosed rings, empty parts, duplicate vertices, mismatched spatial attributes, discontinuous parts, empty Z values, bad envelopes, and incorrect geospatial extents. This can cause errors in applications.</td>
<td></td>
</tr>
<tr>
<td>Topology</td>
<td>Based on data, topology will be checked.</td>
<td></td>
</tr>
</table>

	If a polygon data continuously covers all District boundary, the dataset should not have any gaps or overlaps.	

METADATA STANDARD WITH EXAMPLES

Dataset documentation is a critical component of the curation process. It consists of the data and metadata (data about the data). It allows users to fully understand the data content and context, including caveats and data limitations. POD’s metadata standard is compatible with the Project Open Data Metadata Schema (PODv1.1 or DCAT-US Schema v1.1)[11]. Metadata description with an example is below:

Label	Definition	Required	Values
Schema Version	URI that identifies the version of the Project Open Data schema being used.	Always	{"conformsTo": "https://resources.data.gov/resources/dcat-us/"}
Dataset	A container for the array of Dataset objects. See Dataset Fields below for details.	Always	{
Metadata Context	URL or JSON object for the JSON-LD Context that defines the schema used.	No	{"@context": "https://project-open-data.cio.gov/v1.1/schema/catalog.jsonld"}
Metadata Catalog ID	IRI for the JSON-LD Node Identifier of the Catalog. This should be the URL of the data.json file itself.	No	{"@id":"https://project-open-data.cio.gov/v1.1/schema/dataset.json"} But it needs to be edited.
Metadata Type	IRI for the JSON-LD data type. This should be dcat:Catalog for the Catalog.	No	{"@type": "dcat:Catalog"}
Data Dictionary	URL for the JSON Schema file that defines the schema used.	No	{"describedBy": "https://project-open-data.cio.gov/v1.1/schema/catalog.json"}

[11] https://resources.data.gov/resources/podm-field-mapping/

Label	Definition	Required	Values
Metadata Type	IRI for the JSON-LD data type. This should be dcat:Dataset for each Dataset.	No	{"dcat":"Dataset"}
Title	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.	Always	User input
Description	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest.	Always	User input
Tags	Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.	Always	User input
Last Update	Most recent date on which the dataset was changed, updated or modified.	Always	Set automatically. Dates should be ISO 8601.Example: {"modified":"2001-01-15"}
Publisher	The publishing entity and optionally their parent organization(s).	Always	Set from user attributes. Example: https://project-open-data.cio.gov/v1.1/schema/#publisher
Contact Name and Email	Contact person's name and email for the asset.	Always	Set from user attributes. Example: https://project-open-data.cio.gov/v1.1/schema/#contactPoint
Unique Identifier	A unique identifier for the dataset or API as maintained within an Agency catalog or database.	Always	set by program Hash(ID) or any other federated management
Public Access Level	The degree to which this dataset could be made publicly-available, regardless of whether it has been made available	Always	See Data Classification below.
License	The license or non-license (i.e. Public Domain) status with which the dataset or API has been published. See Open Licenses for more information.	If-Applicable	Creative Commons Attribution 4.0 (CC-BY-4.0)

Rights	This may include information regarding access or restrictions based on privacy, security, or other policies. This should also serve as an explanation for the selected “accessLevel” including instructions for how to access a restricted file, if applicable, or explanation for why a “non-public” or “restricted public” data asset is not “public,” if applicable. Text, 255 characters.	If-Applicable	Leave Empty. We still need to develop business rules to define the access constraints and who has it (groups definition)
Spatial	The range of spatial applicability of a dataset. Could include a spatial region like a bounding box or a named place.	If-Applicable	If-Applicable, a geographic polygonal region. Use the following schema: https://project-open-data.cio.gov/v1.1/schema/#spatial
Temporal	The range of temporal applicability of a dataset (i.e., a start and end date of applicability for the data).	If-Applicable	If-Applicable, a timestamp window. Use this format: https://en.wikipedia.org/wiki/ISO_8601#Time_intervals
Distribution	A container for the array of Distribution objects. See Dataset Distribution Fields below for details.	If-Applicable	This field describes how this dataset can be accessible. See example: https://project-open-data.cio.gov/v1.1/schema/#distribution
Frequency	The frequency with which the dataset is published.	No	User defined. Format should comply: https://en.wikipedia.org/wiki/ISO_8601#Durations https://project-open-data.cio.gov/iso8601_guidance/#accrualperiodicity
Data Standard	URI used to identify a standardized specification the dataset conforms to.URI that serves as a unique identifier for the standard or Data model	No	Use this definition: https://project-open-data.cio.gov/v1.1/schema/#dataset-conformsTo. Look at distribution model
Data Dictionary	URL to the data dictionary for the dataset (taxonomies and ontologies). Note that documentation other than a data dictionary can be referenced	No	it is a schema pointing to the specific glossary used for this taxonomy. Example: {"describedBy": "http://www.agency.gov/veg

	using Related Documents (references).		etables/schema.json"}
Data Dictionary Type	The machine-readable file format (IANA Media Type also known as MIME Type) of the dataset's Data Dictionary (describedBy).	No	{"describedByType": "application/schema+json"} https://www.iana.org/assignments/media-types/media-types.xhtml
Collection	The collection of which the dataset is a subset.	No	use the Uri in the form: {"isPartOf":"http://dx.doi.org/10.7927/H4PZ56R2"}
Release Date	Date of formal issuance.	No	Set automatically. Dates should be ISO 8601.Example: {"issued":"2001-01-15"}
Language	The language of the dataset.	No	This should adhere to the RFC 5646 standard. Example: {"language":["en-US"]} or if multiple languages, {"language":["es-MX","wo","nv","en-US"]}
Homepage URL	This field is not intended for an agency's homepage (e.g. www.agency.gov), but rather if a dataset has a human-friendly hub or landing page that users can be directed to for all resources tied to the dataset.	No	Leave empty for now. {"landingPage":"http://www.agency.gov/vegetables"}
Related Documents	Related documents such as technical information about a dataset, developer documentation, etc.	No	Array of strings (URLs). We are looking into bibliographic reference managers URI. Example (as now): {"references":["http://www.agency.gov/legumes/legumes_data_documentation.html"]}
Category	Main thematic category of the dataset.	No	Array of strings.{"theme":["vegetables","produce"]}

Label	Definition	Required	All this fields should be filled up automatically after the dataset is processed
Metadata Type	IRI for the JSON-LD data type. This should be dcat:Distribution for each Distribution.	No	dcat:Distribution
Access URL	URL providing indirect access to a dataset, for example via API or a graphical interface.	If-Applicable	Url for accessing this dataset (API, GUI or sftp)
Data Standard	URI used to identify a standardized specification the distribution conforms to.	No	context that needs to be the same as the one on the dataset above
Data Dictionary	URL to the data dictionary for the distribution found at the downloadURL. Note that documentation other than a data dictionary can be referenced using Related Documents as shown in the expanded fields.	No	it is a schema pointing to the specific glossary used for this taxonomy. Example: {"describedBy": "http://www.agency.gov/vegetables/schema.json"}
Data Dictionary Type	The machine-readable file format (IANA Media Type or MIME Type) of the distribution's describedBy URL.	No	{"describedByType": "application/schema+json"}
Description	Human-readable description of the distribution.	No	This can be created with the description above plus the way to access it. Example: {"description":"Vegetable data as a zipped CSV file with attached data dictionary"}
Download URL	URL providing direct access to a downloadable file of a dataset.	If-Applicable	Direct access point to the dataset.
Format	A human-readable description of the file format of a distribution.	No	Example: {"format":"JSON"}
Media Type	The machine-readable file format (IANA Media Type or MIME Type) of the distribution's downloadURL.	If-Applicable	Example: {"mediaType":"text/csv"} . It should comply any of these: http://www.iana.org/assignments/media-types/media-types.xhtml
Title	Human-readable name of the distribution.	No	Example: {"title":"listofvegetables.csv"}

DATA QUALITY

Bureaus should implement an internal quality assurance or quality checking process for open datasets. The rigor of the quality check is dependent upon the dataset the agency intends to publish and the standards provided below are guidelines and advice, but not requirements.

