City of Gainesville

OPEN DATA PUBLICATION HANDBOOK

January 2019

CONTENTS

1. PURPOSE

This Open Data publication process handbook is intended to provide guidance to City of Gainesville departments and to give an overview of procedures for identifying, prioritizing, publishing and maintaining open data.

This handbook will be updated as needed by the Open Data Governance Team. All questions should directed to Strategic Initiatives department at DG_Strategic_Planning@cityofgainesville.org

2. ROLES AND RESPOSIBILITIES

Publication of datasets will be done in close cooperation between the Open Data Governance Team and individual city department. The following roles and responsibilities will be used to facilitate the implementation of this procedure:

2.1 Open Data Governance Team

The Open Data Governance Team is the internal decision-makers regarding the curation, release, and management of datasets onto dataGNV.

2.2 Open Data Manager

The Open data Manager is responsible for managing all aspects of the city's open data program, and act as a frontline for the open data portal. The open data manager oversees system integration between departmental data sources and the open data portal to facilitate availability of datasets. The Open Data Manager is also responsible for the quality assessment of datasets.

2.3 Department Data Stewards

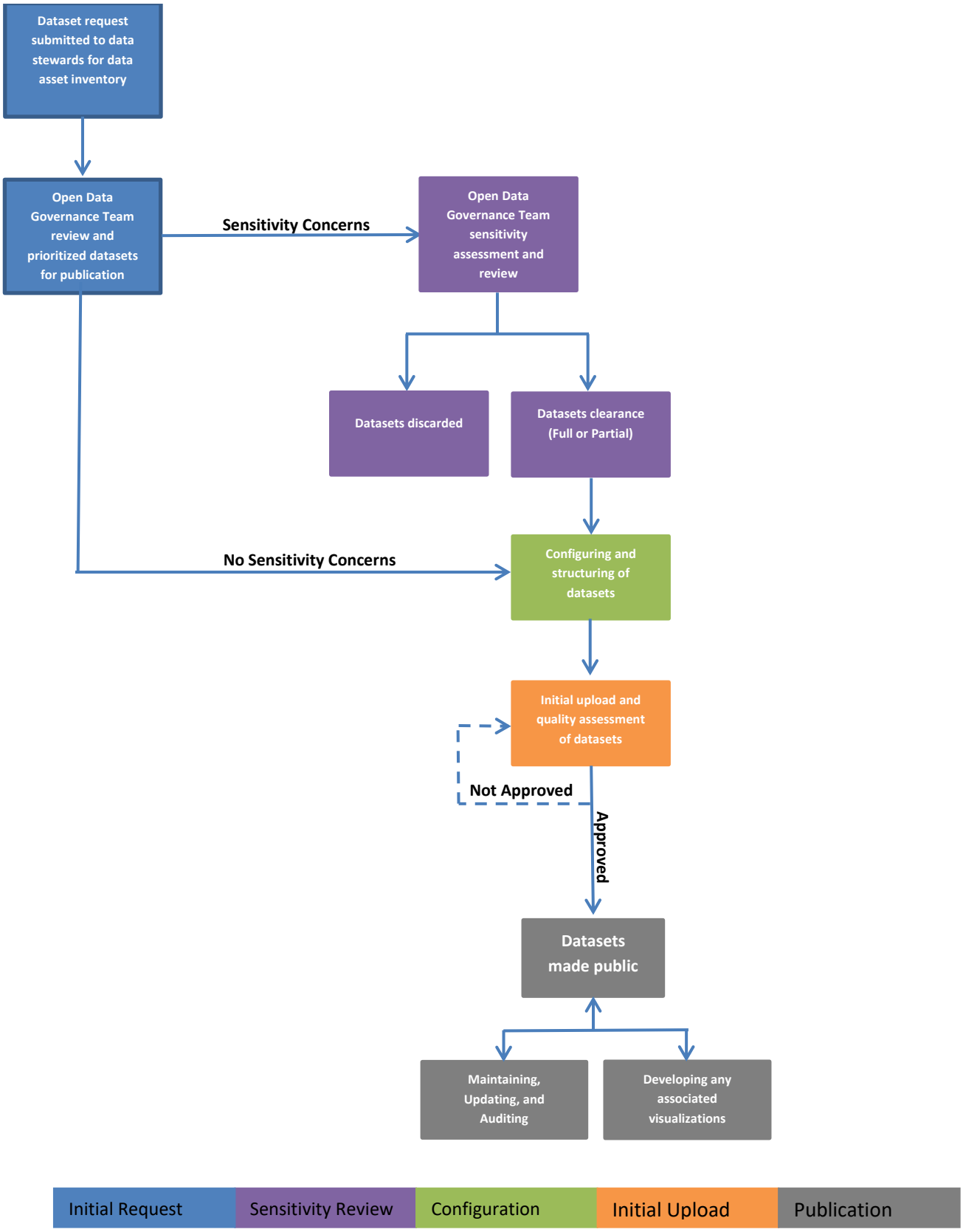
Data Stewards are appointed by department directors based on their department's data administration needs. Data Stewards will be expected to work closely with departmental staff to ensure data is properly inventoried, and validated. It is recommended that data stewards: 1) have access to and familiarity with departmental systems, 2) have an understanding of the programs and goals of the department, 3) are effective communicators, 4) have an interest in and understand the value of open data. Data Stewards responsibilities include:

