

Technical Standards Manual

Technical Standards Manual (September 2022) Update

Revisions Overview

- **Format changes**

- This revision restructures the Technical Standards Manual (TSM) around the lifecycle of an Open Data dataset. Rather than dividing content into rules, policies, and guidelines, as in the [previous version](#), the manual instead presents the policies and guidelines in the context of identifying, structuring, documenting, and publishing Open Data. This change is intended to make the TSM:
 - more intuitive for Open Data Coordinators (ODCs) to use as a guide and find specific pieces of information
 - easier for the NYC Open Data Team to regularly update by linking out to external resources instead of incorporating new policy updates directly within the TSM body
 - a more readable document, allowing it to be a resource for members of the public to better understand how the NYC Open Data program works.
- Throughout, references to specific Local Laws have been replaced by references to the Open Data Law, or NYC Administrative Code Title 23, Chapter 5.

- **Content changes**

- [Section 1](#): The Foreword replaces the Executive Summary and Introduction from the previous version(s).
 - *New content*: Includes discussion of the NYC Open Data strategic plan.
- [Section 2](#): Roles & Responsibilities Within Your Agency, contains information on the responsibilities of agencies and their Open Data Coordinators.
 - *New content*: [2.3](#) includes new guidance for ODCs: information about convenings and trainings, and the transfer of knowledge requirement.
- [Section 3](#): Identifying Datasets, compiles all content relevant to the identification of datasets, including the operational definitions of data and public dataset, and an overview of the annual compliance process.
 - *New content*: [3.1](#) “What is a dataset?” provides more in-depth content about what data is mandatory for publication.
- [Section 4](#): Preparing datasets for publication, compiles all content relevant to the publication process, including the prioritization of datasets, metadata requirements, privacy considerations, and publication policies.
 - *New content*:
 - [4.1](#): Prioritizing datasets, contains more detail on prioritization criteria.
 - [4.4.2](#): Data dictionaries, includes the new tracking/versioning requirement.
 - [4.4.3](#): Data quality, contains guidance for documenting data quality issues

- [4.5](#): Privacy, contains new privacy resources for ODCs.
- *Removed content:*
 - Removed explicit mention of agencies self-hosting data, as law requires all data shared publicly to be available on Open Data, and direct availability on Open Data maximizes utility.
- [Section 5](#): Maintaining datasets, compiles all content relevant to the maintenance of datasets after publication, including dataset updates, automation, and retention and archiving.
 - *New content:*
 - [5.1](#): Consolidation
 - [5.2](#): Timely updates, includes requirements new for the TSM
 - [5.3](#): Platform-wide updates
 - [5.4](#): Data interoperability
 - [5.5](#): Dataset retention and archiving
- [Section 6](#): Open Data and the public, compiles all content relevant to Open Data public engagement in one place.

Technical Standards Manual

Table of Contents

1. [Foreword](#)
 2. [Roles & Responsibilities](#)
 - 2.1 - Chief Open Platform Officer (COPO)
 - 2.2 - Open Data Team
 - 2.3 - City Agencies
 - 2.4 - Open Data Coordinators (ODCs)
 3. [Identifying datasets](#)
 - 3.1 - What is a Dataset?
 - 3.2 - Annual Compliance Process
 - 3.3 - Technology Systems
 4. [Preparing datasets for Publication](#)
 - 4.1 - Prioritizing datasets
 - 4.2 - Components of Released Data
 - 4.3 - Geocoding
 - 4.4 - Metadata
 - 4.5 - Privacy
 - 4.6 - Dataset Publishing
 5. [Maintaining datasets](#)
 - 5.1 - Consolidation
 - 5.2 - Timely Updates
 - 5.3 - Platform-wide Updates
 - 5.4 - Data Interoperability
 - 5.5 - Dataset Retention and Archiving
 6. [Open Data and the Public](#)
 - 6.1 - Open Data Help Desk
 - 6.2 - Responding to Public Dataset Requests
 - 6.3 - Dataset Questions and Errors
 - 6.4 - Guidance for Responding to Public Inquiries
 - 6.5 - Civic Engagement Requirements
- Appendix A - [Definitions](#)
- Appendix B - [Contacting the Open Data Team](#)

1. Foreword

In late 2009, New York City published its first citywide, centralized collection of official data, featuring nearly 200 datasets from 30 City agencies, on what was then called the [NYC DataMine](#). This effort was launched jointly with the BigApps competition – a civic innovation challenge where New Yorkers competed to build applications to improve the City. The DataMine evolved into a more robust Open Data portal in 2011 – NYC Open Data – and in 2012, Open Data was signed into law.

