OPEN DATA HANDBOOK

DISTRICT OF COLUMBIA GOVERNMENT

December 2018

TABLE OF CONTENTS

TABLE OF CONTENTS	**2**
INTRODUCTION	**3**
BACKGROUND	**3**
DATA SUBMISSION PROCESS	**3**
DATASET RULES WITH EXAMPLES	**9**
METADATA STANDARD WITH EXAMPLES	**14**
APPENDIX A: EDI DATA DICTIONARY	**25**
APPENDIX B: DATA SUBMISSION DATA DICTIONARY	**28**

INTRODUCTION

The Office of the Chief Technology Officer (OCTO), within the District of Columbia (DC) government, manages the District's data program. This includes open data, business intelligence, data curation, and Geographic Information Systems (GIS). The open data handbook explains the process and steps the OCTO data program undertakes when an agency 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 DC's open data portal. The purpose is three-fold, as follows:

1. Provides transparency and accountability for DC's data program
2. Allows agencies and the public to understand the overall data curation process.
3. Documents data policies, rules, requirements, and guidelines for open data consistency and standardization.

BACKGROUND

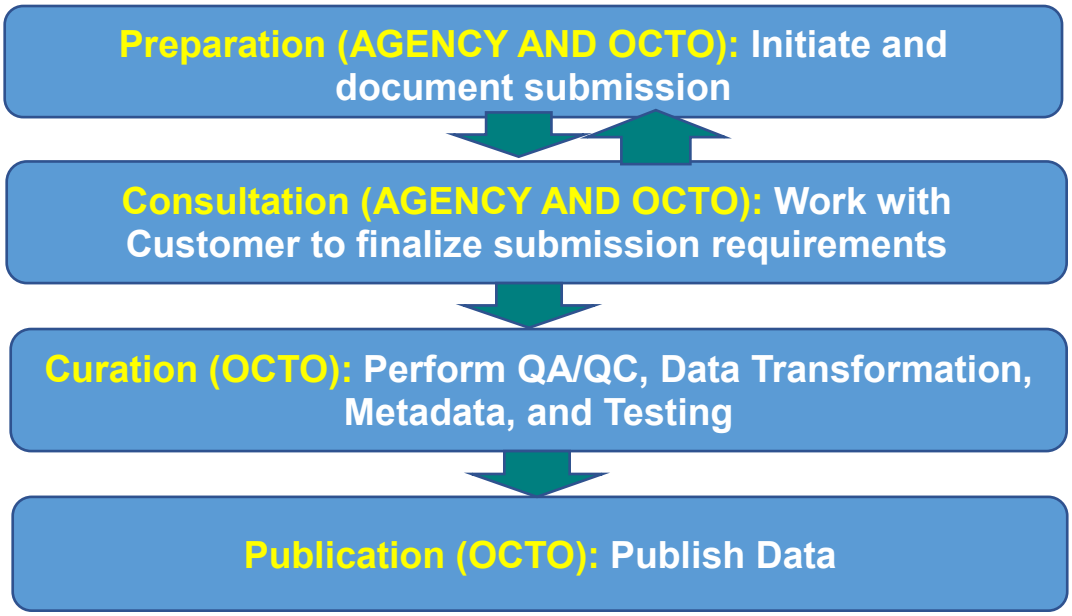
Originally, the District started developing an enterprise data program through its GIS effort. The DC GIS program used a federated model to create a common set of procedures, standards, and services empowering agencies to participate in the data effort in a collective framework. Several years ago, the program expanded to include and support date warehousing, big data, business intelligence, and the open data portal (http://opendata.dc.gov).

DC issued an open data policy (https://octo.dc.gov/page/district-columbia-data-policy) effective April 27, 2017. The policy's goal is to develop guidelines, governance, and a framework to maximize the District's investment and benefit in data. One of the key products is an enterprise data catalog (EDI) which is a comprehensive data catalog of District-wide data sets that will drive a lot of the resulting data collection and submission efforts. For more information, go to http://opendata.dc.gov/pages/cdo-annual-report.

For this guide, OCTO researched exemplary and innovative open data programs in other jurisdictions (New York, New York City, Seattle, Maryland, and San Francisco). This was to discover and apply lessons learned from other's open data standards and strategies to DC's. In particular, New York State and San Francisco had comprehensive information on data standards, rules, process, documentation, and application development that directly helped develop this handbook.

DATA SUBMISSION PROCESS

DC's data submission process involves four steps as generally summarized below (a more detailed flow diagram is presented at the end of the Submission section):



Overall, the data analysts guide the data owner and other analysts as needed to run the data through the submission process, with different groups leading each part (see all caps in diagram above). Their application occurs in a unified database infrastructure consisting of a data warehouse and geospatial database. While there are specific processes and guidelines for each database, they share an overall setting where consistency and standardization are promoted and supported for their individual data curation processes.

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

Preparation

The service ticket will then get assigned to a data analyst within the OCTO data curation program. Once assigned, the service ticket can be closed. That analyst will work with the data owner or representative and other resources to start constructing a Data Submission Packet (DSP). A DSP consists mainly of the following components:

Component	What is it?	Why is it important?
Service Ticket	A service request that initiates the data submission process. It contains the requestor information and type of action.	This part documents the request and lets the data curation team know to start the process. Having a ticket in the system promotes accountability and the opportunity to develop metrics as a potential performance measure.
Data Submission Form	The form records information (see Appendix B for the data dictionary) and specifications for the submittal. It is broken into data, metadata, transformation, contact, and miscellaneous	This part allows data curation staff to perform the data curation process fluidly. By having the needed information, staff can independently and efficiently curate the data.

	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.	
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 descriptors for any domains.	This part permits the database administrator to define and create the required tables for the data.
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.	This part makes creating the ETL process a lot easier and documented. The ETL process can be one of the most cumbersome parts of the curation process.
Data Review Summary Report	Each dataset gets compared to the overall dataset rules (see Appendix C) and relevant requirements in the data submission form. A report is printed out of the results. This will be a major development effort next FY.	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.
Additional Documentation	Additional documentation may include a survey instrument, data collection tool, study or report specific to the data, explanatory documentation for complex datasets, and other relevant documents.	Additional 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.

Together, these components in a DSP allow the data analyst to gain an in-depth knowledge of the dataset before working directly with the data owner to refine and finalize the requirements. They will work with the data owner to ensure the EDI information is properly linked to the submission requirements. The next step in the data submission process is consultation.

Consultation

Consultation involves reviewing and finalizing the DSP imnformation, specifications, and requirements. First, if needed, the service ticket is examined for content. Next, the data submission form is thoroughly reviewed for content. Then the data dictionary is evaluated against the source data, including domain values. 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. Finally, the data summarization report is examined for potential transformations that ensure data meets rules.

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.