- *Identifying Data Sources*: Completing an initial data request worksheet for each dataset in the respective department's inventory
- *Developing Data Inventories:* Cataloging and maintaining departmental data inventories; updating departmental data inventories as often as necessary.
- *Assisting in Data publication:* Assisting with access to, configuration and validation of, datasets for publication as open data; Updating frequently existing departmental open datasets and related metadata.
- *Communicating and Collaborating:* Helping to foster an environment that encourages open data and collaboration with other city departments and government agencies by acting as an open data advocate within each respective department;

3. PROCESS OVERVIEW

The Open Data Governance Team will review the data asset inventory to determine datasets appropriate for publication on Gainesville's Open Data Portal. Request for unpublished data will be submitted to departments/data stewards. Prior to submitting the request for data, the process will be communicated to the city's leadership team. The Open Data Governance Team will review and prioritize datasets for publication based on strategic priority areas: Strong Economy, Better Future, Greater Equity, and Community Model. Following data prioritization, ongoing efforts will be made to publish data on dataGNV, the city's open data portal. Data publication entails five steps: Initial request, sensitivity review, dataset configuration and formatting, initial upload, and publication. See figure below for illustration of this workflow.

3.1 Open Data Publication



3.2 Dataset Request

A dataset inventory will occur annually as well as on ad hoc basis throughout the year as new datasets arise. As a result, the Open Data Manager will send a data spreadsheet request to identified data steward in each department. Each data steward will use this spreadsheet to list all datasets created or managed by their department, and return this for review and prioritization. Note that all datasets will be inventoried including datasets that may contain sensitive or restricted datasets, or datasets that are not considered 'valuable' to the open data program. Each dataset will include the following information:

- Department (e.g., Strategic Initiative, GPD)
- Strategic Priority Area (Strong Economy, Better Future, Greater Equity, and Community Model.)
- Name (e.g., Building Permits, Crime Incidents)
- Format (e.g., CSV, ArcGIS Feature Class)
- Granularity
- Source (e.g., SQL DB, Access DB, SDE GeoDatabase)
- Update Frequency (e.g., daily, weekly, annually)
- Data Steward (e.g., Pierre Nathan)
- Sensitive data (yes/no/not sure; description)
- Description and other notes

While filling the spreadsheet, data stewards will also express a data "wish list", specifying datasets their department procures or wishes to procure from other departments. Any current interdepartmental data exchanges should also be included in this list to identify opportunities to create efficiencies in existing data sharing processes. One of the main advantages of the open data program is that it facilitates efficient inter-departmental data sharing by housing city data on a centralized share drive. The data "wish list" will include these fields:

- Dataset (e.g., Building Inspections)
- Requesting Department (e.g., Fire Department)
- Owner Department (e.g., Department of Doing)
- Description of Use (e.g., Used to predict fire risks and vulnerabilities)
- Requested Update Frequency (e.g., weekly)