Emphasizing the importance of data literacy and access to government records, the City has worked to make Open Data more accessible to New Yorkers. These efforts have included more guidance for open data novices; community-developed standards for easy-to-read data documentation; the launch of [NYC Open Data Week](#), an annual festival celebrating the free, public data that the City makes available; and the hosting of free Open Data 101 classes taught by volunteer Open Data Ambassadors.

As of 2022, there are more than 3,000 datasets on NYC Open Data, and a network of approximately 100 Open Data Coordinators (ODCs) – one from each City agency, office, or commission – who are responsible for identifying, documenting, and sharing their organization’s data. NYC Open Data is managed by the Open Data Team at the Office of Technology and Innovation (OTI), which oversees the process of keeping New York City’s open data up-to-date and accessible.

In September 2019, the Open Data Team outlined its [vision for the next decade of open data](#), co-created in collaboration with more than a dozen members of the City’s robust open data community. This strategic plan highlights three main priorities and outlines 27 specific initiatives. The three priorities are listed below, and update on the progress of these initiatives can be found in subsequent [Open Data annual reports](#).

- **Improving user experience:** making it easier for all New Yorkers to request, find, understand, and use data
- **Strengthening the City’s capacity:** supporting City agencies and other publishers of NYC Open Data with better tools, more resources, clearer standards, and stronger connections with existing Open Data operations
- **Building communities:** supporting New Yorkers with tools and information to help them solve local problems while amplifying the voices of people who find solutions using NYC open data.

More than a decade after the first datasets were published, NYC Open Data has evolved considerably, but the fundamental purpose remains the same: ensuring data is available and accessible for all New Yorkers and making government operations more efficient and data-informed.

The Open Data Team compiles and updates this Technical Standards Manual in order to promote consistent policies and guidelines for the City’s ODCs, and consistent organization and standardization for the City’s public datasets. As more agencies publish data of various types, the standards laid out in this document become more essential for consistency across datasets. Our hope is that issuing these standards publicly not only supports agencies in participating in the NYC Open Data program, but improves the quality of datasets, helps New Yorkers to better understand the process and structure that the Open Data Team uses, and allows everyone to provide feedback to further the development of open data as a valuable public resource.

In this September 2022 update, the Technical Standards Manual is reformatted to highlight how policies, guidance, and technical standards apply throughout the entire lifecycle of a dataset: from identification, to publication, to ongoing maintenance. In moving to this new format, the Open Data Team intends to make more frequent updates to the Technical Standards Manual, ensuring the document properly reflects current best practice. As part of our next update, we will be adding more information about the role of the Open Data Team, further details on data publishing standards, and additional guidance on creating and updating data documentation.

2. Roles & Responsibilities

2.1 – Chief Open Platform Officer (COPO)

Per Section 20-f of the New York City Charter, the director of the Office of Data Analytics serves as both NYC’s Chief Analytics Officer and Chief Open Platform Officer. The COPO has the following responsibilities:

- Overseeing the operation and ongoing evolution of the City’s Open Data Team and program
- Issuing policies to improve the City’s capacity to identify, publish, and maintain datasets
- Coordinating with City executive leadership on Open Data governance and technical controls
- Publishing the annual Compliance Plan

2.2 – Open Data Team

As of 2022, the New York City Open Data Team is part of the Office of Data Analytics at the Office of Technology and Innovation (OTI). The team’s responsibilities include:

- Setting citywide strategy and standards for Open Data, including around data quality, documentation, public engagement, and implementation of the NYC Open Data Law
- Administering the network of Open Data Coordinators (ODCs) that exists across City government, including convening and training this group
- Supporting agencies and their designated Open Data Coordinators with the identification, structuring, documentation, publication, updating/automation, and promotion of agency public data
- Operating nyc.gov/opendata, ensuring the platform that underpins this service is operating smoothly and managing the site’s content
- Promoting NYC Open Data as a resource for all New Yorkers, including trainings about its usage and applications
- Facilitating agency response to public inquiries about NYC Open Data
- Identifying and adopting best practices from other open data programs across the country and around the world

2.3 - City Agencies

Data made available through NYC Open Data is generally sourced from City agencies, offices, or commissions, with each dataset providing a window into a government operation. City agencies are core to the functioning of the Open Data program, as they are ultimately responsible – primarily through their appointed Open Data Coordinators (ODCs) – for identifying, publishing, and maintaining their assets that qualify as public datasets.