Curation

Before the data goes through the curation process, the agency receives an email "receipt" that states OCTO received the data and provides an approximate date for publication. This allows agencies to document submission for management and performance purposes.

This stage involves staff performing QA/QC, devising data transformation workflows, developing metadata, as well as cataloging and testing the dataset. The general steps are:

- Test connectivity between the destination data curation database(s) and source databases.
- Analyze and review the source database schema(s).
- Develop the data mappings, objects, views, and other related entities in the destination database.
- Create database schema, tables, views, indexes, and other related entities.
- Develop the data transformation workflows between the two data containers.
- Move database elements and data transformation workflows to testing environment.
- Test workflows and data with the application or specified use case.
- Catalog dataset if needed in EDI
- Create metadata

There will need to be periodic review submitted datasets as structure changes and technology enhancements occur. That review will go through the same process as a separate submission using information from the first submission.

Submitted datasets may also become obsolete or dated where archiving will be an option. OCTO will review archiving options with the data owner. If authorized, the dataset will be included in the data program's archive with a note added to the metadata.

Publication

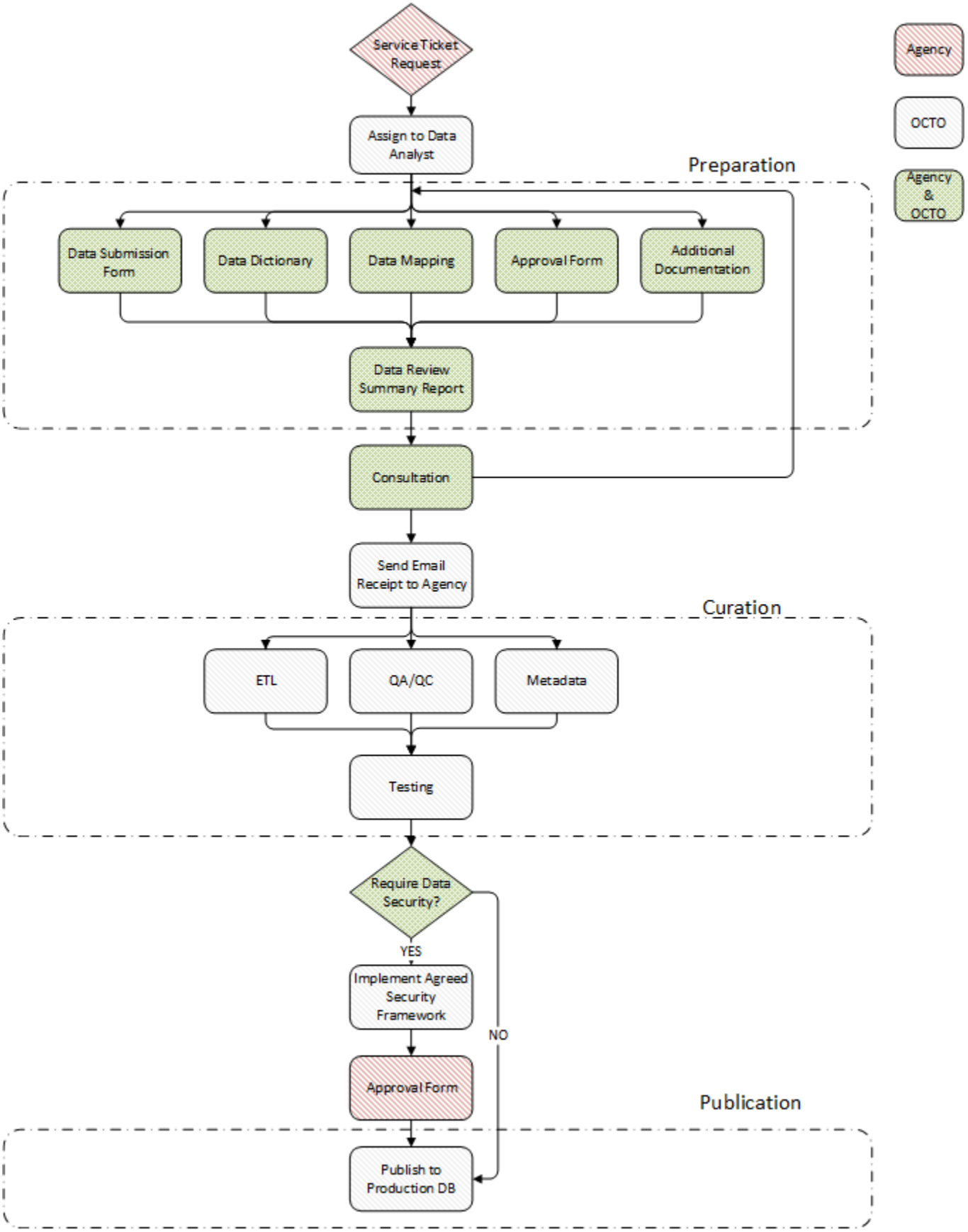
After curation, datasets are moved to production systems, such as map services, business intelligence (BI), open data, and other applications. For open data and geospatial applications, the

data curation team publishes the data to map services. This uses the geography of the data for optimal and maximized GIS operations. This process is:

1. Identify what map service the layer belongs in unless it's part of a new map service.
2. Create a layerfile that sets the symbology, labels, alias names, definition query if needed, and other parameters.
3. Add the layerfile to the map document defining the map service.
4. Publish map service to test environment
5. If it passes QA/QC, publish the map service to production and failover environments.
6. Add to DC's open data portal (http://opendata.dc.gov)

This process also applies to tabular and select imagery data. Once the dataset is published, this completes the submission process. It is available to production systems, open data, and other applications.

Data Submission Workflow



DATASET RULES WITH EXAMPLES

Standardized and consistent data is important for efficient and uniform analysis, minimizing data anomalies due to formatting, structure, and other factors. The rules are:

<table>
<tr><th>1.1</th><th>Data Format</th><th></th></tr>
<tr><th>Rule</th><th>Description</th><th>Examples</th></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>COMPANY</td><td>AIRTIME</td><td></td><td></td></tr>
<tr><td>10/2/2017</td><td></td><td>AAA</td><td>BBB</td><td>CCC</td></tr>
<tr><td>Music</td><td>A Network</td><td>0:20:30</td><td></td><td></td></tr>
<tr><td>Missing</td><td>B Network</td><td>0:01:00</td><td></td><td></td></tr>
<tr><td>Sport Event - LIVE</td><td>C Network</td><td>0:40:00</td><td></td><td></td></tr>
<tr><td>Teen Kids</td><td>A Network</td><td></td><td>0:23:04</td><td></td></tr>
</table>
<table>
<tr><td>DATE</td><td>TITLE</td><td>COMPANY</td><td>AIRTIME</td><td>NETWORK</td></tr>
<tr><td>10/2/2017</td><td>Music</td><td>A Network</td><td>0:20:30</td><td>AAA</td></tr>
<tr><td>10/2/2017</td><td>Missing</td><td>B Network</td><td>0:01:00</td><td>AAA</td></tr>
<tr><td>10/2/2017</td><td>Sport Event - LIVE</td><td>C Network</td><td>0:40:00</td><td>AAA</td></tr>
<tr><td>10/2/2017</td><td>Teen Kids</td><td>A Network</td><td>0:23:04</td><td>BBB</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.</td><td>
<table>
<tr><td>1st QUARTER REPORT</td><td></td><td></td><td></td><td></td></tr>
<tr><td>FY 2018</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>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>
<tr><td>Summarized Data</td><td>Avoid including subtotal and total of values in cells as part of the column. Typically, applications can compute these values and having totals or subtotals skews results.</td><td>
<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>Total</td><td></td><td>0:43:34</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>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td></td><td>
<table>
<tr><td>Teen Kids</td><td>A Network</td><td>0:23:04</td></tr>
</table>
</td></tr>
<tr><td>Grouped Data</td><td>Avoid grouping data in cells as part of the column with headers. Do not separate columns using header titles.</td><td>
<table>
<tr><td>SCHOOL</td><td>TEST SCORE</td><td>ADS</td></tr>
<tr><td>NE</td><td></td><td></td></tr>
<tr><td>A Middle School</td><td>80</td><td>2</td></tr>
<tr><td>B High School</td><td>82</td><td>6</td></tr>
<tr><td>NW</td><td></td><td></td></tr>
<tr><td>A High School</td><td>85</td><td>1</td></tr>
<tr><td>B Middle School</td><td>77</td><td>4</td></tr>
</table>
<table>
<tr><td>SCHOOL</td><td>TEST SCORE</td><td>ADS</td></tr>
<tr><td>A Middle School</td><td>80</td><td>2</td></tr>
<tr><td>B High School</td><td>82</td><td>6</td></tr>
<tr><td>A High School</td><td></td><td></td></tr>
<tr><td>B Middle School</td><td>85</td><td>1</td></tr>
</table>
</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>Text values should be all upper case, lower case, or initial caps.</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>
<tr><td>1.2</td><td>Data Field</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. Underscores separate words. Avoid use of abbreviations. Instead, use title case for field names and be sure that the names match that in the Data Dictionary. Aliases reflect real-world understandable, plain English, names limited to 30 characters, initcaps words and use spaces to separate words.</td><td>
<table>
<tr><td>Seats Offered</td><td>Seats Matched</td><td>MtcRate</td><td>Matched: Early Action, Dual Language Guarantee, or Transfer</td><td>Matched: In-boundary with Sibling Enrolled</td></tr>
<tr><td>37</td><td>9</td><td>24%</td><td>1</td><td>0</td></tr>
</table>
<table>
<tr><td>SEATS_OFFERED</td><td>SEATS_MATCHED</td><td>MATCH_RATE</td><td>EARLY_DUAL_LANG_GUAR_TRANS</td><td>INBOUND_SIBLING_ENROLLED</td></tr>
<tr><td>37</td><td>9</td><td>24%</td><td>1</td><td>0</td></tr>
</table>
</td></tr>
<tr><td>Leading or Trailing Spaces</td><td>Text fields must be trimmed of leading or trailing space(s).</td><td></td></tr>
</table>

<table>
<tr>
<td>Numeric Field Values</td>
<td>Do not mix text in a field that is intended to contain numeric or date data.</td>
<td>
<table>
<tr><td>PROJECT</td><td>PROGRESS</td></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><td>PROJECT</td><td>PROGRESS</td></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><td>PROJECT</td><td>COST</td></tr>
<tr><td>A Construction</td><td>$500,500.223</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><td>PROJECT</td><td>COST</td></tr>
<tr><td>A Construction</td><td>500500.22</td></tr>
<tr><td>B Highway</td><td>55250000.00</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><td>NAME</td><td>KEY_ INDEX</td></tr>
<tr><td>A AGENCY</td><td>Negative 50</td></tr>
<tr><td>B AGENCY</td><td>(10)</td></tr>
</table>
<table>
<tr><td>NAME</td><td>KEY_INDEX</td></tr>
<tr><td>A AGENCY</td><td>-50</td></tr>
<tr><td>B AGENCY</td><td>-10</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, column must be assigned as text format preventing the loss of leading zeroes.</td>
<td>
<table>
<tr><td>SIC</td><td>INDUSTRY</td></tr>
<tr><td>0191</td><td>General Farms, Primarily Crop</td></tr>
<tr><td>1521</td><td>General Contractors-Single-Family Houses</td></tr>
</table>
</td>
</tr>
<tr>
<td>Date and Time Fields</td>
<td>Date fields should be in MM/DD/YYYY format. Time should be stored as military (i.e. 24-hour time) in HH:MM:SS or HH:MM format. 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 EST.</td>
<td>
<table>
<tr><td>UPDATED_DATE</td><td>UPDATED_TIME</td></tr>
<tr><td>09/15/2018</td><td>10:15:20</td></tr>
<tr><td>09/15/2018</td><td>16:25:45</td></tr>
</table>
OR
<table>
<tr><td>UPDATED_DATE</td><td>UPDATED_TIME</td></tr>
<tr><td>09/15/2018</td><td>10:15 AM</td></tr>
<tr><td>09/15/2018</td><td>04:25 PM</td></tr>
</table>
</td>
</tr>
</table>

<table>
<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><td>NAME</td><td>ZIPCODE</td></tr>
<tr><td>A Building</td><td>20001</td></tr>
<tr><td>B Building</td><td>20003</td></tr>
</table>
<table>
<tr><td>NAME</td><td>ZIPCODE</td></tr>
<tr><td>A Building</td><td>20001-0112</td></tr>
<tr><td>B Building</td><td>20003-2214</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><td>NAME</td><td>PHONE</td></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><td>NAME</td><td>PHONE</td></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>
<tr><td>Unique Identifier</td><td>For Open Date and GIS publication, OBJECTID (auto generated sequential number) will be added. OBJECTID <u>will not</u> 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 database as the primary key.</td><td>
<table>
<tr><td>EDI_ID</td><td>GIS_ID</td><td>Object ID</td><td>SERVICE_ID</td></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.</td><td>
<table>
<tr><td>NAME</td><td>URL</td></tr>
<tr><td>OCTO</td><td>https://octo.dc.gov</td></tr>
<tr><td>DDOT</td><td>https://ddot.dc.go</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><td>EMAIL</td></tr>
<tr><td>John Doe@example.com</td></tr>
<tr><td>John Doe at dc.gov</td></tr>
<tr><td>Jane Doe@ dc.gov</td></tr>
</table>
<table>
<tr><td>EMAIL</td></tr>
<tr><td>JohnDoe@example.com</td></tr>
<tr><td>Jane.Doe@dc.gov</td></tr>
</table>
</td></tr>
<tr><td>1.3</td><td>Addressing</td><td></td></tr>
</table>