3.3 Prioritizing New Datasets

Following the data request and inventory process, the Open Data Governance Team will evaluate and prioritize datasets. Each dataset will be rate as high, medium, or low in term of value and ease of publication as defined below:

- **Value:**
    - *Low* – The dataset historically has had little or no public demand. Information in the dataset does not help evaluate the needs identified by the Strategic

Framework, leadership priorities or departmental objectives. The dataset is not a source of performance measure

- *Medium* – At least one of the following is true concerning the dataset. The dataset has been previously requested, or has the potential to be requested by the public, business or academic community, or the dataset may be useful to other departments and has been included on a department data "wish list" as defined in the previous section
- *High* – The dataset is frequently requested. The dataset is tied to a performance management goal, the city manager's office initiative, or would provide immediate benefit to multiple departments, or the community for making data informed decisions, or other purpose.

- **Ease of Publication:**

  - *Low* – One or more of the following is true concerning the dataset. The dataset contains significant sensitivity concerns that will require an extended sensitivity review process, or a significant revision of the dataset to resolve. The dataset requires significant configuration to generate the file in a format the open portal accepts. Setting up the automation process will require a change in business process and will require significant effort.
  - *Medium* – The dataset configuration will require moderate effort, and can be generated in a format that the portal accepts. Sensitivity concerns are minimal and can be successfully addressed by the Open Governance team or the City Attorney Office.
  - *High* – Effort to channel the dataset through the open data process is low; most of the following are true concerning the dataset. Very minimal (if any) sensitivity concerns are found in the dataset, and are quickly and easily mitigated by the Data Stewards. Metadata is already accessible, or easy to produce. The data is already in a format that can be uploaded to the portal, or is easily configured and converted. An existing automation process used to maintain data on the portal can be implemented for the dataset.

3.4 Sensitivity Assessment

Datasets rated high are prioritized for upload to the open data portal and assessed for sensitivity related to privacy and security. The Open data governance committee will conduct a sensitivity review and identify potential concerns as either none, minor, significant, or overwhelming.

- *No concerns:* The dataset can be configured and structure by data steward for publication
- *Minor concerns*: such as a column that can be easily removed, are communicated to the department/Data steward for mitigation. The dataset is then re-reviewed by the Strategic Initiative department before publication.

- *Significant concerns*: represent issues that are not easily resolved but do not immediately disqualify the dataset, such as the possibility that anonymized individuals in a dataset could be easily re-identified when the dataset is combined with another dataset on the portal. Once the Open Data Committee identifies such concerns, the dataset is forwarded to the City attorney office for review and clearance (partial or full) for publication.
- *Overwhelming concerns*: immediately disqualify a dataset from being posted to open data. Datasets identified as overwhelmingly concerning are not uploaded to the open data portal.

3.5 Configuring and Formatting Datasets

Once datasets are prioritized and approved for publication, data stewards will configure and format the dataset for publication according the standard outlined below. This is to ensure that data release to the public is as accurate, complete and up to date as possible. Each dataset is unique and required a different process to check and access quality. However, certain commonalities apply across all datasets. The data steward and the Open Data Manager should collaborate to find the best way to configure and format dataset.

3.5.1 Common Data quality's key principles:

Key Principles	Description	Example: Active business
Consistency	Variables should be processed and documented the same way across datasets. Make geographic boundaries like address, district, and census tract standard if your dataset have them. Also, the field use in one dataset must match the definition in other datasets.	• The address of businesses location should have a consistent format across the datasets (e.g."2001 SW 16TH ST GAINESVILLE, FL" vs "2001 SOUTHWEST 16TH STREET GAINESVILLE FLORIDA". • The spelling of the business type should be consistent throughout across the dataset (e.g. "Restaurant" vs "restaurant")
Uniqueness	Entries must be unique with no duplicate of the key identifier.	We should have no more than one identifier number per business, and this must match identifier in other dataset (e.g. building permit).This allows for multiple records of the same business in the same dataset, but with different values in related fields
Accuracy	Accuracy is the degree to which your dataset represent reality. Ways to check accuracy include comparing to similar datasets as well as doing spot checks or audits. If your dataset contains many entries, use the accuracy rate of a sample inspection to estimate the accuracy of your dataset and/ or fields.	• If a business location in the Active business dataset is in Gainesville, then the zip code in the address must start by 326… • Annually, a staff member calls a sample of active businesses phone number to check for accuracy.
Completeness	Often defined as % of data values that are complete, e.g. % of Active business with a phone number. Mandatory elements should be 100% complete. You can also define completeness for a dataset (versus a single field).	We have a universe of 300 active business and 250 valid addresses. Our dataset has a completeness of 83%
Validity	Each data field has a syntax, range or set of rules it should conform to. Some of these rules are outline below in the data structure.	See Data structure section