Checking for data quality and establishing a process for quality assurance builds more confidence in our data and helps to avoid publishing datasets with errors or datasets that would later require corrections. These guidelines are not requirements, but instead present options and avenues for bureaus to check their data quality. Bureaus may opt to build their own internal dataset quality and evaluation procedures.

Data quality checklist

Use the following checklist as guidance for checking the quality and completeness of your dataset before publishing it. For large datasets, consider looking at the first few rows, last few rows, and a few rows in the middle at random. If you are working with a database, you can write select statements to select *MIN* and *MAX* values, or to select the *TOP* few or last few values to validate that there are no extreme or unexpected values (e.g. default 1/1/1753 dates).

<table>
<tr><th>data quality rules</th><th>Description</th><th>Examples</th></tr>
<tr><td><b>1. Reasonableness of values</b></td><td></td><td></td></tr>
<tr><td>1.1 - Overall, does the data you are looking at make sense?</td><td>Look for common knowledge and expertise to validate your data</td><td>Your data looks off or does not match your expectations.</td></tr>
<tr><td>1.2 - Do values match the column headings (dates in date fields, emails in email columns)</td><td>Verify expected range of values in each column. Define ranges for testing.</td><td>
<table>
<tr><th>MONTHLY PUBLIC EVENTS</th><th>ATTENDANCE</th><th>DATE</th></tr>
<tr><td>Event A</td><td>35</td><td>2/13/2020</td></tr>
<tr><td>Event B</td><td>20499</td><td>2/18/2020</td></tr>
<tr><td>Event C</td><td>59</td><td>12/6/2019</td></tr>
</table>
<table>
<tr><th>MONTHLY PUBLIC EVENTS</th><th>ATTENDANCE</th><th>DATE</th></tr>
<tr><td>Event A</td><td>35</td><td>2/13/2020</td></tr>
<tr><td>Event B</td><td>20</td><td>2/18/2020</td></tr>
<tr><td>Event C</td><td>59</td><td>2/26/2020</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th></th><th></th><th></th></tr>
<tr><td>1.3 - Do data values match any formatting rules, such as numbers in number columns, only alphabetical characters in business or organization name fields?</td><td>Verify that names in text fields are correct, as much as possible</td><td>
<table>
<tr><th>BUSINESS NAME</th><th>PERMIT NUMBER</th><th>DATE</th></tr>
<tr><td>RED $PACE</td><td>3420</td><td>03-24-2020</td></tr>
<tr><td>BLUE SPACE</td><td>2!23</td><td>05-30-2020</td></tr>
</table>
<table>
<tr><th>BUSINESS NAME</th><th>PERMIT NUMBER</th><th>DATE</th></tr>
<tr><td>RED SPACE</td><td>3420</td><td>03-24-2020</td></tr>
<tr><td>BLUE SPACE</td><td>2123</td><td>05-30-2020</td></tr>
</table>
</td></tr>
<tr><td>1.4 - Are values the expected or appropriate length, such as 5 or 9-digit zip codes, 10-digit phone numbers? The SQL LEN function can locate numbers that are the incorrect length</td><td>Whenever possible, use parsing tools to verify known formats like zip codes, phone numbers, emails, etc.</td><td>
<table>
<tr><th>BUSINESS NAME</th><th>PHONE NUMBER</th><th>ZIP CODE</th></tr>
<tr><td>RED SPACE</td><td>503-55-5555</td><td>97201-65</td></tr>
<tr><td>BLUE SPACE</td><td>888-888</td><td>9720</td></tr>
</table>
<table>
<tr><th>BUSINESS NAME</th><th>PHONE NUMBER</th><th>ZIP CODE</th></tr>
<tr><td>RED SPACE</td><td>503-555-5555</td><td>97201-5350</td></tr>
<tr><td>BLUE SPACE</td><td>971-888-888</td><td>97204</td></tr>
</table>
</td></tr>
<tr><td>1.5 - Are there unexpected negatives, commas or other formatting where there should not be?</td><td>Check any unexpected anomaly in data due to misplaced negative signs or commas.</td><td>
<table>
<tr><th>DATE</th><th>RESPONSE TIME</th><th>COST OF SERVICE</th></tr>
<tr><td>01-01-2020</td><td>15</td><td>1,000,000</td></tr>
<tr><td>01-02-2020</td><td>-5</td><td>10.00</td></tr>
<tr><td>01-03-2020</td><td>12</td><td>120.00</td></tr>
</table>
<table>
<tr><th>DATE</th><th>RESPONSE TIME</th><th>COST OF SERVICE</th></tr>
<tr><td>01-01-2020</td><td>15</td><td>1,000.00</td></tr>
<tr><td>01-02-2020</td><td>5</td><td>10.00</td></tr>
<tr><td>01-03-2020</td><td>12</td><td>120.00</td></tr>
</table>
</td></tr>
<tr><td>1.6 - Do any dates fall outside a specific time period or return a default date value, such as 1/1/1700?</td><td>Verify expected dates in cell content.</td><td>
<table>
<tr><th>DATE</th><th>RESPONSE TIME</th></tr>
<tr><td>01-01-20200</td><td>15</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td></td><td><table><tr><td>01-02-2020</td><td>5</td></tr></table>
<table><tr><th>DATE</th><th>RESPONSE TIME</th></tr><tr><td>01-01-2020</td><td>15</td></tr><tr><td>01-02-2020</td><td>5</td></tr></table></td></tr>
<tr><td><b>2. Duplicate values or records</b></td><td></td><td></td></tr>
<tr><td>2.1 - Verify that there are not duplicate records or rows within the dataset</td><td>Row records must be unique. Remove any duplicate</td><td><table><tr><th>DATE</th><th>RESPONSE TIME</th><th>COST OF SERVICE</th></tr><tr><td>01-01-2020</td><td>15</td><td>1,000.00</td></tr><tr><td>01-02-2020</td><td>5</td><td>10</td></tr><tr><td>01-03-2020</td><td>12</td><td>120</td></tr><tr><td>01-02-2020</td><td>5</td><td>10</td></tr></table>
<table><tr><th>DATE</th><th>RESPONSE TIME</th><th>COST OF SERVICE</th></tr><tr><td>01-01-2020</td><td>15</td><td>1,000.00</td></tr><tr><td>01-02-2020</td><td>5</td><td>10</td></tr><tr><td>01-03-2020</td><td>12</td><td>120</td></tr></table></td></tr>
<tr><td>2.2 - Compare the total number of records or rows in the dataset and any reports within the source system (e.g. COUNT total rows) or based upon familiarity with the dataset itself. Do 10,000 rows make sense, or is it more likely that there are duplicate records?</td><td>Verify the total number of records and compare with the expected number of entries for the dataset.</td><td></td></tr>
<tr><td>3. Inconsistent values</td><td></td><td></td></tr>
<tr><td>3.1 - Check for inconsistently reported values or a lack of standardization across the dataset. Using a wildcard (*) search if possible can pull up similar but slightly different values in a SQL statement.</td><td>Look at the data dictionary. Compare known fields with the expected content (i.e. dates, phone numbers, emails, urls, etc.)</td><td></td></tr>
<tr><td>3.2 - Cities, counties, and other location fields are common areas for inconsistency. Does the dataset contain Lane, Lane Co, Lane</td><td>Make sure geolocation of an address or region keeps the same nomination.</td><td><table><tr><th>COMMUNITY CENTER</th><th>VISITS</th></tr><tr><td>MT. Tabor</td><td>356</td></tr></table></td></tr>
</table>