<table>
<tr><td>Address Data</td><td>Address data will always be run through the DC 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>200 I STREET SE</td><td>308596</td><td>399755.50</td><td>134641.60</td><td>38.87960769</td><td>-77.00281784</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 field called ADDRESS and then geocoded against the MAR.</td><td>
<table>
<tr><th>NAME</th><th>STREENUMBER</th><th>STREET</th><th>QUADRANT</th></tr>
<tr><td>OCTO</td><td>200</td><td>I ST</td><td>SE</td></tr>
<tr><td>DDOT</td><td>55</td><td>M ST</td><td>SE</td></tr>
</table>
<table>
<tr><th>NAME</th><th>ADDRESS</th></tr>
<tr><td>OCTO</td><td>200 I ST SE</td></tr>
<tr><td>DDOT</td><td>55 M ST SE</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>OCTO</td><td>200 I ST SE, Washington, DC 20003</td></tr>
<tr><td>DDOT</td><td>55 M ST SE, Washington, DC 20003</td></tr>
</table>
<table>
<tr><th>NAME</th><th>ADDRESS</th><th>CITY</th><th>STATE</th><th>ZIPCODE</th></tr>
<tr><td>OCTO</td><td>200 I ST SE</td><td>Washington</td><td>DC</td><td>20003</td></tr>
<tr><td>DDOT</td><td>55 M ST SE</td><td>Washington</td><td>DC</td><td>20003</td></tr>
</table>
</td></tr>
<tr><td>1.4</td><td>Miscellaneous</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>
</table>

Topology	Based on data, topology will be checked. 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 EDI and metadata (data about the data). It allows users to fully understand the data content and context, including caveats and data limitations. DC's metadata standard with an example is below:

Required Element	Required elements must be provided
Recommended Element	Metadata authors are strongly encouraged to fill in recommended elements.
Optional Element	Optional elements are proved at the discretion of the agency

1.1	**Item Description**	
Element	**Description**	**Examples**
Title	The name by which the data set is known.	Zoning
Tags	Set of terms that can be used to search for resource. Terms should be provided as a comma-separated list	zoning, land use, planning, districts
Description (Abstract)	A summary of what is in the data set.	Zoning. The dataset contains polygons representing zoning districts, as submitted to the DC Office of Zoning, created for the DC Office of Zoning (DCOZ). Zoning districts were identified from official adopted zoning and unofficial zoning documents and heads-up digitized from the 1995 orthophotographs. All data is stored and exported in Maryland State Plane coordinates NAD 83 meters. METADATA CONTENT IS IN PROCESS OF VALIDATION AND SUBJECT TO CHANGE
Summary (Purpose)	Why the data set was developed.	This data is used for the planning and management of Washington, D.C. by local government agencies.
Dataset Credit	Organizations credited with the data set, part of the creation of the data set; one organization per line.	District of Columbia Office of Zoning

Use Constraints	Restrictions and legal prerequisites for using the data set after access is granted (ex. transferring to others). This depends on the level of access for the data set.	This work is licensed under a Creative Commons Attribution 4.0 International License.
1.2	**Topics & Keywords**	
Topic Categories	The primary themes associated with the resources' content: • Administrative and Other Boundaries • Aerial Photography and Scanned Maps • Basemap • Business and Economic • Development • Communication • Cultural and Society • Demographic • Education • Elevation • Environment • Facility and Structure • Financial • Government Operations • Health • Historic • Location • Planning Land Use and Zoning • Property and Land • Public Safety • Public Services • Recreation • Technology • Transportation • Utility and Communication	Planning & Cadastral
Place Keywords	Words or phrases summarizing an aspect of the place of the data set.	Washington, D.C.; District of Columbia; DC; DC GIS; DCGIS
Theme Keywords	Words or phrases summarizing an aspect of the theme of the data set.	Zoning Planning Districts Land use regulation Residence districts Special purpose districts Mixed use districts Commercial districts

		Industrial districts
1.3	**Citation**	
Created Date	As part of the citation information, the creation date of the data set. This is formatted to “yyyy-mm-dd.”	5/6/2018
Publication Date	As part of the citation information, the publication date of the data set. This is formatted to “yyyy-mm-dd.	5/6/2018
Revised Date	As part of the citation information, the revised date of the data set. This is formatted to “yyyy-mm-dd.”	5/6/2018
2.1	**Resource Details**	
Status	Development status as reflected in domain choices: • Completed • Historical Archive • Obsolete • On Going • Planned • Required • Under Development	On Going
Supplemental Information	Any supplemental information that further describes or applies to the dataset.	The long title of the Zoning Regulations, as adopted, shall be as follows: REGULATIONS CONTROLLING AND RESTRICTING THE HEIGHT, BULK, NUMBER OF STORIES, AND SIZE OF BUILDINGS AND OTHER STRUCTURES, THE OPEN SPACES AROUND THEM, THE DENSITY OF POPULATION, AND THE USES OF BUILDINGS, STRUCTURES, AND LAND IN THE DISTRICT OF COLUMBIA, AND FOR SAID PURPOSES DIVIDING THE DISTRICT OF COLUMBIA INTO ZONING DISTRICTS. The regulations set forth in this title shall be known and may be cited by the short form title of the 'Zoning Regulations of the District of Columbia.' Unless otherwise noted, the authority for this chapter is §1 of An Act providing for the Zoning of the District of Columbia and the regulation of the location, height, bulk and uses of buildings and other structures and of uses of land in the District of Columbia, and for other purposes, approved June 20,1938 (52 Stat 797), as amended by An Act approved March 4, 1942, P.L. V-468, (56 Stat 122); and by 55492 of the District of Columbia Self Government and Governmental Reorganization Act, P.L. 93.198, 87 Stat. 810. SOURCE: §1101.1, 1102.1 and 9301.1 of Regulations effective May 12, 1958. INTERPRETATION AND APPLICATION: In

		their interpretation and application, the provisions of this title shall be held to be the minimum requirements adopted for the promotion of the public health, safety, morals, convenience, order, prosperity, and general welfare for the following purposes: (a) To provide adequate light and air; (b) To prevent undue concentration of population and the overcrowding of land; and (c) To provide distribution of population, business and industry, and use of land that will tend to create conditions favorable to transportation, protection of property, civic activity, and recreational, educational, and cultural opportunities; and that will tend to further economy and efficiency in the supply of public services. The regulations set forth in this title and the zoning maps are designed with consideration of the following: (a) The character of the respective districts; (b) The suitability of each district for the uses permitted in each district under this title; and (c) The encouragement of the stability of districts and of land values in those districts. The provisions of this title shall govern whenever the regulations in this title do the following: (a) Require larger yards, courts, or other open spaces; b) Require a lower height or bulk of buildings or a smaller number of stories; (c) Require a greater percentage of lot to be unoccupied; or (d) Impose other higher standards than are required in or under any statute or by any other municipal regulations.
2.2	**Resource Extent**	
Time Period of Content Date	A single date for which the data set corresponds to the ground or for which the user can be confident of accuracy and completeness. This date is typically set to the publication date or the date of last update. It is formatted to “yyyy-mm-dd.”	12/31/2017
Currentness Reference	The basis on which the Time Period of Content Date is determined. This is typically set to the publication date of the dataset.	publication date
2.3	**Resource Contacts**	
Contact Person	A person within the development or managing group who can receive inquiries regarding the content or development of the data set. This may not be the same as the distributor.	Nyambi Nyambi