3.5.2 Use Vertical rather than Horizontal Orientation

Horizontal data orientation should be restructured to vertical whenever feasible. Vertical datasets are more easily understood and sortable, as well as more useful for creating visualizations. This is especially true for datasets containing data by year, especially numerous years. Years should have their own rows, rather than columns, in the data. Update of the data become much more difficult as data files get wider with numerous columns. Moreover, horizontal orientation (as opposed to vertical) restricts the ability to perform analytics and observe time based trends in a single view. Vertical orientation of the submitted data file provides maximum flexibility for reuse and application, as the data can then be sorted by year or to create visualizations with the data rolled up by year. Vertical data orientation not only makes the data machine-readable, but human readable. Any grouping of variables is possible with vertical orientation facilitating complex analyses of the data.

Example: RTS Ridership- by route, by month

Not Acceptable – horizontal data orientation - each subsequent year of data requires the addition of another column

Route ID	Route description	January 2015	February 2015	March 2015
1	Downtown to Butler Plaza: via Archer RD	56665	54706	54791
2	Downtown to Walmart Supercenter	10101	9522	9331
5	Downtown to Oaks Mall (via University Ave)	38470	40587	40377

Acceptable – vertical data orientation - additional years can be appended to the data file as needed

Route ID	Route description	Passengers	Month	Year
1	Downtown to Butler Plaza: via Archer RD	56665	January	2015
2	Downtown to Walmart Supercenter	10101	January	2015
5	Downtown to Oaks Mall (via University Ave)	38470	January	2015
1	Downtown to Butler Plaza: via Archer RD	54706	February	2015
2	Downtown to Walmart Supercenter	9522	February	2015
5	Downtown to Oaks Mall (via University Ave)	40587	February	2015
1	Downtown to Butler Plaza: via Archer RD	54791	March	2015
2	Downtown to Walmart Supercenter	9331	March	2015
5	Downtown to Oaks Mall (via University Ave)	40377	March	2015

3.5.3 Column Headers

- Should correctly defined the type of data
- Can't be ambiguous, and codes should not be used. However, if any codes absolutely must be used, they must be fully explained in the dataset documentation.
- Must be unique and short
- Should be in lower case except the first letter
- Contain only alphanumeric
- Should not be merged cells

Example: Active Businesses

Not Acceptable – multi-row header; column names uppercase; merged cells

BUSINESS	CONTACT	
NAME	NAME	NUMBER
SUNRISE	KEITHS DINER	352-226-3808

Acceptable- single row header; column names in title case

Business	Contact	Business Phone
SUNRISE	KEITHS DINER	352-226-3808

Not Acceptable – cryptic column names

Busnme	Ctnme	Busph
SUNRISE	KEITHS DINER	352-226-3808

Acceptable – clear, reasonably long column names

Business Name	Contact Name	Business Phone
SUNRISE	KEITHS DINER	352-226-3808

3.5.4 Rows

- Should be ordered from the leftmost column to the rightmost one
- Each row should be one observation.
- A group of rows related to one entity should repeat the entity for all rows in the group.

Example: Active Businesses

Not Acceptable – empty fields in-group of rows

Name	Business_Type	Contact
DHAIRYA, INC.	RESTAURANT	352-226-3808
	RETAIL MERCHANT	352-378-0983
	CONDO RENTAL	352-377-3322

Acceptable – field repeated for all rows in group

Name	Business_Type	Contact
DHAIRYA, INC.	RESTAURANT	352-226-3808
DHAIRYA, INC.	RETAIL MERCHANT	352-378-0983
DHAIRYA, INC.	CONDO RENTAL	352-377-3322

3.5.5 Blank, "N/A" or Other Unknown Cells

Blank fields, when left unexplained, often lead to confusion – particularly when the column is numeric

- If the Blank field represents zero, then the field should be zero
- If the blank field represents "not collected" or "unknown", then this should be explained in the metadata or data dictionary.