<table>
<tr><td>County, LANE C, or Mt. Angel, Mount Angel, M. Angel?</td><td></td><td><table><tr><td>MOUNT SCOTT</td><td>250</td></tr><tr><td>COLUMBIA</td><td>538</td></tr></table><table><tr><th>COMMUNITY CENTER</th><th>VISITS</th></tr><tr><td>MOUNT TABOR</td><td>356</td></tr><tr><td>MOUNT SCOTT</td><td>250</td></tr><tr><td>COLUMBIA</td><td>538</td></tr></table></td></tr>
<tr><td>3.3 - Are agency names and acronyms used interchangeably or applied inconsistently?</td><td>Avoid using interchangeable definitions, acronyms and short names.</td><td>E.g. BHR, Bureau of Human Resources, Human resources, HR</td></tr>
<tr><td>3.4 - For other data collection points, look for multiple versions that represent the same value, such as Male, M, m or F, Female, f.</td><td>Check known data fields defining collection points from the data dictionary consistent.</td><td>E.g. Dr. Doctor, Sr. Senior, Jr., Junior, etc.</td></tr>
<tr><td><b>4. Null/missing values</b></td><td></td><td></td></tr>
<tr><td>4.1 - Check your data for null values. How many records contain null values for each field? Consistency helps users of the data to know what to expect when values are purposely null.</td><td>Make sure Null or corrupted cells are identified and create rules to manage those cases in your dataset</td><td><table><tr><th>DATE</th><th>RESPONSE TIME</th><th>COST OF SERVICE</th></tr><tr><td>01-01-2020</td><td>15</td><td>0</td></tr><tr><td>01-02-2020</td><td></td><td>10.00</td></tr><tr><td>01-03-2020</td><td>12</td><td>120.00</td></tr></table><table><tr><th>DATE</th><th>RESPONSE TIME</th><th>COST OF SERVICE</th></tr><tr><td>01-01-2020</td><td>15</td><td>NULL</td></tr><tr><td>01-02-2020</td><td>NULL</td><td>10</td></tr><tr><td>01-03-2020</td><td>12</td><td>120</td></tr></table></td></tr>
<tr><td>4.2 - Is the number of nulls acceptable?</td><td>Too many Nulls may decrease data quality and value</td><td></td></tr>
<tr><td>4.3 - Is there a pattern as to where there are null values?</td><td>Identify any pattern of Null values. If a data field is consistently missing values, it might be possible to eliminate it without affecting the integrity of the dataset.</td><td></td></tr>
<tr><td>4.4 - Where null values are allowed, are they treated consistently throughout the dataset (i.e., empty</td><td>Use consistent ways to set a Null value. The recommendation is “NULL”</td><td></td></tr>
</table>

values, “NULL”, “NA”, and not a combination of each)?		
5. Test calculated or derived values		
5.1 - If the dataset contains any calculated or derived values, perform a spot check by replicating the calculating. Are the replicated results the same as those within the dataset?	Validate any calculated or derived value in the dataset. Document formulas whenever possible and viable. Build a data lineage document if the dataset is critical.	
5.2 - Check for any outliers (negatives where the calculation can only result in a positive, unreasonably large or small numbers) to ensure the dataset does not contain any miscalculated fields. Using a MAX, MIN or TOP select statement can assist in retrieving outlier records from a database.	Make sure any calculated value brings a reasonable outcome. Multiple layers of calculation or cross-table formulas may multiple small errors. Switching operation or updating formulas may break connection between cells.	
5.3 Check for Statistical Significance. Statistical analysis may show misleading values when looking at only one statistical component (i.e. only average when looking at popular values)	Statistical significance refers to the claim that a result from data generated by testing or experimentation is not likely to occur randomly or by chance but is instead likely to be attributable to a specific cause. When analyzing a data set and doing the necessary tests to discern whether one or more variables have an effect on an outcome, strong statistical significance helps support the fact that the results are real and not caused by luck or chance.	
5.4 Check for Reasonable inference. Does the conclusion derived from data make sense and proper biases and data profile verification have happened? This is particularly critical in large projects and socially impactful decision making.	When deriving conclusions from data, make sure that inferences draw on highly diverse and feature-rich data of unpredictable value and create new opportunities for discriminatory, biased, and privacy-invasive profiling and decision-making.	

5.5 Trash-in Trash-out. Check for quality of data sources when harvesting from different sources. Is your calculation off or biased?	Data quality of third party sources can impact the final quality of calculations, particularly when using complex formulas and statistical approaches(i.e. in machine learning or AI).	
6. Slice data		
6.1 - Slice data into subgroups based upon categories or specific units of time. Does there appear to be internal consistency amongst values and subgroups, without any extreme outliers or over-representation of specific fields?	If data contains multiple categories in a single column, the analyst may slice or separate those groups in order to prevent over-representation	For instance, if a mix of sample data that represents mostly white higher income neighborhoods, while intending to represent all demographics.
6.2 - Identify any potential patterns or shifts over time that seem inconsistent either with previous observations.	Datasets overtime may shift or represent other types of anomalies. Create a new dataset for the new dynamics.	For instance, mobility patterns before and after COVID19 emergency.
7. Check the schema		
7.1 - Review the listed columns to ensure they are in the proper order and are comprehensive, and there are no formatting issues associated with the column names or fields.	Make the best effort for sorting columns in a priority order; even when analysis software may not consider the order, it is important for human analysts.	
7.2 - Does your data file contain columns that are not in your dataset?	Verify your data dictionary, which represents the table schema, for any mismatch with real data.	
7.3 - Are columns missing?	Make sure all described columns are part of the dataset.	
7.4 - Check any formatting for views or tables	Make sure the table and transformations are well formed	
8. Geospatial Data Quality		
8.1 - Agencies publishing spatial data should check their data layers before publishing to ensure they do not contain corrupt geometry.	Make sure geospatial data represented in layers is properly mapped into the corresponding geometries.	