2.3.1 - Open Data Coordinator Requirement

The Open Data Law requires each city agency to appoint an Open Data Coordinator (ODC) who manages open data efforts for their agency. For more information about Open Data Coordinator responsibilities, see section 2.3.1.

Should a designated ODC leave the role for any reason, agencies are required to appoint a new ODC. The agency should contact the Open Data Team regarding any ODC personnel changes to inform the Open Data program's record-keeping and enable onboarding of the new ODC into the role. See section 2.3.5 for more information on transfer of ODC knowledge.

2.3.2 - Data Ownership

While the Open Data Team facilitates the publication and standardization of open data, all data and datasets remain the responsibility of the agency from which they originate. Upon publication to NYC Open Data, datasets become a public resource available to anyone, without restriction or licensing requirements (see NYC Open Data's [Terms of Use](#) for more information).

Agencies are responsible for ensuring their data has been reviewed internally by appropriate agency management for confidentiality, privacy, security, and similar concerns, consistent with the Open Data Law, before the data is published. For more information on vetting datasets, see [section 4.1](#). For more information on privacy requirements, see [section 4.5](#). Agencies are also responsible for updating data according to the frequency identified in the dataset's metadata.

Data administered on behalf of other agencies

Data administered on behalf of other agencies shall be maintained and supported by the agency that submits the data for publication. All updates, structural changes, and enhancements to the dataset will need to be coordinated by the agency administering the data for public use.

2.3.3 - Open Data Support and Service

Agencies should take appropriate steps to ensure continuity of service when using NYC Open Data to host their data. Open Data is not intended to be the sole source of operation-critical data sharing or hosting, or act as a substitute for an agency's data systems and utilities.

NYC Open Data is not supported 24/7 and should not be used by agencies as a critical dependency for operational business. The Open Data Team assumes no responsibility for supporting or prioritizing operationalized data on NYC Open Data.

2.3.4 - Technology Systems

When considering any purchase, implementation, or reorganization of a technology system or database, agencies should ensure that compliance with the Open Data Law is taken into consideration. Agencies should be mindful of storing data in a machine-readable format, capturing data in a consistent manner, and ensuring that data can be easily extracted. Data in agency systems must be evaluated for inclusion on Open Data.

2.4 - Open Data Coordinators (ODCs)

2.4.1 - Responsibilities of the ODC

An Open Data Coordinator (ODC) is the representative of their agency in the NYC Open Data program and the public touchpoint for their agency's open data. In addition to being fluent with data, ODCs should be excited

about government transparency, well-networked within their agency, and deeply familiar with their agency's business practices.

ODCs are tasked with the following responsibilities:

1. Identifying datasets and enabling their delivery to NYC Open Data, including as-needed updates to the underlying data and associated documentation/metadata
2. Developing their agency's portions of the annual citywide compliance report
3. Addressing feedback from users of Open Data
4. Increasing awareness and strategic use of their agency's datasets through public outreach and presentations

ODCs are expected to regularly seek out new datasets for publication to NYC Open Data. To do that effectively, they need to stay up to date with the goals and routine operations of their agency. Here is a brief list of steps that ODCs should regularly take to stay informed about their agencies and identify new datasets:

1. Review the goals and indicators in their agency's chapter of the Mayor's Management Report (MMR)
2. Follow the legislative requirements and public commitments of their agency
3. Scan their agency's public communications for reports, statistics, charts, graphs, and other potential sources of new Open Data
4. Meet with agency divisions to identify new data owners and data systems
5. Meet with staff responsible for maintaining their agency's website to ensure that data underlying dashboards, maps, and other publicly available digital tools are reviewed and prioritized for Open Data publication
6. Review received FOIL requests where data has been provided to ensure the same information is published on Open Data