3.5.6 Text Fields must be trimmed of whitespace

Text fields must be trimmed of whitespace (space-padding), otherwise searching, sorting and filtering the data on the Open Data platform will not work as expected. The following VB script can be used in Excel to remove trailing whitespace from all fields. Please note that any CSV must be properly imported into Excel first, and then saved to CSV after the edit.

```
Sub NoSpaces() Dim c
As Range
For Each
c In Selection.Cells c = Trim(c)
Next End
Sub
```

3.5.7 Numeric Variable

- Do not mix text in a field that is intended to contain numeric data. Mixing text and numeric data in the same column will result in the entire column being stored as text
- Numeric data that represents money should be expressed as a full number where possible (E.g. "5000000" instead of "5 (million)").
- Numeric data that represents money should not include currency symbols, or commas for place-separators.
- Negative values should be preceded with a minus sign, not placed within parentheses
- Provide at least 2 decimal places of precision, and no rounding if possible, Indicate in the metadata how percentages and unit measurement are expressed.

Example: Gainesville Checkbook Expenditure

Not Acceptable– monetary values containing currency symbols, place-separators, varying decimal places, and parentheses

Fund	Department	Expenditure Detail	Transaction Amount
The General Fund	Police	Travel & Training	(63.5)
Fleet Management Services Fund	Fleet Management	Diesel Fuel	$1,000.34

Acceptable- monetary values containing no decimal places or two decimal places; negative values expressed with leading minus sign.

Fund	Department	Expenditure Detail	Transaction Amount
The General Fund	Police	Travel & Training	-63.50
Fleet Management Services Fund	Fleet Management	Diesel Fuel	1000.34

3.5.8 Date and Time

Standardizing date format is the only way to display trends over time. It is critical for conducting analyses, time series, and inform decision-making.

- Must be in the local (Eastern Standard)
- Must be written with the slash (e.g. 09/02/2013)
- Time must be follow by "PM" or "AM" if written in 12 hour time (e.g. 9:00:00 AM)
- Full dates are much more preferable to month, year for analyzing trends over time and should be provided any time if available. If only monthly data is available, the next best option is to provide a full date set to the last day of the month, e.g. 10/31/2017 – the display can be masked to show only month and year while still retaining the ability to trend.
- "FY" must precede the year for financial period (e.g. FY 2019)
- When the full date and time is available, it should be provided in a single field.

Example: 311 Service Request

Acceptable- full date and time.

Id	Issue Type	Service request Date and Time
4958300	Trash/Debris (Private Property)	09/27/2018 09:58:00 PM
4957977	Parking in Yard (Other Than Driveway)	09/27/2018 07:39:00 PM

3.5.9 Location Variable

Clean address data is very valuable as it can add another dimension to your data – geographic locations that can be mapped. To ensure accurate mapping, addresses must be broken into four columns: street Address, City, State and Zip Code. If latitude and longitude have not been provided, then a complete and clean address must be submitted for the platform to auto-generate latitude and longitude for those

datasets which lend themselves to mapping. If geocoding and mapping of address data is desired, then street address data must not contain P.O. Boxes or other information that cannot be accurately geocoded such as “corner of”, “in front of”, “across from”, etc. If this is the case then latitude and longitude should be provided. Geographic coordinates (latitude, longitude) should be specified in decimal degree.