8.1 - Check for null, self-intersecting, short-segment, incorrect ring order, incorrect segment orientation, unclosed rings, empty parts, duplicate vertices, mismatched spatial attributes, discontinuous parts, empty Z values, bad envelopes, and incorrect geospatial extents, and adjust data layers accordingly.	Run your geometry verification scripts and make sure geospatial data is ready.	

Correcting/Cleaning Data

If the data quality check identifies any issues, unexpected values, or other quality concerns, the data coordinator or individual publishing the data should work closely with any subject matter experts for the data and any database administrators or IT staff who are assisting with extracting or formatting the dataset. For datasets built upon views or select statements from a relational database, performing a walkthrough of the select statement or using the DESCRIBE function to review the formatting of the view itself may identify where a field is mismatched, incorrectly calculated, or formatted improperly.

Inconsistencies that are part of a larger quality issue with the dataset and cannot be corrected through changing a view or extraction process should be documented within the “Limitations” metadata field to educate users on any limitations or quality considerations for the dataset itself. Some limitations or quality issues that should be documented are:

- Substituted or imputed values in place of missing values
- Missing values that are omitted entirely
- Known inconsistencies or missing data for specific record types or rows

APPENDIX A - REFERENCE DATA CONTROL

A1. Private or Restricted Data Dictionary

Below is a list of the reference information in a data dictionary. This list represents data that have privacy or public records requests considerations.

Data Element Name	Data Type	Element Definition	Validation Rules	Required
DATE_CREATED	Date	Record Creation Date	Date	Y
DATE_MODIFIED	Date	Record Modified Date	Date	N
URI_ID	string	Unique Resource Identifier	Authorized ID number	Y
LAST_MODIFIED_BY	String	Record Last Modified User	Personal Name	Y
DATA_INVENTORY_ID	String	Data Inventory Id	Authorized Catalog number	Y
BUREAU_CODE	String	Bureau Abbreviation	See Bureau Codes.	Y
BUREAU_NAME	String	Bureau Name	See Bureau Codes.	Y
DATA_SET_NAME	String	Dataset Name	String less than 80 characters. No special symbols.	Y
DATABASE_NAME	String	Database Name	String less than 80 characters. No special symbols.	N
LINKED_APPLICATION_NAME	String	Linked Application Name	String less than 80 characters. No special symbols.	N
DATA_CATALOG_NAME	String	Data Catalog Name	String less than 80 characters. No special symbols.	N
DATA_DIAGRAM_FILE_PATH	String	Data Diagram File Path	String Long describing an access path as a directory or URL	Y
DATA_SET_AUDIENCE_VALUE	String	Dataset Audience Value	String Long	N
DATA_SET_CLASSIFICATION_NAME	String	Dataset Security Classification	String. See Data Classification.	Y
DATA_SET_CLASSIFICATION_REASON	String	Dataset Classification Justification	String Long	N
DATA_PROGRAM_OWNER	String	Data Owner Bureau Program Abbreviation	See Bureau Codes.	Y
DATA_SET_CATEGORY	String	Dataset Topic Category	String. See Data Categories	N
DATA_SET_TYPE	String	Dataset Type	String. See Data Categories	Y
EARLIEST_DATE_OF_RECORDS	Date	Earliest Date of Available Record	Date	Y
RECENT_DATA_OF_RECOR	Date	Recent Date of Record	Date	Y

DS				
UPDATE_INTERVAL	String	Dataset Update Frequency	Format should comply: https://en.wikipedia.org/wiki/ISO_8601#Durations	Y
RETENTION_SCHEDULE	String	Data Retention Schedule	Format should comply: https://en.wikipedia.org/wiki/ISO_8601#Durations	Y
PUBLICATION_BARRIERS	String	Publication Barrier	String	N
BARRIER_SUMMARY	String	Description of Barrier	String	N
PUBLIC_INTEREST_IN_DATA_VALUE	String	Public Interest in Dataset	String	N
DATA_SET_URL	String	Dataset Website Link	URL	N
KEYWORDS	String	Dataset Keywords/Tags	Array of categories in string format	Y
NAME_DC_DESIGNEE	String	Data Champion Designee Name	Individual Name or ID	N
EMAIL_DC_DESIGNEE	String	Data Champion Designee Email	E-mail	N
CREATED_DATA	Date	Data Collection or Creation Date	Date	N
MODIFIED_DATA	Date	Data Modification Date	Date	N
DCS_LAST_MOD_DTTM	Date	Data Last Modified Date	Date	N
REVIEWED_BY_CHAMPION	String	Flag for data champion review	Boolean flag	Y
DATA_INTEGRITY_PASSED	String	Data Integrity Assessment Flag	Boolean flag	Y
PRIVACY_ASSESSMENT_PASSED	String	Privacy Assessment Flag	Boolean flag	Y
SOCIAL_IMPACTS_PASSED	String	Social Impacts Assessment Flag	Boolean flag	N
ISSUE_IDENTIFIED	String	Identified Issue	Boolean flag	Y
DERIVATIVE_DATA_SET	String	Derivative Dataset	Boolean flag	Y
ISSUE_NOTES	String	Issue Notes	String	N
CURR_STATUS	String	Current Status	Document flow control status	Y
SUBMISSION_DATE	Date	Data submission date	Date	Y
PUBLISHED_DATE	Date	Data publication date	Date	Y
COMMENTS	String	Comments	String	Y
LATEST_ASSESSMENT	Date	Latest Assessment Date	Date	Y
OPENDATA_PORTAL	String	Defines whether dataset is published in Open Data Portal.	Boolean flag	Y

A2. Privacy and Public Records Data Dictionary

Bureaus are encouraged to include other fields that are relevant to the specific dataset, such as mapping fields to forms or specific data collections, calculations or imputed values, or any other information that provides context on the data within the dataset.

Report any issue or question on privacy or public records exemptions to a data champion.

Data Element Name	Data Type	Element Definition	Validation Rules
ADDRESS_EMPLOYEE_HOME	String	Employee home address	Must be removed.
ADDRESS_PUBLIC_HOME	String	home address from an individual from the public	Provide (unless some other privacy interest, such as providing the information would lead to harassment)
BANK_ACCOUNT_CODE	String	Bank account code or pin number	Must set Confidential and must remove.
BANK_ACCOUNT_NUMBER	String	Bank account number	Must set Confidential and must remove.
BIOMETRICS	String	Biometric information	Must be removed. Set as Confidential.
BIRTH_DATE	Date	date of birth	Must be removed.
DEVICE_ID_NUMBER	String	Electronic device ID number	Recommend remove. Set as Restricted. If medical device, then must be removed and set as confidential.
DEVICE_SERIAL_NUMBER	String	Electronic device serial number	Recommend remove. Set as Restricted.
DRIVERS_LICENSE	String	Driver's license	Must be removed.
EMAIL_ADDRESS_BUSINESS	String	contact email from a business	Provide.
EMAIL_ADDRESS_EMPLOYEE PERSONAL	String	personal city employee email	Must be removed.
EMAIL_ADDRESS_PUBLIC_EMPLOYEE	String	city employee public email address	Provide. A series of emails linked together by email responses and forwarding should be treated as a single document. A review of the series of emails should be done to identify privileged communications/redactions. If exempt material is found, that part should be redacted and the remainder can be disclosed.