2.4.2 - Seeking out Open Data Stakeholders

ODCs should seek to develop good relationships with relevant stakeholders in their agency, both to develop support for open data within the context of their agency, and to be able to leverage stakeholder subject-matter expertise and resources to make their work more productive. Some typical stakeholders for open data are listed below:

A. Agency Head's office (Commissioner or equivalent)

By law, Agency Heads appoint the ODC for their agency. Open Data provides one of the most revealing looks into an agency's operations, and is frequently reviewed by advocates, elected officials, and the media. In many cases, Open Data is the first time that formerly internal datasets are shared with the public. ODCs should meet with their Agency Head to share the importance of this work, to inform them of plans for upcoming releases, to communicate legal requirements around the publication of Open Data, and to ensure Open Data is receiving sufficient support from other agency staff.

B. General Counsel/FOIL Officer/Agency Privacy Officer

General Counsels, Agency Privacy Officers (APOs), and Freedom of Information Law (FOIL) Officers are invaluable resources for ODCs. At a very minimum, ODCs should review each new dataset with their agency counsel and/or APO prior to release. The direct input of these stakeholders is also helpful for completing the Annual Open Data Compliance report (see [section 3.2](#)). More broadly, the expertise of this group on personally identifiable information (PII) in specific agency contexts, legislative mandates/restrictions, and FOIL will be helpful on an ongoing basis.

C. Data owners

Data owners are the agency subject-matter experts who create, manage, and/or use a specific dataset or system in the course of their day-to-day work. Working with data owners is essential for ODCs to understand the context in which each dataset is created – which is critical for structuring datasets logically, writing useful and appropriate documentation, and effectively responding to public inquiries. ODCs should meet and work with data owners regularly to prepare datasets and their corresponding metadata for publication to NYC Open Data.

D. Data or data analytics staff

While some ODCs also work in data or data analytics roles, the work of any ODC can help to improve an agency's data, so staff who do related work will make for natural collaborators. Though some datasets are ready to publish without much ODC effort, many require the restructuring of tables, reconciliation of fields, vetting of values, and creation of documentation. These same tasks are often required to make data usable for a data science or data analysis project. ODCs should talk with data analysts, data engineers, and data scientists to find areas of overlapping work and understand compelling internal uses for making the quality and documentation improvements required to make data publicly available. Given their wide-ranging use of agency data, these staff can also be helpful as stand-ins for members of the public since they will often perform similar analyses and ask similar questions as a member of the public might.

E. Information Technology

Public datasets are often captured in the regular course of business, and can be found within various systems, databases, and applications. ODCs should work closely with IT staff to help identify and access this data. Additionally, IT staff can help in the setup of processes to automatically publish data from an agency system directly to Open Data.

F. Communications/External Affairs

Communications professionals are often the most keenly aware of which issues or initiatives are a priority for their agency at any given time. Agency communications teams are a good resource for ODCs to receive help with writing easily understood data dictionaries, publicly announcing data releases, meeting agency civic engagement commitments, responding to public inquiries, drafting public facing commentary, and advising on intra-agency coordination.

2.4.3 - ODC Convenings

The Open Data Team will hold mandatory ODC convenings, generally once per quarter. These meetings are to share new information about the Open Data program, allow ODCs to connect and collaborate with each other, conduct specialized training, and engage external stakeholders and the public.

2.4.4 - ODC Training

Training in the form of an “Office Hours” session with the Open Data Team is available upon request through an email to opendata@oti.nyc.gov. This session can cover any a broad topic – like the Open Data Coordinator role itself, or a specific one – like reviewing or troubleshooting a specific dataset.

The Open Data Team will also organize group trainings to explain program changes or initiatives in greater detail and with more discussion than what's possible during an ODC convening. Depending on the topic(s) covered, some trainings will be mandatory for all ODCs or mandatory for new ODCs who have not previously

been trained.

2.4.5 - Transfer of Knowledge

To facilitate ODC transitions, ODCs must create and maintain proper documentation of their work. At a minimum, this documentation should include:

- Details about datasets scheduled for future publication
- A listing of source data systems and contacts
- Technical details about datasets that the agency has already made public on Open Data

For more information about this process, email the Open Data Team at opendata@oti.nyc.gov.