Example: 311 Service Request

Not Acceptable – not formatted for mapping

Issue Type	Street Number	Street Address	City	Zip Code	LONGITUDE	LATITUDE
Trash/Debris	320	Ne 12th St	Gainesville, FL	32635	-82.310306708689	29.6542076513162
Parking in Yard	101	Nw 28th Ter	Gainesville, FL	32608	-82.324393037335	29.6520246551837

Acceptable – formatted for mapping

Issue Type	Street Address	City	State	Zip Code	LONGITUDE	LATITUDE
Trash/Debris	320 Ne 12th St	Gainesville	FL	32635	-82.310306708689	29.6542076513162
Parking in Yard	101 Nw 28th Ter	Gainesville	FL	32608	-82.324393037335	29.6520246551837

3.5.10 Dataset Metadata

Field	Label	Definition	Required	Examples: 311 Service Requests
title	Title	Human-readable name of the data. Should be in plain English and include sufficient detail to facilitate search and discovery.	Always	311 Service Requests
description	Description	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the data is of interest.	Always	Non-emergency requests like pothole repairs or damaged street signs can be reported through the 311GNV app (https://play.google.com/store/apps/details?id=com.seeclickfix.gnv311.app&hl=en) or online (http://www.cityofgainesville.org/openGNV/311GNV.aspx).
Theme	Category	Main thematic category of the dataset. Choose between" Better Future", "Community Model", "Greater Equity", "Strong Economy".	Always	Community Model
keyword	Tags/Keywords	Tags (or keywords) are short way to categorize dataset. These help users discover your dataset; please include terms that would be used by technical and non-technical users. Enter one or more keywords separated by commas.	Always	311, request, citizen, repair, non-emergency, reports
publisher	Data Provided By	The publishing entity and optionally their parent organization(s).	Always	City of Gainesville - Administrative Services
contactPoint	Contact Information	Contact person's email for the asset.	Always	dg_strategic_planning@cityofgainesville.org
identifier	Row Identifier	A unique identifier for the dataset or API as maintained within a department catalog or database.	If-Applicable	N/A
row Label	Row Label	Describe what each row in the dataset represents	If-Applicable	N/A
license	License	The license or non-license (i.e. Public Domain) status with which the dataset or API has been published.	If-Applicable	Public Domain
landingPage	Source Link	This field is not intended for a department's homepage, 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. Often, departments will have a programmatic website that describes the collection of data and methodologies in more detail.	No	http://www.cityofgainesville.org/openGNV/311GNV.aspx
Semantics &RDF	Row Class & Subject Column	Row Class and Subject Column are related to RDF ("Resource Description Framework"). These are used to further categorize datasets, and designate relationships through a graph database.	No	N/A

3.5.11 Data Dictionary

The data dictionary is one of the most important parts of documentation for your dataset. This document provides all of the information a user will need to interpret a dataset without you. This is good practice as it helps reduce the number of inbound questions and makes sharing data much easier. Generally, this is develop once, unless the underlying structure of the data changes

3.5.12 Column Metadata:

There should be documentation for every field in the datasets. Describe your columns so people understand how the data should be interpreted.

Example**:** Crime Incidents 2011-Present

Column Name	Column Description	API Field Name
ID	Incident identification number.	id
Incident_Type	Type of incident that occurred.	narrative
Report _Date	Date and time that the incident report was filed.	report_date
Offense _Date	Date and time of the offense.	offense_date
City	City where the incident occurred.	city
State	State where the incident occurred.	state
Location	Location in latitude and longitude.	location

3.5.13 Define methodologies where appropriate

If you have a more complicated methodology than can be explained in the general data description, please describe it in your data dictionary. For example, surveys that involve observation and special collection methods should be described in appropriate detail so the user understands the nature of the data. It is also appropriate to reference existing methods documentation via a Source link if you've already documented on a website for example. The main purpose is to make sure the user can easily ascertain how the data was created.

3.6 Initial upload and Quality assessment

3.6.1 Initial upload

Datasets are uploaded to Gainesville's open portal in two different ways:

- *Data Stewards upload data directly to the portal***:** Following configuration, data stewards will upload the new dataset to the Open Data Portal, and will input metadata (e.g., dataset title, categories, column descriptions) to contextualize datasets and dataset attributes. Once the dataset is uploaded, the Open Data Manager will review the data and conduct a quality assessment before released for publication.
- *Data stewards send Data to Strategic Initiatives department for upload to the Open portal***:** Data stewards send the new dataset and metadata (e.g., a readme text file, column descriptions) to the strategic Initiative department for upload to the Open data portal. Once uploaded, the dataset is share with the data steward for review before release for publication.