EMAIL_ADDRESS_PUBLIC_PERSONAL	String	personal email from an individual from the public	Must be removed.
EMPLOYEE_BADGE_IDCARD	String	City employee badge ID	Must be removed.
EMPLOYEE_DATE_ADMISSION	String	Date of admission of city employee	Recommend remove. Set as Restricted.
EMPLOYEE_DATE_DISCHARGE	String	Date of discharge of city employee.	Recommend remove. Set as Restricted.
FINANCIAL_ACCOUNT	String	Any financial accounts	Must be removed.
GENETIC_INFORMATION	String	Genetic information	Must be removed. Set as Confidential.
HEALTH_PLAN_BENEFICIARY_NUMBERS	String	Health plan beneficiary number	Must set Confidential and must remove.
LICENSE_PLATE_EMPLOYEE	String	Employee license plate number	Recommend remove. Set as Restricted.
LICENSE_PLATE_PUBLIC	String	license plate number from a vehicle owned by an individual from the public	Provide. Must remove if received from PPB (LEDS).
MEDICAL_HISTORY	String	Medical history	Must be removed. Contact City Attorney.
MINORS_BIRTH_DATE	String	date of birth of a minor	Must be removed. Set as Confidential.
MINORS_EMAIL	String	Minor's email	Must be removed. Set as Confidential.
MINORS_NAME	String	the name of a minor	Must be removed. Set as Confidential.
PAYMENT_CARD_CAV2, _CVC2, _CVV2, or _CID	String	Payment card CAV2, CVC2, CVV2, or CID	Must set Confidential and must remove.
PAYMENT_CARDHOLDER_EXPIRATION_DATE	String	Payment cardholder expiration date	Must set Confidential and must remove.
PAYMENT_CARDHOLDER_NAME	String	Payment cardholder name	Must set Confidential and must remove.
PAYMENT_CARDHOLDER_NUMBER	String	Payment cardholder number	Must set Confidential and must remove.
PERSONNEL_DISCIPLINE_ACTION	String	Applies to completed disciplinary actions when a sanction is imposed, and materials or documents that support that particular disciplinary action, fall within the scope of this exemption. Also applies during investigation to determine disciplinary action.	Contact City Attorney. Recommend remove or withhold.
PERSONNEL_DISCIPLINE_ACTIONS_NO_SANCTION	String	Applies to personnel investigation of a public safety employee which does not result in discipline.	Contact City Attorney. For Public Safety Officers: Recommend remove or withhold

PERSONNEL_FILE_DATA	String	Data from a city personnel file.	Contact City Attorney. Must remove information listed in other categories.
PUBLIC JOB_INFO	String	Job information of an individual from the public.	Recommend remove. Set as Restricted.
PUBLIC_AGE	String	Age of an individual from the public	Recommend remove. Set as Restricted.
PUBLIC_CRIMINAL_RECORD	String	Criminal record of an individual from the public.	Must be removed. Set as Confidential.
PUBLIC_DATE_DEATH	String	Date of dead of an individual from the public.	Recommend remove. Set as Restricted.
PUBLIC_EDUCATION	String	Education of an individual from the public.	Recommend remove. Set as Restricted.
PUBLIC_ETHNICITY	String	Ethnicity of an individual from the public.	Recommend remove. Set as Restricted.
PUBLIC_FINANCIAL_STATUS	String	Financial Status of an individual from the public.	Recommend remove. Set as Restricted.
PUBLIC_GENDER	String	gender of an individual of the public	Recommend remove. Set as Restricted.
PUBLIC_GRADES	String	Grades of an individual from the public.	Recommend remove. Set as Restricted.
PUBLIC_IDENTITY_NAME	String	Identity or login name of an individual from the public.	Recommend remove. Set as Restricted.
PUBLIC_INTERNET_PROTOCOL	String	IP address of an individual from the public	Recommend remove. Set as Restricted.
PUBLIC_MAIDEN_NAME	String	Maiden name of an individual from the public	Recommend remove. Set as Restricted.
PUBLIC_NAME	String	Name of an individual from the public	Recommend remove. Set as Restricted.
PUBLIC_PERSONAL_WEBSITE	String	Personal website of an individual from the public	Recommend remove. Set as Restricted.
PUBLIC_RACE	String	Race of an individual	Recommend remove. Set as Restricted.
PUBLIC_SALARY	String	Salary of an individual from the public.	Recommend remove. Set as Restricted.
PUBLIC_SCHOOL_ATTENDANCE	String	School attendance of an individual from the public	Recommend remove. Set as Restricted.
PUBLIC_SECURITY_DATA	String	Records or information that would reveal or otherwise identify security measures of individuals, buildings or other property, information processing, communication, and telecommunication systems.	Contact City Attorney. Must be removed.
PUBLIC_WORKPLACE	String	Place of work of an individual from the public	Recommend remove. Set as Restricted.

PUBLIC_WORSHIP_PLACE	String	Place of worship of an individual from the public	Recommend remove. Set as Restricted.
PUBLIS_SAFETY_PLANS_DATA	String	security plans that could impact the physical safety of any individual or jeopardize law enforcement activities.	Contact City Attorney. Must be removed.
RECORDS_PERTAINING_TO_LITIGATION	String		Contact City Attorney. Must be removed.
RESUME_APPLICANT_DATA	String	City position applicant data. It may include home address, personal home number, or other personal identifiable information, sexual self-identification, hobbies, and interests.	Contact City Attorney. Seek City Attorney advice; withhold except for finalists for high level positions such as Bureau directors.
RESUME_EMPLOYEE_DATA	String	City employee resume. It may include home address, personal home number, or other personal identifiable information, sexual self-identification, hobbies, and interests.	Contact City Attorney. For Current Employees: Provide, but must remove personal identifying information (home address, personal phone number, etc), hobbies, and interests.
SOCIAL_SECURITY_NUMBER	String	Social Security number	Must be removed. Set as Confidential.
TELEPHONE_NUMBER_BUSINESS	String	Telephone number of a business	Provide.
TELEPHONE_NUMBER_EMPLOYEE	String	Personal telephone number of a city employee	Must be removed.
TELEPHONE_NUMBER_PUBLIC	String	Public telephone number of an individual from the public	Recommend remove if combined with name and address.
TESTING_MATERIALS_DATA	String	Testing and evaluation materials	Recommend remove. Set as Restricted.
TESTING_SCORES_DATA	String	Testing and evaluation scores	Recommend remove. Set as Restricted.
WHISTLEBLOWER_NAME	String	name of an individual who is a whistleblower. There might be an ongoing investigation.	Contact City Attorney. Recommend remove. Depends on whether the investigation is ongoing.
WITNESS_NAME	String	Name of a witness on a criminal investigation	Recommend remove. Set as Restricted.