3. Identifying Datasets

3.1 - What is a Dataset?

NYC's [Open Data Law](#) sets operational definitions for which agency assets are considered 'data' and 'public datasets' for the purposes of the City's Open Data program. The data of New York City agencies that falls under the definition of 'public data' is **open by default**. Agencies must provide a compelling justification why data assets that meet the definition of 'public datasets' should not be published on NYC Open Data.

Understanding how these definitions relate to the information used in and created by agency operations is crucial for ODCs as they work to identify new datasets for publication to NYC Open Data.

'Data' is defined as:

"final versions of statistical or factual information:

- (1) in alphanumeric form reflected in a list, table, graph, chart or other non-narrative form, that can be digitally transmitted or processed; and*
- (2) regularly created or maintained by or on behalf of and owned by an agency that records a measurement, transaction, or determination related to the mission of an agency."*

Data does not include:

"information provided to an agency by other governmental entities, nor shall it include image files, such as designs, drawings, maps, photos, or scanned copies of original documents, provided that it shall include statistical or factual information about such image files and shall include geographic information system data."

Additionally, the definition of 'public dataset' in the Open Data Law excludes datasets for which any of the following conditions hold:

- The dataset contains information designated as sensitive, private, or confidential as defined by the [Citywide Data Classification Policy](#).
- The Agency Privacy Officer (APO) determines identifying information contained in the dataset may not be disclosed pursuant to the City’s Identifying Information Law; the dataset contains information that is exempt pursuant to the [Public Officers Law](#), or any other provision of a federal or state law, rule or regulation or local law.

A plain-language summary is below:

‘Public data’ is...	‘Public data’ is not...
<ul style="list-style-type: none"> • Machine-readable • Regularly maintained by a City agency, even if the source data is not owned by that agency (for example, an agency analysis of US Census data) 	<ul style="list-style-type: none"> • Anything owned by another entity/agency (for example, US Census data underpinning an analysis that an agency runs) • A one-off report or analysis • Scanned documents and photos • Sensitive or Private Information • Information that is exempt pursuant to the Public Officers Law, or any other provision of a federal or state law, rule, or regulation

3.1.1 Partial Publication of Public Datasets

An agency does not necessarily need to withhold an entire dataset from publication if that dataset contains some information that an agency withholds under the aforementioned laws. Many datasets on NYC Open Data are derived from internal-only equivalents that contain private or sensitive information. ODCs should consult with their agency’s legal counsel or, in the case of identifying information, their Agency Privacy Officer, to determine whether their agency may release a modified version of the dataset.

3.1.2 Currency of Public Datasets

What was once a public dataset is not always a public dataset. As agency missions evolve over time, new programs and data systems may replace older ones. Agencies are required to review the status of their data over time and maintain currency with the public disclosure requirements outlined in the Open Data Law and guidance in this manual.

3.1.3 Voluntary Publication of data that does not meet the “Public Dataset” definition

While agencies are not *required* to publish datasets which do not meet the legal definition of a “public dataset,” they are still welcome to voluntarily publish these datasets on NYC Open Data. For example, a dataset may be summarized in a report, and the table used to generate that report might be voluntarily published in addition to the underlying source data.

Nothing in the Open Data legislation, policies, or standards prohibits an agency from voluntarily disclosing information not otherwise defined as a public dataset, nor prohibit an agency from making such voluntarily disclosed information accessible through NYC Open Data. As with every dataset release, Open Data Coordinators are required to review whether voluntarily disclosed information contains identifying information with their Agency Privacy Officer, who must approve the disclosure pursuant to the Identifying Information Law.

3.2 - Annual Compliance Process

The Open Data Law requires OTI to submit a compliance plan to the Mayor and the City Council describing public datasets under the management of each agency, and a prioritized plan to make those datasets available on NYC Open Data. The first compliance plan was submitted in July 2013, and annual updates are now required by the Open Data Law on or before September 15 of each year. The Open Data Team will supply a template to each agency to facilitate development of that agency's portion of the compliance plan and updates to their prior commitments. Preparing compliance plans requires close cooperation between agencies and the Open Data Team.