Contact Organization	The name of the organization that can receive inquiries regarding the content or development of the data set.	DC Office of Zoning
Contact Person Position	The position of the Point of Contact Person.	
Contact Role	The association between the responsible party and the resource: • Resource Provider • Custodian • Owner • User • Distributor • Originator • Point of Contact • Processor • Publisher • Author	Point of Contact
Contact Address	The mailing address of the Contact Person. Ensure that the address type is set to the appropriate choice – mailing or physical address, or both.	441 4th St NW Suite 210 South
Contact City	The city of the Contact Person.	Washington
Contact State or Province	The state or province of the Contact Person.	DC
Contact Postal Code	The ZIP or other postal code of the Contact Person.	20001
Contact Voice Phone	The voice telephone number, including area code, of the Contact Person.	(202) 727-0314
Contact Fax Phone	The facsimile telephone number, including area code, of the Contact Person.	
Contact E-mail Address	The Internet mail address of the Contact Person.	Nyambi.Nyambi@dc.gov
Contact Person Hours of Service	The hours of service of the Contact Person.	10 am - 5 pm
2.4	**Resource Maintenance**	
Update Frequency	The frequency with which changes and additions are made to the data set after the initial data set is completed.: Continual • Daily • Weekly • Monthly	As Needed

	• Quarterly • Biannually • Annually • As Needed • Irregular • Not Planned • Unknown	
Next Update	The scheduled revision date. This is formatted to “yyyy-mm-dd.”	5/6/2019
2.5	**Resource Constraints**	
Legal (Access) Constraints:	Restrictions or limitations on obtaining the dataset; for example, to assure the protection of privacy or intellectual property: • Copyright • Patent • Patent Pending • Trademark • License • Intellectual Property Rights • Restricted • Other Restrictions	
Security Constraints:	Identify any handling restrictions on the dataset: • Unclassified • Restricted • Confidential • Secret • Top Secret	
2.6	**Spatial Reference**	
	This section describes the spatial reference or projection of the data set. It includes the spatial reference parameters, such as projection name, horizontal datum and units, and ellipsoid. For datasets with a vertical, this section descriptors such as altitude datum and units. ESRI automatically populates this section in ArcCatalog when creating or updating metadata if the spatial reference has been defined for the data set.	The scheme used to define horizontal coordinates as depicted by the domain choices.

2.7	Resource Quality	
Attribute Accuracy	An explanation of how accurately the entities have been identified or how accurately values have been assigned in the data set. This can be the results of quantitative analysis, steps taken to ensure accuracy during development, or known deficiencies. This may be summed up by stating "Dataset believed to be complete and accurate, subject to change." This can include the sentence "In process of validation by source and/or responsible agency" or "Validated by source and/or responsible agency" depending on the status of the dataset validation.	Data believed to be complete and accurate as of the time period cited. Subject to change. Digital version compared with original source data for accuracy and completeness. In process of validation by source and/or responsible agency.
Logical Consistency	An indication of topological problems such as overshoots, undershoots, unwanted intersections, unclosed polygons, missing or duplicate labels, etc. This can include the sentence "In process of validation by source and/or responsible agency" or "Validated by source and/or responsible agency" depending on the status of the dataset validation.	same
Completeness	Information about omissions, selection criteria, generalization, definitions used, and other rules used to derive the data set that may affect the completeness of content of the data. This may include the statement "Data believed to be complete and accurate, subject to change." This can include the sentence "In process of validation by source and/or responsible agency" or "Validated by source and/or responsible agency" depending on the status of the dataset validation.	Data believed to be complete and accurate, subject to change. Digital version compared with original source data for accuracy and completeness. In process of validation by source and/or responsible agency.
Horizontal Positional Accuracy	An explanation of the accuracy of the horizontal positions (coordinates) of spatial objects and a description of the tests or line of reasoning used to arrive at the estimate.	In process of validation by source and/or responsible agency.
Vertical Positional Accuracy	An explanation of the accuracy of the vertical positions (coordinates) of spatial objects and a description of the tests or line of reasoning used to arrive at the estimate. This only applies to datasets with a vertical component such as topography or spot elevations. Use "N/A" for datasets not containing a vertical component.	N/A
2.8	**Resource Lineage**	
Type of Source Media	As part of source information, the type of media or presentation form of the source information.	Map

Source Citation Abbreviation	As part of source information, a short title or descriptor of the source.	Zoning map of DC 1996
Source Contribution	As part of source information, the part of the layer the source information contributed to in data development.	geography and attributes
Source Title	As part of source information, the title or descriptor of the source.	Zoning map of DC 1996
Source Originator	As part of source information, the name of the organization or individual that developed the source.	DCOZ
Source Publication Date	As part of the source information, the publication date of the source. This is formatted to “dd/mm/yyyy.”	1996
Source Online Linkage	As part of source information, the online web url for the source originator or document.	http://dcoz.dc.gov
Process Description	As part of the process step block, information about the sources of data used to construct the data set and steps used to process the data. This includes detailed process descriptions, how the data was created and checked including any special conditions, parameters, or findings, where needed to accurately describe and portray the data set.	Base data came from two sources: the street centerline developed by Spatial Systems (SSI), and georeferenced scanned maps from the Real Property Tax Administration. Attribute data was provided by zoning regulations and orders. This map is unofficial, not yet adopted by the Zoning Commission. Planimetric data was not available when this data was first created. Maintenance for Zoning layer: Notification of new, approved zoning orders is sent to the Office of Zoning's contractor, Michael Baker (Baker). Baker staff make the requested changes to the Zoning layer within the Zoning geodatabase. Zoning layer boundaries are snapped to the Street Centerline GIS layer from the DC Atlas. Once completed, the updated layer is posted to the server at the Office of Zoning. The related tables "Zoning_Cases" and "Zoning_Orders" are also updated with information about the zoning order (such as the order number, square and lot of the requested change, previous and new zoning for the area, and dates of the hearing and approval). The Case Number serves as the primary key to link the geospatial layer to these tables. The zoning map represented by this layer is unofficial and has not been adopted by the DC Office of Zoning. Once the DC OZ adopts the map, the only official Zoning map will reside on the DC Office of Zoning's server. Extracts from the map may be distributed to other DC government offices, but these extracts will not be official Zoning maps. All data is stored and exported in Maryland State Plane coordinates NAD 83 meters.