A3. Data Classification

Data classification comes from the Bureau of Technology Services Administrative Rule 2.18. The open data program recommends an additional classification level when disclosing data may create major damage or injuries.

VALUE	DESCRIPTION
Public	Information approved for general public access. This would include general public information, published reference documents (within copyright restrictions), open source material and press releases. This type of information should still be protected against threats to the integrity of the information.
Restricted	Information which is intended strictly for use within the City. Although most 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. This would include information associated with internal email systems, City user account activity information and certain personnel information.
Confidential	Information that is sensitive in nature requires significant controls and protection. Unauthorized disclosure of this information could have a serious adverse impact on the City or individuals and organizations who interact with the City. This information includes but is not limited to: 1) cardholder data subject to the Payment Card Industry- Data Security Standard (PCI DSS), 2) personally identifiable information as defined by the Oregon Identity Theft Protection Act (ORS 646A.600) or the Fair and Accurate Credit Transactions Act of 2003 (also known as the "Red Flag Rules"). This information may be subject to public disclosure laws, 3) Protected Health Information (PHI) as defined by the Health Accountability and Portability Act (HIPAA) and the HI-TECH Act.
Restricted Confidential	Datasets for which the originating agency has determined that unauthorized disclosure could potentially cause major damage or injury, including death, to residents, agency workforce members, clients, partners, stakeholders, or others identified in the information, or otherwise significantly impair the ability of the agency to perform its statutory functions. Includes any dataset designated by a federal agency at the level "Confidential" or higher under the federal government's system for marking classified information.

A4. Bureau Codes

Portland Open Data uses Cost Objects to extract bureau, divisions and offices and programs codes. This is a top level code list of Bureau Codes as December, 2020

BUREAU CODE	DESCRIPTION BUREAU OR BUSINESS AREA
AT	Office of the City Attorney
AU	Office of the City Auditor
BO	City Budget Office
CB	Office for Community Technology
DR	Bureau of Fire & Police Disability & Retirement
DS	Bureau of Development Services
EC	Bureau of Emergency Communications
EM	Portland Bureau of Emergency Management
ES	Bureau of Environmental Services
FM	Fund & Debt Management
FR	Portland Fire & Rescue
GR	Office of Government Relations
HC	Portland Housing Bureau
HN	Office of Human Relations
MF	Office of Management & Finance
MY	Office of the Mayor
NI	Office of Community and Civic Life
OE	Office of Equity & Human Rights
PA	Commissioner of Public Affairs
PK	Portland Parks & Recreation
PL	Portland Police Bureau
PN	Bureau of Planning & Sustainability
PS	Commissioner of Public Safety

PU	Commissioner of Public Utilities
PW	Commissioner of Public Works
SA	Special Appropriations
SD	Office of Sustainable Development
TR	Portland Bureau of Transportation
WA	Portland Water Bureau
XX	Archive
ZD	Prosper Portland

APPENDIX B - GLOSSARY

Aggregation: Process in which information is gathered and expressed in a summary form, for purposes such as statistical analysis or anonymization. Information summed over a large population is typically free of privacy implications, so aggregation can be used to mitigate privacy concerns.

Application Programming Interface (API): Provides other products and services universal access to our data on Oregon's Open Data Portal. It allows developers to use our data for the creation of applications or other products.

City Agency - Any city bureau or office that manages data on behalf of the city. Internal programs are not agencies.

Comma Separated Value (CSV) File: Used for the digital storage of data structured in a tabular form. Each line of the CSV file corresponds to a row in the table. Within a line, fields are separated by commas and each field belongs to one table column. A CSV file is used to move tabular data between different computer programs.

Data: Data means final versions of statistical or factual information, including statistical or factual data about image files that:

A) Is in alphanumeric form reflected in a list, table, graph, chart or other non narrative form that can be digitally transmitted or processed
B) Is controlled by and regularly created or maintained by, or on behalf of, a state agency
C) And records a measurement, transaction, or determination related to the mission of the agency

Data Business Rules - Set of statements that define or constrain an aspect of data processing. These rules are intended to asset business structure or to control or influence the behavior of the business.

Data Catalog - A data catalog belongs to a database instance and is composed of metadata containing database object definitions like base tables, synonyms, views or synonyms and indexes.

Data Champion - A group of data advocates in City Bureaus and Agencies, appointed internally or shared among two or more agencies, with the following tasks:

(1) To promote data management best practices in Bureaus and other City agencies.
(2) Answer or coordinate research to answer questions about the collection, use, sharing, security and access controls for data that is gathered using a technology or program in the Bureau.
(3) Support bureaus and offices to compile documentation, policies, standards to assure bureau data management.
(4) Perform Data Assessments through the open data submission process.
(5) Support for general data preparation and consulting on data management issues.
(6) Provide bureau training on specific bureau data management needs.
(7) Support on Data compliance of existing standards, laws and regulations.

Data Dictionary - A description of data in business terms including other information needed to use the data (for instance, data types, details of the structures, security restrictions, etc.).Often the content of a data dictionary comes directly from the logical data model.

Data Profiling - it is the process for enriching data with contextual information, including its structure, validation, creation, restrictions, relationship with other data, and how to operate with it. The requirement to profile data must be balanced with the City's data quality, security and privacy regulations.

Data Quality - It refers both to the characteristics associated with high quality data and to the process used to measure or improve the quality of data. Data quality can be defined by dimensions of quality or characteristics that are important to business processes and measurable features important to data consumers. Common dimensions of quality are:

(1) Accuracy - The degree that data correctly represents 'real-life' entities.
(2) Completeness - Whether or not all required data is present.
(3) Consistency - Assurance that data values are consistently represented within a data set and between data sets, and consistently associated across data sets.
(4) Integrity - Also known as coherence, it refers to the consistency between data objects via a reference key contained in data objects, or the internal consistency within a data set such that there are no voids or missing parts.
(5) Reasonability - Whether a data pattern meets expectations.
(6) Timeliness - Refers to how frequent data is likely to change and for what reasons; while data values are the most up-to-date.

(7) Uniqueness - Refers to the state where data objects are unique and not duplicated within the data set.
(8) Validity - Refers to whether data values are consistent with a defined domain of values.

Data Retention Time - Refers to how long data is kept available. For Data Retention Schedules look at the Portland Archives website: https://www.portlandoregon.gov/archives/69741

Dataset: a named collection of related records, maintained on a storage device, that contains data organized, formatted or structured in a specific or prescribed way. The most basic representation of a dataset is data elements presented in tabular form. A dataset may also present information in a variety of non-tabular formats, such as an extended mark-up language (XML) file, a geospatial data file, or an image file.

High-Value Data: Data qualifies as high-value if it can be used to increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation.

Licensing - Refers to the agreement to use or allow use of data sets. It includes the set of permissions derived from intellectual property rights by third-parties for using, reusing and redistributing data. The city allows data with minimum or no restrictions for most of the data produced by it.

Linked Data - Structured data which is interlinked with other data so it becomes more useful as it is properly contextualized and connected to other data sets.

Machine-Readable: Refers to data that can be easily processed by a computer without losing any semantic meaning.