3.2.1 - Components of the Annual Compliance Plan

The citywide annual compliance plan currently includes the following components:

- A summary description of all public datasets under the control of each agency.
- Prioritization of public datasets for inclusion on NYC Open Data, and a timeline for their publication or update.
 - For datasets that cannot be made available on the single web portal, the plan shall clearly state the reasons why.
- A summary of FOIL metrics that details the number of the FOIL responses that included the release of data, a listing of data shared via FOIL that are not available on NYC Open Data, and whether any datasets were voluntarily published on NYC Open Data in response to a FOIL request.
- A summary of all the public datasets maintained by the agencies on their websites and whether these datasets are posted or scheduled for publication on NYC Open Data.
- A summary of civic engagement activities that each agency has completed during the past year.

3.2.2 - Compliance Process Resources

The Open Data Team will be in frequent contact with agencies throughout the compliance process. ODC trainings (see [section 2.3.4](#)) will also focus on ensuring that ODCs understand the requirements of the annual compliance process. The Open Data Team will regularly update the [ODC Resource Page](#) so that relevant content remains accurate and actionable.

4. Preparing Datasets for Publication

4.1 - Prioritizing Datasets

Agencies and ODCs may find they have a large number of datasets that need to be published and must work to prioritize them appropriately. For the purposes of prioritizing the public release of datasets, agencies should consider the questions below:

Does the dataset...	Additional Detail
...create equitable access to public information and resources?	NYC Open Data is inspired by the vision of “Open Data for All” – that government data should be accessible to everyone.
...foster agency/interagency efficiency?	Agencies often make use of NYC Open Data datasets to inform decisions made in the course of business. Publishing data that’s used by other agencies creates efficiency in interagency data sharing.
...further the mission of the agency?	Publishing mission-related data is important for promoting transparency and enabling the public to keep government accountable
...improve public knowledge of the agency and its operations?	The public will have a better idea of what an agency’s purpose is and how it seeks to accomplish it by looking at the data it releases.
...increase agency accountability and responsiveness?	Public access to open data allows more people to understand how agencies gather information and measure their performance. This enables the public to potentially ask questions, make suggestions, and catch mistakes.
...respond to a need or demand identified by the public?	If the public has inquired about a particular topic, the release of related datasets should be prioritized over those that are of less public interest. Public inquiries related to a dataset may include correspondence with the agency, FOIL requests, or the creation of automated tools to scan an agency’s website for data.

4.2 - Components of Released Data

4.2.1 - Acceptable Data Formats

Datasets are required to be provided in a machine-readable format. Common file extensions include:

- .csv
- .xls
- .xlsx
- .shp
- .txt (must include an agreed upon delimiter)

Questions on acceptable data formats should be addressed to the Open Data Team. The Portable Document Format (.pdf) file type is not an acceptable dataset format.

4.3 - Geocoding

For datasets that include row-level address information, agencies must include both core address fields and core geospatial reference fields. These fields will appear on NYC Open Data according to a standard column naming convention.

4.3.1 – Core Address Fields

Agencies are responsible for separating core address information into five standard column fields:

- "NUMBER"
- "STREET"
- "UNIT"
- "POSTCODE/ZIPCODE"
- "BOROUGH"

Note that "NUMBER" and "STREET" can also be posted in a single "ADDRESS" field.

4.3.2 - Core Geospatial Reference Fields

Agencies are also required to include eight core fields of geospatial reference information in datasets that have street addresses:

- "LATITUDE" (minimum of 6 digits following the decimal point)
 - e.g.: "40.693588"
- "LONGITUDE" (minimum of 6 digits following the decimal point)
 - e.g.: "-73.985719"
- "COMMUNITY BOARD"
- "COUNCIL DISTRICT"
- "CENSUS TRACT"
- "BIN" (Building Identification Number)
- "BBL" (Borough-Block-Lot)

- “NTA” (Neighborhood Tabulation Area)

Agencies should use [Geosupport](#) if their datasets do not already contain the eight core geospatial reference fields. Geosupport is a publicly available tool maintained by the Department of City Planning that also serves as the City's geocoder of record. Street address data entered into Geosupport should return all required geospatial fields. Agencies can contact the Open Data Team (opendata@oti.nyc.gov) for assistance with the geocoding process.