Process Date	As part of the process step block, the date the processes occurred. This is formatted to "dd/mm/yyyy."	20010806
Process Contact Person	As part of the process step block, the person who can receive inquiries and has the best knowledge regarding the content, procedures, and parameters used in the process step. This may not be the same as the data set contact person.	Resource Name
Process Contact Organization	As part of the process step block, the name of the organization that performed the processing. Typically, this is the process contact person's organization.	D.C. Office of Zoning
Process Contact Person Position	As part of the process step block, the position of the Point of Contact Person.	
Process Contact Address	As part of the process step block, the mailing address of the process Contact Person. Ensure that the address type is set to the appropriate choice – mailing or physical address, or both.	441 4th St NW Suite 210 South
Process Contact City	As part of the process step block, the city of the process Contact Person.	Washington
Process Contact State or Province	As part of the process step block, the state or province of the process Contact Person.	DC
Process Contact Postal Code	As part of the process step block, the ZIP or other postal code of the Contact Person.	20001
Process Contact Voice Phone	As part of the process step block, the voice telephone number, including area code, of the process Contact Person.	(xxx) xxx-xxxx
Process Contact Fax Phone	As part of the process step block, the facsimile telephone number, including area code, of the Contact Person.	
Process Contact E-mail Address	As part of the process step block, the Internet mail address of the Contact Person.	Resource email
Process Contact Hours of Service.	As part of the process step block, the hours of service of the process point of contact.	10 am - 5 pm
2.9	**Distribution**	

Distributor Contact Person	A person who can receive inquiries regarding the distribution of the data set.	
Distributor Organization	The name of the organization that can receive inquiries regarding the distribution of the data set.	DC Office of Zoning
Distributor Position	The position of the Distributor Contact Person	
Distributor Address	The mailing address of the Distributor. Ensure that the address type is set to the appropriate choice – mailing or physical address, or both.	441 4th St NW
Distributor City	The city of the Distributor.	Washington
Distributor State or Province	The state or province of the Distributor.	DC
Distributor Postal Code	The ZIP or other postal code of the Distributor.	20001
Distributor Voice Phone	The voice phone number, including area code, of the Distributor.	
Distributor Fax Phone	The facsimile telephone number, including area code, of the Distributor.	
Distributor E-mail Address	The Internet mail address of the Distributor.	dcgis@dc.gov
Distributor Hours of Service	The hours of service of the Distributor.	10am – 5pm
Distribution Liability	Statement of the liability assumed by the distributor.	This work is licensed under a Creative Commons Attribution 4.0 International License.
Ordering Instructions	Instructions for obtaining the data set. Include instructions for obtaining custom orders. This depends on the availability of the data set.	DC datasets can be downloaded from "http://opendata.dc.gov".
2.10	**Entity Attribute**	
	This section describes the attribution of the dataset. The primary elements requiring population are the attribute definition. Each column in the data set needs a description clarifying the contents of the column. A definition source is required for ESRI items shape, shape.area, shape.perimeter, and fid stating “ESRI.” Users are encouraged to enter the value domain of the column in the appropriate tags to help users translate the contents of the column where needed, especially when lookup tables do not exist.	
Attribute Definition		FID

	Description of each column in the data set.	FID OID 4 0 0 Internal feature number. ESRI Sequential unique whole numbers that are automatically generated Etc.
3.1	**Metadata Reference**	
Metadata Date	The date that the metadata were created or last updated.	5/6/2018
Metadata Contact Person	The name of the person responsible for the metadata information.	Resource Name
Metadata Contact Organization	The name of the organization responsible for the metadata information.	D.C. Office of Zoning
Metadata Contact Position	The position of the Metadata Contact Person	
Metadata Contact Address	The mailing address of the Metadata Contact. Ensure that the address type is set to the appropriate choice – mailing or physical address, or both.	441 4th St NW
Metadata Contact City	The city of the Metadata Contact.	Washington
Metadata Contact State or Province	The state or province of the Metadata Contact.	DC
Metadata Contact Postal Code	The ZIP or other postal code of the Metadata Contact.	20009
Metadata Contact Voice Phone	The voice phone number, including area code, of the Metadata Contact.	(xxx) xxx-xxxx
Metadata Contact Fax Phone	The facsimile telephone number, including area code, of the Metadata Contact.	Same
Metadata Contact E-mail Address	The Internet mail address of the Metadata Contact.	zoning.info@dc.gov
Metadata Contact Hours of Service	The hours of service of the metadata contact.	9 am - 5 pm
Metadata Review Date	The review Data of the Metadata by the Metadata Contact Person. This is formatted to "dd/mm/yyyy."	5/6/2018

Appendix A – EDI Data Dictionary

Field Name	Data Type	Domain	Description
QB_DATE_CREATED	Date		Quickbase Record Creation Date
QB_DATE_MODIFIED	Date		Quickbase Record Modified Date
EDI_ID	Integer		EDI Unique Identifier
QB_RECORD_OWNER	String		Quickbase Record Owner
QB_LAST_MODIFIED_BY	String		Quickbase Record Last Modified User
FORMDATAINVENTORYID	Integer		Data Inventory Id
AGENCYCODE	String		Agency Abbreviation
AGENCYNAME	String		Agency Name
DATASETNAME	String		Dataset Name
DATABASENAME	String		Database Name
LINKEDAPPLICATION NAME	String		Linked Application Name
DATACATALOGNAME	String		Data Catalog Name
DATADIAGRAMFILEPATH	String		Data Diagram File Path
DATASETAUDIENCEVALUE	String		Dataset Audience Value
DATASETCLASSIFICATIONNAME	String	DATASET_CLASSIFICATION	Dataset Security Classification
DATASETCLASSIFICATIONREASON	String		Dataset Classification Justification
DATAOWNER	String		Data Owner Agency Abbreviation
DATASETCATEGORY	String	DATASET_CATEGORY	Dataset Topic Category
DATASETTYPE	String	DATASET_TYPE	Dataset Type
EARLIESTDATEOFRECORDS	Date		Earliest Date of Available Record
RECENTDATAOFRECORDS	Date		Recent Date of Record
UPDATEINTERVAL	String		Dataset Update Frequency
RETENTIONSCHEDULE	String		Data Retention Schedule
PUBLICATIONBARRIERS	String		Publication Barrier
BARRIERSUMMARY	String		Description of Dataset
PUBLICINTERESTINDATAVALUE	String		Public Interest in Dataset
DATASETURL	String		Dataset Website Link
KEYWORDS	String		Dataset Keywords/Tags
NAMEGCDESIGNEE	String		General Counsel Designee Name
EMAILGCDESIGNEE	String		General Counsel Designee Email
CREATEDDATA	Date		Record Created Date
MODIFIEDDATE	Date		Record Modified Date
DCS_LAST_MOD_DTTM	Date		Date, Record Was Last Modified
REVIEWEDBYOCTO	String		Flag for OCTO Review
ISSUEIDENTIFIED	String		Identified Issue
DERIVATIVEDATASET	String		Derivative Dataset
ISSUENOTES	String		Issue Notes
CURR_STATUS	String		Current Status
PUBLISHED_DATE	Date		EDI Record Published Date
COMMENTS	String		Comments
LATEST_IT_ASSESSMENT	String		Latest IT Assessment Date

OPENDATA_PORTAL	String		Defines whether dataset is published in Open Data Portal.