Metadata: Describes characteristics about the data such as the title, description, and keywords (data about data). Metadata facilitates a common language when discussing a dataset's attributes.

Open Data: Open data can be freely used, modified, and shared by anyone for any purpose. Open Data has the following general features:

(1) Availability and Access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
(2) Re-use and Redistribution: the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
(3) Universal Participation: everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

Personally Identifiable Information (PII): Information that can be used by itself or in conjunction with other information to identify an individual.

Priority: Prioritize datasets based on their organizational and/or public value, the quality of the data, the limitations of the tools and processes, alignment with the City of Portland's Data Strategy and Agency Strategic Plans, and demand for this data from other agencies, government entities, or the public.

Publishable Data: any and all data and datasets collected by a state agency, excluding:

(A) Data to which a city bureau or office may deny access pursuant to any provision of a federal, state or local law, rule or regulation, or another applicable policy or restriction.
(B) Data that contains a significant amount of information to which a state agency may deny access pursuant to any provision of a federal, state or local law, rule or regulation.
(C) Data that reflect the internal deliberative process of a city bureau or office, including but not limited to negotiating positions, future procurements or pending or reasonably anticipated legal or administrative proceedings.
(D) Data stored on a personal computing device owned by a city bureau or office, or data stored on a portion of a network that has been exclusively assigned to a single city employee or to a single computing device owned or controlled by the City.
(E) Materials subject to copyright, patent, trademark, confidentiality agreements or trade secret protection.
(F) Materials that have commercial value or the disclosure of which could reduce a state agency's competitive advantage.

(G) Proprietary applications, computer code, software, operating systems and similar materials.
(H) Employment records, internal employee directories or lists, facilities data, information technology and other data related to internal state agency administration.
(I) Any other data the publication of which is prohibited by law.

Quality: Determine the quality of datasets by noting their adherence to data standards, detail of the metadata, the completeness and accuracy of the file, and whether they are in an open format. The standard often used to measure open format is the Five Stars of Open Data rating system.

Readiness: Data readiness is defined by the level of effort required to publish data as an open dataset. Readiness factors include the ability of the system to export source data, the technical debt associated with publishing a dataset, and other factors that impact the level of work required to produce an open dataset.

Tabular Data: Structured data that exists in a table format, with rows and columns. Datasets on data.oregon.gov are displayed in a tabular format

APPENDIX C. MAPPING METADATA FIELDS TO GEODATABASE ISO 19115

Source https://resources.data.gov/resources/podm-field-mapping/

POD v1.1	**ISO XPath**	**Sample**
identifier	CGIS? DOI? our URI?	ODPilot100102
accessLevel	gmd:resourceConstraints/gmd:MD_SecurityConstraints/gmd:classification/gmd:MD_ClassificationCode	public
contactPoint {fn, hasEmail}	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:pointOfContact/gmd:CI_ResponsibleParty/gmd:individualName //gmd:CI_ResponsibleParty/gmd:contactInfo/gmd:CI_Contact/gmd:address/gmd:CI_Address/gmd:electronicMailAddress	"fn":"Mary Koo", "hasEmail":"mary.koo@portlandoregon.gov"
description	**//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:abstract/gco:CharacterString**	Received 911 calls received at the emergency call center
title	**//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:title/gco:CharacterString**	BOEC - April2018 - Filtered-In 911 calls
dcat	*	Dataset
keyword	**//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:descriptiveKeywords/gmd:MD_Keywords/gmd:keyword (gco:CharacterString or gmx:Anchor)**	["Emergency","911"]
modified	**//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date/gmd:CI_Date/gmd:dateType/gmd:CI_DateTypeCode == "revision" + gmi:MI_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date/gmd:CI_Date/gmd:date/gco:Date**	2020-11-05
publisher	CI_Citation/gmd:citedResponsibleParty/gmd:CI_ResponsibleParty/gmd:organisationName/gmd:organisationName	BOEC
license	*	https://creativecommons.org/licenses/by/4.0/

rights	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:resourceConstraints/gmd:MD_LegalConstraints/gmd:accessConstraints/gmd:MD_RestrictionCode	This dataset is available as is, it does not contain private or sensitive information.
temporal	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:temporalElement/gmd:EX_TemporalExtent/gmd:extent/gml:TimePeriod/gml:beginPosition + //gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:temporalElement/gmd:EX_TemporalExtent/gmd:extent/gml:TimePeriod/gml:endPosition	2018-04-01T00:00:00Z/2018-05-01T00:00:00Z
accrualPeriodicity	/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:resourceMaintenance/gmd:MD_MaintenanceInformation/gmd:maintenanceAndUpdateFrequency/gmd:MD_MaintenanceFrequencyCode	R/P1M
issued	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date/gmd:CI_Date/gmd:dateType/gmd:CI_DateTypeCode == publication + //gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date/gmd:CI_Date/gmd:date (gco:Date or gco:DateTime)	2020-11-05
language	**//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:language**	["En-US"]
landingPage	//gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/gmd:function/gmd:CI_OnLineFunctionCode == information + //gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/linkage/URL	https://pdx-open-data-open-pdx.hub.arcgis.com
conformsTo	*	Data Standard
describedBy	//gmd:aggregationInfo/gmd:MD_AggregateInformation/gmd:aggregateDataSetName/gmd:CI_Citation/gmd:citedResponsibleParty/gmd:CI_ResponsibleParty/gmd:contactInfo/gmd:CI_Contact/gmd:onlineResource/gmd:CI_OnlineResource/gmd:CI_OnlineResource/gmd:linkage/gmd:URL WHERE: //gmd:aggregationInfo/gmd:MD_AggregateInformation/gmd:initiativeType/gmd:DS_InitiativeTypeCode == dataDictionary	Data Dictionary
describedByType	*	Data Dictionary Type

isPartOf	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:aggregationInfo/gmd:MD_AggregateInformation/gmd:aggregateDataSetIdentifier/gmd:MD_Identifier/gmd:code (gco:CharacterString or gmx:Anchor) + //gmd:identificationInfo/gmd:MD_DataIdentification/gmd:aggregationInfo/gmd:MD_AggregateInformation/gmd:associationType/gmd:DS_AssociationTypeCode == largerWorkCitation	Collection
issued	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date/gmd:CI_Date/gmd:dateType/gmd:CI_DateTypeCode == publication + //gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:date/gmd:CI_Date/gmd:date (gco:Date or gco:DateTime)	Release Date
references	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:aggregationInfo/gmd:MD_AggregateInformation/gmd:associationType/gmd:DS_AssociationTypeCode == crossreference + //gmd:identificationInfo/gmd:MD_DataIdentification/gmd:aggregationInfo/gmd:MD_AggregateInformation/gmd:aggregateDataSetName/gmd:CI_Citation/gmd:citedResponsibleParty/gmd:CI_ResponsibleParty/gmd:contactInfo/gmd:CI_Contact/gmd:onlineResource/gmd:CI_OnlineResource/gmd:linkage/gmd:URL	Related Documents
theme	//gmd:identificationInfo/gmd:MD_DataIdentification/gmd:topicCategory/gmd:MD_TopicCategoryCode	Theme
	distribution	
accessURL	//gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/gmd:function/gmd:CI_OnLineFunctionCode == information, search, order or offlineAccess + //gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/linkage/URL	
conformsTo	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributionFormat/gmd:MD_Format/gmd:specification	
describedBy	//gmd:contentInfo/gmd:MD_FeatureCatalogueDescription/featureCatalogueCitation/CI_Citation/citedResponsibleParty/CI_ResponsibleParty/contactInfo/CI_Contact/onlineResource/CI_OnlineResource/lin	https://project-open-data.cio.gov/v1.1/schema/catalog.json