Agencies with datasets that do not have address fields but include other locational data are encouraged, but not required, to populate as many core geospatial reference fields as possible using Geosupport.

4.4 – Metadata

4.4.1 - Required Metadata Elements

Metadata is contextual information that makes public datasets easier to use and understand. The Open Data Team collects metadata information about datasets from ODCs prior to dataset publication. Per the Open Data Law, all datasets on NYC Open Data must be accompanied by the following elements:

- Scheduled publication date
- Actual publication date
- Most recent update date
- URL
- Whether it complies with data retention standards (which mandates that row-level data be maintained on the dataset)
- Whether it has a data dictionary
- Whether it meets the geocoding standard, does not meet the geocoding standard, or is ineligible for the geospatial standard
- Whether updates to the dataset are automated
- Whether updates to the dataset “feasibly can be automated”

These required metadata elements are collected in the each dataset’s data dictionary and collated in the [Published Asset Inventory](#) dataset.

Note: It is the responsibility of the ODC to ensure metadata is accurate and up to date. When a dataset is updated, the metadata should also be reviewed and updated as necessary.

4.4.2 - Data Dictionaries

All datasets on NYC Open Data must be accompanied by a plain-language data dictionary, with the goal of making the data more understandable. The data dictionary accompanying each dataset should be thoroughly defined and prepared, as it helps users understand the data, agency specific terminology/codes, and gives context on how the data should be used.

Data dictionaries must be updated as datasets are updated, reflecting the latest version of the published dataset. Templates for data dictionaries can be found on the [ODC resource page](#).

4.4.3 – Data Quality

Agencies must ensure that their datasets are thoroughly reviewed for data quality and usability. In particular, new datasets must be reviewed for compliance with the [NYC Open Data Quality Standards](#) prior to publication. The standards and checklist laid out in this document will also be used by the Open Data Team to review new datasets in order to ensure quality and consistency across the inventory of NYC Open Data. Any issues identified in a dataset that cannot be reasonably addressed must be thoroughly documented in the dataset's data dictionary. Additionally, data quality issues that are identified or reported throughout the lifecycle of a dataset should be investigated, corrected, and/or documented.

4.5 - Privacy

New York City is committed to protecting the privacy of New Yorkers' identifying information. Since datasets sometimes contain identifying information, it is crucial for Open Data Coordinators to understand what information is considered "identifying information" and how the City requires agencies to address it, before sharing these datasets on NYC Open Data.

Agencies must treat any information published on NYC Open Data in accordance with all applicable federal, state, and local laws, including the City's Identifying Information Law and the Chief Privacy Officer's [Citywide Privacy Protection Policies and Protocols](#) (CPO Policies), which are described below. Open Data Coordinators should consult their agency counsel and Agency Privacy Officer (APO) before publishing new datasets to NYC Open Data, to ensure that any information disclosed is in compliance with all relevant laws and regulations.

The sections below explain where to find information regarding the City's Identifying Information Law and how ODCs can collaborate with APOs.

4.5.1 - The City's Identifying Information Law & CPO Privacy Policies

Sections [23-1201 through 23-1205](#) of the Administrative Code of the City of New York are collectively known as the Identifying Information Law (IIL), which governs how agencies collect, retain, and disclose identifying information. "Identifying information" is any information obtained by or on behalf of the City that, when used on its own or in combination with other information, can identify or locate an individual. Additionally, the [CPO Privacy Policies](#) explain additional agency responsibilities regarding the handling of identifying information including minimizing disclosure and privacy risk where possible. Agencies must follow the steps outlined in the IIL and CPO Privacy Policies when publishing public datasets to NYC Open Data.

4.5.2 - Privacy Resources for ODCs

The IIL and CPO Privacy Policies make agencies responsible for assessing each of their datasets, coordinating with their Agency Privacy Officer to determine if any of the data qualifies as identifying information under the IIL, and obtaining APO approval to disclose any identifying information as part of a dataset. The following resources will assist with that process:

Agency Privacy Officers (APOs)

An APO is a City employee appointed by their agency head to serve as a privacy resource for the agency, and has specific duties related to the agency's IIL compliance. City employees can view a [listing of APOs on the CityShare intranet](#). ODCs should have their APOs review all new datasets before publication to help determine whether it contains any identifying information. To ensure that information published on NYC Open Data is in accordance with the IIL's requirements, an ODC must obtain APO approval before including any identifying information in a public dataset.