DATASET_CLASSIFICATION	VALUE	DESCRIPTION
	Open	Datasets that do not fall within the definitions of Public Not Proactively Released, For District Government Use, Confidential, Restricted Confidential.
	Public Not Proactively Released	Dataset that is not protected from public disclosure or subject to withholding under any law (including FOIA), regulation, or contract. Nevertheless, publication of the dataset on the public Internet and exposure to search engines would: 1. Have the potential to jeopardize the safety, privacy, or security of residents, agency workforce members, clients, partners, or anyone else identified in the information; 2. Require subjective redaction; 3. Impose an undue financial or administrative burden on the agency; or 4. Expose the District to litigation or legal liability.
	For District Government Use	Dataset that the originating agency determines is subject to one or more FOIA exemptions, is not highly sensitive, and may be distributed within the District government without restriction by law, regulation, or contract.
	Confidential	Dataset that the originating agency has determined is protected from disclosure by law, including FOIA, regulation, or contract and that is either highly sensitive or is lawfully, regulatorily, or contractually restricted from disclosure to other public bodies. Such datasets generally include datasets that contain data that qualifies for designation by a federal agency or District agency as: 1. Attorney-Client Privileged; 2. Criminal Justice Information; 3. Critical Infrastructure Information; 4. Family Educational Rights and Privacy Act (FERPA); 5. Federal Tax Information (FTI); 6. For Official Use Only (FOUO); 7. Law Enforcement Sensitive; 8. Legally privileged; 9. Payment Card Information (PCI); or 10. Protected Health Information (PHI) under the Health Insurance Portability and Accountability Act (HIPAA); 11. Sensitive but Unclassified.
	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.

DATASET_TYPE	VALUE	DESCRIPTION
	Database	Contained in a database or system
	Spreadsheet	Stored as a spreadsheet
	Image	Formatted as a raster dataset or picture image
	Other	Other dataset type
	Resource	Resource type
	Chart	Presented as a chart or graph
	Geographic Information	Stored as a geospatial layer
	Text	Stored at a CSV or text file
	Unknown	Unknown type

DATASET CATEGORY	VALUE	DESCRIPTION
	Administrative and Other Boundaries	Administrative and other areas such as wards and service boundaries
	Aerial Photography and Scanned Maps	Aerial photography, scanned maps, and raster imagery
	Basemap	Derivative layers used to create the basemap
	Business and Economic Development	Business and economic activities, conditions, and employment
	Communication	Communications infrastructure and services
	Cultural and Society	Characteristics of society and culture
	Demographic	Characteristics of human populations
	Education	Characteristics of education
	Elevation	Datasets that deal with topography and land formations
	Environment	Environmental resources, protection and conservation
	Facility and Structure	Facility and Man-made construction
	Financial	Financial and banking services
	Government Operations	Datasets that are part of government operations such as service requests and building permits
	Health	Health, health services, human ecology, and safety
	Historic	Historic themed data
	Location	Positional information and services
	Planning Land Use and Zoning	Information used for land use and zoning
	Property and Land	Property and Land
	Public Safety	Public safety activities and services
	Public Service	Public services activities
	Recreation	Recreation and outdoor activities and facilities
	Technology	Technology services and activities
	Transportation	Means and aids for conveying persons or goods
	Utility and Communication	Energy, water and waste systems and communications infrastructure and services

Appendix B – Data Submission Data Dictionary

Field Name	Data Type	Domain	Description
DATASET_ID	Integer		Unique Data Identifier
DATASET_EDI_ID	Integer		Enterprise Data Inventory ID
DATASET_EDI_EXIST	Boolean		Indicate A Dataset Exists In EDI
DATASET_EDI_NAME	String		Dataset Name
DATASET_EDI_CLASSIFICATION	String	EDI_CLASSIFICATION	Dataset Classification Level
DATASET_EDI_SOURCE_FORMAT	String	EDI_SOURCE_FORMAT	Dataset Source Format
DATASET_EDI_DESCRIPTION	String		Dataset Description
DATASET_SECURITY_CHECK	Boolean		Indicate Security Check Requirement
DATASET_TASK_TYPE	String	DATASET_TASK_TYPE	Task Types
DATASET_TASK_DETAILS	String		Task Detail Notes
DATASET_APPLICATION_DEPENDENCY	String		Dataset Application Dependency Note
DATASET_PRIMARY_KEY	String		Dataset Primary Key Field
CLASSIFICATION_EDI_LEVEL	Integer	EDI_LEVEL	Dataset Security Classification Level as Stored In EDI
CLASSIFICATION_EDI_REASON	String		Dataset Classification Reason
CLASSIFICATION_DERIVATIVE	Boolean		Indicate Dataset Is Derivative
CLASSIFICATION_METHODOLOGY	String		If Yes to Derivative, Notes on How to Process Dataset
CLASSIFICATION_SECURITY	String		Additional Security Notes
GEOSPATIAL_PROJECTION	String	PROJECTION	Geospatial Projections
GEOSPATIAL_GEOCODE	Boolean		Indicate If Geocoding Is Required
GEOSPATIAL_OUTSIDE_DC	Boolean		Indicate Locations Outside DC Exist
GEOSPATIAL_GEOMETRY_TYPE	String	GEOMETRY_TYPE	Geometry Type
GEOSPATIAL_TOPOLOGY	Boolean		Indicator If Topology Check Is Required
ETL_METHODOLOGY	String	ETL_METHODOLOGY	ETL Methodology for Extracting the Data from Source
ETL_UPDATE_FREQUENCY	String	UPDATE_FREQUENCY	ETL Update Frequency
ETL_DBCONNECTION_DBNAME	String		ETL Database Name
ETL_DBCONNECTION_SCHEMANAME	String		ETL Database Schema Name
ETL_DBCONNECTION_TABLENAME	String		ETL Database Table Name
ETL_DBCONNECTION_USERNAME	String		ETL Database Username
ETL_API_WEB_SERVICE_URL	String		ETL API Web Service Url
ETL_FILESHARE_LOCATION	String		ETL File Share Location
ETL_EMAIL_ATTACHMENT_SUBJECT	String		ETL Email Subject
ETL_EMAIL_ATTACHMENT_FILENAME	String		ETL Email Attachment Filename
ETL_EMAIL_ATTACHMENT_EMAILADDRESS	String		ETL Email Address
ETL_OTHER	String		Other ETL Type
METADATA_EDI_AGENCY_NAME	String	AGENCY_NAME	Agency That Submits the Data