	kage/URL	
describedByType	*	application/json
description	//gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/gmd:functiongmd:CI_OnLineFunctionCode == download + //gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/gmd:description	Distribution with description
downloadURL	//gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/gmd:function/gmd:CI_OnLineFunctionCode == download + //gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/linkage/URL	distribution url with ‘download’ function
format	//gmd:distributionInfo/gmd:MD_Distribution/gmd:distributionFormat/gmd:MD_Format/name/gco:CharacterString	Distribution format name
mediaType	*	
title	//gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/gmd:function/gmd:CI_OnLineFunctionCode == download + //gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/gmd:name/gco:CharacterString	distribution + name

APPENDIX D. DATA MANIFEST

Portland Open Data relies on documenting the data workflow control in structured files called Data Manifest. Data Manifest has the following tree structure:

```
-/control
        /datastrategy.json
        /control.json
 /data
        /data.json
        /metadata.json
        /transformation.json (optional)
        /miscellaneous.json (optional)
 /descriptor
        /dictionary.json
        /mapping.json (optional)
        /documentation.json (optional)
```

D1. Data Control

The section dedicated to the control of the data workflow is constituted by two descriptive files:

a) datastrategy.json - describing the tasks during data curation and clearance requirements for publication.
b) control.json - file that describes and controls the data workflow from start to publication.

D1.1 Datastrategy.json

The file datastrategy.json is a canonical machine-readable schema for describing action items within a government agency's digital strategy, and for reporting on its progress. This schema has been adapted from the federal model[12]. These files can enable automation, performance measures and audits[13].

datastrategy.json will describe the data curation process that the City or a Bureau follows in the publication of an open dataset.

[12] https://github.com/GSA/digital-strategy
[13] https://labs.data.gov/dashboard/docs/main#automated_metrics

This schema is compiled by two sections:

1. agencies.json - machine-readable listing of city bureaus, their primary domain, and abbreviation (e.g., Portland Housing Bureau)
2. items.json - machine-readable representation of the action items from the digital strategy

D1.2. Agency List

The agency list contains a timestamp of when the file was last updated and the schema version as well as a listing of common federal agencies. Each agency has three fields:

- name - The Human-readable name of the agency (e.g., Bureau of Planning and Sustainability)
- id - The agencies abbreviation or id (e.g., PN)
- url - the agency's primary domain (e.g., www.portland.gov/bps/)

In JSON this is represented as:

```
{
   "generated":"2020-11-24 10:46:19",
   "agencies":[
      {
         "name":"Bureau of Planning and Sustainability (BPS)",
         "id":"PN",
         "url":"www.portland.gov/bps/"
      },
      {
         "name":"Portland Bureau of Transportation",
         "id":"TR",
         "url":"www.portland.gov/pbot/"
      },
      ...
   ]
}
```

D1.3 Items

The items act as a machine-readable representation of the agency-specific action items outlined in the digital strategy, as well as a base schema for reporting on its progress. At the root level, the schema contains a timestamp indicating when it was last updated and the schema version, as well as a list of all action items.

Each action item can have the following properties:

- id - a unique identifier for that action item, e.g., 2.1
- parent - where applicable, the parent action item, (e.g., 2.2.1's parent would be 2.1). Useful for grouping and formatting
- text - the human-readable text of the action item
- due - when the action item is due (relative to the release of the digital strategy)
- due_date - date calculated as the absolute due date for the action item
- fields - a list of all fields associated with that action item
- multiple - whether multiple responses are allowed per action item (e.g., listing multiple systems with each of the action-item's field being answered once per system)

The field object is made up the following:

- type - the HTML input type that best represents the field (e.g., select, text, textarea)
- name - HTML friendly name for the field
- label - Human readable label for the field
- option - where applicable, an array of label, value pairs describing the potential options (e.g. for a drop down)
- value - when used as an agency progress report, the agency-reported answer to the field, or if multiple answers, an array of agency-reported answers. Multiple values will be represented as an array in JSON, as nested `value` nodes in XML.

In JSON this would be represented as:

```
{
   "generated":"2020-07-12 11:00:27",
   "items":[
      {
         "id":"2.1",
         "parent":null,
         "text":"Verify that there are not duplicate records or
rows within the dataset.",
         "due":"5 Days",
         "due_date":"2020\/07\/17",
         "fields":[
            {
               "type":"select",
               "name":"2-1-status",
               "label":"Overall Status",
               "options":[
                  {
                     "label":"Not Started",
                     "value":"not-started"
                  },
                  {
                     "label":"In Progress",
                     "value":"in-progress"
                  },
                  {
                     "label":"Completed",
                     "value":"completed"
                  }
               ],
               "value":null
            }
         ],
         "multiple":false
      },
      ...
   ],
}
```

D1.4 Control.json

The fields of the control.json file are:

Field ID	**Field name**	**Field description**
THE CASE TICKET ID SECTION		
CASE_ID	Case ID number	Number of the case for submission
CANDIDATE_DATASET_NAME	Name of candidate dataset	Name or short description of the candidate dataset
DATASET_DESCRIPTI	General description	Longer description of the candidate dataset

ON		

BUREAU AUTHORIZATION FORMS		
BUREAU_NAME	Bureau	Bureau name
CLIENT_NAME	Client name	Name of the bureau data owner
CLIENT_EMAIL	e-mail	email of the bureau data owner
SUBMITTER_NAME	Submitter name	Submitter name
SUBMISSION_DATE	Date of submission	date of submission after preparation
CASE_STATUS	Case status	Case status: Preparation, verification, submission, publication, rejected, published
BUREAU_AUTHORIZATION	Bureau authorized by	Bureau data steward authorizing the submission
DATA_INTEGRITY_PASSED	Data integrity check	Flag for approval of data integrity assessment
SUBMISSION_DATE	Data Submission date	Date of data submission for publication and permanent storage
PUBLISHED_DATE	Data publication	Date of data gets published and available at the Open Data Portal

DATA ASSESSMENT REPORTS		Array of assessments done. This section keeps a history of assessments. Failed evaluations can be resubmitted and changed.
ASSESSMENT_TYPE	Assessment type	Data integrity, data privacy, data equity
EVALUATOR_NAME	Name of the evaluator	Name of the person performing the assessment
EVALUATION_DATE	Date of evaluation	Date of evaluation or assessment
EVALUATION_DESCRIPTION	Description of assessment	compilation of business rules assessed in this procedure
ISSUE_IDENTIFIED	Flag identifying an issue	
EVALUATION_NOTES	Results and comments of assessment	Results of the tests of the assessed business rules