3.6.2 Quality assessment

After a dataset and associated metadata have been uploaded to the portal, but before it is released publicly, a quality assessment (QA) review takes place to find and correct any errors in the data itself or in the metadata. Datasets that pass both a second sensitivity review stage and QA are made public on the portal.The Gainesville's open data program use a quality standard checklist (See appendix A) during the dataset QA process**.**

The Gainesville's open data program is not, however, a data editing program. Data Stewards will not be asked to substantially change or improve datasets pulled from city systems before uploading them to the portal. Data Stewards are not asked to impute missing values, though they may do so if they wish. The primary goal of the Open Data Program is to share our existing data with the public and among city staff. With this in mind, the program also provides an opportunity, to gain feedback about the quality of our data from other departments and the public. Any improvement to internal processes for data creation and curation that results from the Open Data Program or associated feedback will be considered welcome.

3.7 Data Publication

In accordance with the publication schedule set by the Open Data governance team, the dataset will be published to the Open Data Portal. Scheduled communications will also occur in conjunction with the dataset publication. Examples of scheduled communications could include a combination of:

- A post to the city's social media accounts;
- A news release from the Communications department;
- An announcement on the Open Data Portal
- Direct communication to civic group, local, state or federal agencies;

3.8 Maintain, Update, and Audit

A successful open data program requires effective maintenance and up-to-date data. In addition to submitting dataset to be published to the portal, Data Stewards must also submit a plan specifying how often the dataset will be refreshed. After the dataset is published, Stewards are responsible (with 'as needed' assistance from the Initiative department) for all manual updates to datasets. Whenever possible, stewards will update an existing dataset instead of creating a new dataset. Two tactics should be employed to ensure that the open data remains up to date:

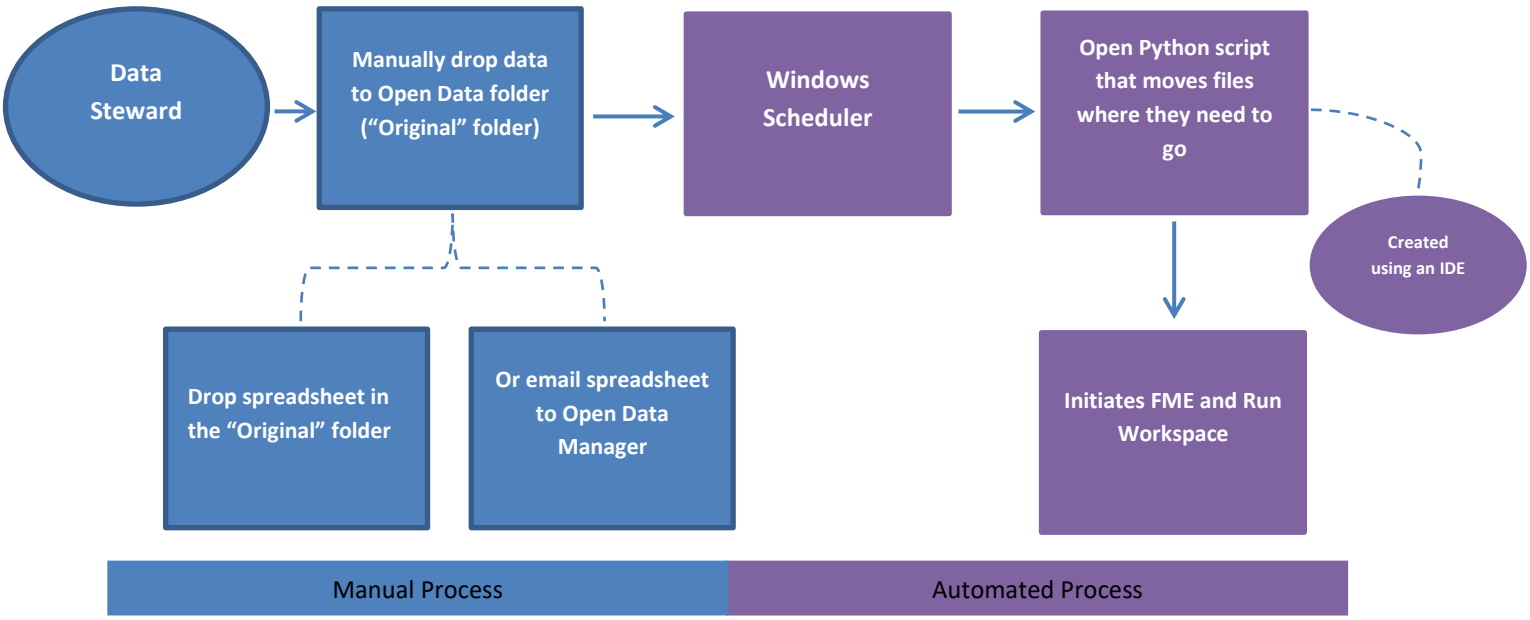
- *Active monitoring of datasets freshness*. Datasets that required manual update will be track on the portal. The Open Data Manager will alert data stewards when a dataset is identified as out of date and needs to be refreshed.
- *Automation.* Automation is the process of updating the Open Data portal programmatically, rather than manually. The Open Data Manager will work with data stewards, the department of technology, and others resources as needed to