METADATA_EDI_OWNER	String		Agency That Owns the Data
METADATA_EDI_KEYWORDS	String		Keywords (Or Phrases) For Searching Purposes
METADATA_CREDIT	String		A Recognition of Who Created or Contributed to The Dataset
METADATA_EDI_DATASET_CATEGORY	String	DATASET_CATEGORY	General Category Within Which the Dataset Should Be Included In
METADATA_CREATION_DATE	String		Date the Data Was Created
METADATA_REVISED_DATE	String		Date the Data Was Revised
METADATA_PUBLICATION_DATE	String		Date the Data Was Published
METADATA_SOURCE	String		Data Source
METADATA_EDI_EARLIEST_RECORDS	String		Earliest Date of Available Record
METADATA_EDI_RECENT_RECORDS	String		Recent Date of Available Record
METADATA_EDI_DATASET_URL	String		The URL To Source Link or Dataset Program Web Page
METADATA_EDI_UPDATE_INTERVAL	String	UPDATE_FREQUENCY	Data Update Frequency
METADATA_EDI_PUBLIC_INTEREST	String	PUBLIC_INTEREST	Public Interest in Dataset
METADATA_EDI_DATASET_URL	String		Dataset Web Address
CONTACT_NAME	String		Contact Name
CONTACT_AGENCY	String	AGENCY_NAME	Contact Agency
CONTACT_PHONE_NUMBER	String		Phone Number
CONTACT_EMAIL	String		Email Address
MICELLANEOUS_NOTES	String		Miscellaneous Notes
MISCELLANEOUS_ATTACHMENTS	Binary		Attachments

DATASET_EDI_CLASSIFICATION:	VALUE	DESCRIPTION
	Open	Datasets that do not fall within the definitions of Public Not Proactively Released, For District Government Use, Confidential, Restricted Confidential.
	Public Not Proactively Released	Dataset that is not protected from public disclosure or subject to withholding under any law (including FOIA), regulation, or contract. Nevertheless, publication of the dataset on the public Internet and exposure to search engines would: 1. Have the potential to jeopardize the safety, privacy, or security of residents, agency workforce members, clients, partners, or anyone else identified in the information; 2. Require subjective redaction; 3. Impose an undue financial or administrative burden on the agency; or 4. Expose the District to litigation or legal liability.
	For District Government Use	Dataset that the originating agency determines is subject to one or more FOIA exemptions, is not highly sensitive, and may be distributed within the District government without restriction by law, regulation, or contract.

	Confidential	Dataset that the originating agency has determined is protected from disclosure by law, including FOIA, regulation, or contract and that is either highly sensitive or is lawfully, regulatorily, or contractually restricted from disclosure to other public bodies. Such datasets generally include datasets that contain data that qualifies for designation by a federal agency or District agency as: 1. Attorney-Client Privileged; 2. Criminal Justice Information; 3. Critical Infrastructure Information; 4. Family Educational Rights and Privacy Act (FERPA); 5. Federal Tax Information (FTI); 6. For Official Use Only (FOUO); 7. Law Enforcement Sensitive; 8. Legally privileged; 9. Payment Card Information (PCI); or 10. Protected Health Information (PHI) under the Health Insurance Portability and Accountability Act (HIPAA); 11. Sensitive but Unclassified.
	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.

DATASET_TYPE:		
	Database	Contained in a database or system
	Spreadsheet	Stored as a spreadsheet
	Image	Formatted as a raster dataset or picture image
	Other	Other dataset type
	Resource	Resource type
	Chart	Presented as a chart or graph
	Geographic Information	Stored as a geospatial layer
	Text	Stored as a CSV or text file
	Unknown	Unknown type

DATASET_TASK_TYPE:		
	Replace	Replace dataset
	Update	Update dataset
	Modify	Modify dataset structure
	Add	New dataset

CLASSIFICATION_EDI_LEVEL:		
	0	Open
	1	Public Not Proactively Released
	2	For District Government Use
	3	Confidential
	4	Restricted Confidential

GEOMETRY_TYPE:		
	Point	Point

	Line	Line
	Polygon	Polygon

GEOSPATIAL_PROJECTION:		
	None	No Projection
	MD State Plane	NAD 1983 State-Plane Maryland FIPS 1900 (Meters)
	Other	Other coordinate system
	State Plane - Other	State Plane coordinate system from another State
	Web Mercader	WGS 1984 Web Mercator (auxiliary sphere)

METADATA_EDI_DATASET _CATEGORY:		
	Administrative and Other Boundaries	Administrative and other areas such as wards and service boundaries
	Aerial Photography and Scanned Maps	Aerial photography, scanned maps, and raster imagery
	Basemap	Derivative layers used to create the basemap
	Business and Economic Development	Business and economic activities, conditions, and employment
	Communication	Communications infrastructure and services
	Cultural and Society	Characteristics of society and culture
	Demographic	Characteristics of human populations
	Education	Characteristics of education
	Elevation	Datasets that deal with topography and land formations
	Environment	Environmental resources, protection and conservation
	Facility and Structure	Facility and Man-made construction
	Financial	Financial and banking services
	Government Operations	Datasets that are part of government operations such as service requests and building permits
	Health	Health, health services, human ecology, and safety
	Historic	Historic themed datasets
	Location	Positional information and services
	Planning Land Use and Zoning	Information used for land use and zoning
	Property and Land	Property and Land
	Public Safety	Public safety activities and services
	Public Service	Public services activities
	Recreation	Recreation and outdoor activities and facilities
	Technology	Technology services and activities
	Transportation	Means and aids for conveying persons or goods
	Utility and Communication	Energy, water and waste systems and communications infrastructure and services

ETL_METHODOLOGY:		
	Direct Database Connection/ODBC	Direct Database Connection or connect through a computer installed driver.
	ETL API Web Service	Application API or customized Web Service

	File Share/FTP	Output file or report placed on a network share or FTP site
	Email Attachment	Output file or report attached to an email with specified email subject
	Other	Other method

ETL_UPDATE_FREQUENCY:		
	Real-time/Near-time	Real-time/Near-time update
	Daily	Daily
	Weekly	Weekly
	Monthly	Monthly
	Quarterly	Quarterly
	Biannually	Biannually
	Yearly	Yearly
	Static	Static
	Unknown	Unknown

METADATA_EDI_PUBLIC_INTEREST:		
	Frequent Request	Frequent Request
	No Requests	No Requests
	Occasional	Occasional Request