implement automatic updates, where appropriate. Datasets that prompt “yes” to any of the questions below are great candidates for automating updates:

a) Is the dataset updated quarterly or more frequently?
b) Are there transformations or any form of manipulation that needs to be done to the dataset prior to uploading?
c) Is the dataset large (greater than 250MB)?
d) Can you only get the changed rows for each subsequent update (rather than the full file)?
e) Is it possible to get data from the source system, rather than from an individual?

Data extracted from a source system will be automated using Python and FME (See figure below for illustration of this workflow). In addition, the Open Data Manager will produce once a year a report of an audit of the data inventory and the Open Data catalog to the Open Data Governance Team. The purpose of this audit is to ensure that datasets are being published per their publication frequency as identified in the data inventory; that datasets are properly tagged and categorized; that dataset metadata is complete and accurate; and that the data inventory and Open Data Catalog perform as expected.

3.9 Automation of Datasets Publication using Python and FME



Process order (Open Data Manager will assist Data Stewards if needed):

1. Set up your workspace by downloading and installing an IDE to write or edit Python script and FME. Also, search your computer programs by looking for "Task Scheduler" and pin this program to your task bar.
2. Create FME workflow to manipulate and to publish the original data.
   - Open up the original data and open up the Socrata version of the dataset (this is for an update to an existing dataset)
   - Write out what needs to be done to the data to make sure it fits into the format of the data in Socrata.
   - Test the formatting of the dataset by using an excel writer.
   - Test in a Socrata dataset by going into Socrata and clicking "Edit" icon in the dataset that need update. This creates a new "working" version of the dataset that is not public facing and can be pushed to the published version if needed (this is for dataset update).
   - Make sure FME workflow is transforming the data correctly.
3. Create script
   This should move the file from "original" to "load" folder for the FME workflow to grab.
4. Create two Basic Task in Task Scheduler
   The Basic Task will be set up to move dataset from "original" to "load" and to initiate FME program automatically at a given date or time.

4. REFERENCES

This handbook was based on elements of open data handbooks from several cities including Boulder, New York, San Francisco, Greensboro, California, and others.

5. APPENDICES

Checklist for Open Data Quality

This Data quality checklist helps ensure quality data for release to the public. Use this check list as a companion to reading and using the Open data quality guide.

- ☐ Variables are documented the same way across the dataset
- ☐ Entries in the dataset are unique with no key identifier duplicate
- ☐ The dataset is the most complete, accurate, and current version appropriate for public release.
- ☐ The data have been spot checked for common errors such as missing and misplaced values
- ☐ Any missing data points are left as null, but the meaning of null is defined in the dataset's metadata.
- ☐ Columns are formatted appropriately.
- ☐ Rows are formatted appropriately
- ☐ Numeric variables are formatted appropriately
- ☐ Date and Time are formatted appropriately
- ☐ Metadata is complete, concise, and free of jargon.
- ☐ Metadata explain the process used to create the data and summarize any changes.
- ☐ Metadata clearly explain any limitations or omissions for each dataset.
- ☐ Metadata clearly identify an update frequency and plan.
- ☐ Location variable are formatted appropriately