The CPO has issued detailed guidance explaining how ODCs and APOs should work together to handle identifying information in public datasets. This guidance is the [CPO Privacy Policies](#).

The Chief Privacy Officer and the Office of Information Privacy

The City's Chief Privacy Officer is appointed by the Mayor to lead OTI's Office of Information Privacy (OIP). OIP supports the CPO in protecting New Yorkers' identifying information by, among other functions, advising City agencies and offices on matters related to privacy and identifying information, and implementing the IIL. ODCs may contact the CPO at PrivacyOfficer@OTI.nyc.gov with any questions or requests for assistance.

4.6 - Dataset Publishing

4.6.1 - Preparing for Data Publishing

ODCs should email opendata@oti.nyc.gov at least one month in advance of the desired release of a new dataset. All datasets will be reviewed by the Open Data Team prior to publication (see [section 4.4.3](#)) so any agency releasing time-sensitive or complex data should contact the team as early as possible. This guidance applies to datasets that have been scheduled for publication in the Open Data Compliance Plan and those released on an ad-hoc basis. This email should include the following information to facilitate the dataset's publication:

- A sample (or final) dataset for testing
- A fully completed data dictionary (see [section 4.4.2](#))
- Any additional relevant information pertaining to the dataset

In instances where file size, format, or other factors prevent sharing via email, alternative methods for file sharing can be determined.

4.6.2 - Methods of Publication

Depending on file size, frequency of update, file format, and other factors, the Open Data Team will work with agencies to make a determination on how a dataset should be published. Outlined below are the most common publishing methods:

Manual publishing - Datasets may be emailed to the Open Data Team and published to NYC Open Data by an Open Data Team member. This is usually a fast and simple method of publishing, but burdensome for datasets with frequent updates and/or large file sizes.

Automated publishing - Automated publishing is the process by which data is pushed to NYC Open Data without human intervention. Where possible, automation is the preferred method of data publication, but usually requires coordination between technical staff at OTI and at the publishing agency. The following methods of automation are the most common:

Database access - If an agency stores its data in a database, the Open Data Team can typically connect to this source and 'automate' publishing to NYC Open Data

File folder access – Agency data files are placed in a secure folder for the Open Data Team to access and publish to NYC Open Data

Note: It is the responsibility of an ODC to see the publishing of a dataset through to completion in a timely manner. This is especially relevant for datasets prioritized in the compliance plan or those with publishing deadlines. Providing access to a database or file storage location does not qualify as publishing, nor does it relieve an agency of its responsibility to publish a dataset.

5. Maintaining Datasets

Publishing data is not the end of the lifecycle of a dataset. Every dataset published on NYC Open Data should be regularly monitored and maintained to ensure accuracy, currency, and accessibility. The basic standards for maintaining datasets on NYC Open Data can be found below.

5.1 - Consolidation

To maintain high standards for data usability and searchability, agencies should generally seek to consolidate multiple datasets that contain the same data collected from different time periods or geographic locations. For example, a series of datasets representing an agency's inventory on an annual basis should be consolidated into one dataset with a year column.

The Open Data Team reserves the right to direct any agency to combine or consolidate related datasets into a single, comprehensive dataset.

5.2 - Timely Updates

ODCs are responsible for ensuring their datasets are current and updated in accordance with the stated dataset update frequency in the metadata. Updating a dataset generally follows the same process as the initial publication. Manual dataset updates take place by contacting the Open Data Team.

For datasets published on both an agency website and NYC Open Data, the following update timelines apply:

- Manually updated datasets that are also shared publicly on agency websites must be updated no later than 10 days after the agency website is updated.
- Automatically updated datasets that are also shared publicly on agency websites must be updated no later than 1 day after the agency website is updated.

An ODC should contact the Open Data Team about any upcoming changes to the schema/column structure of a dataset before those changes are implemented, as these changes have the potential to impact dataset functionality and will also need to be logged in the dataset's data dictionary. This is especially important for automated datasets for which development work needs to be coordinated with the Open Data Team.

For assistance tracking updates to datasets, ODCs can use the [Open Data Asset Inventory](#) as a resource.

5.3 - Platform-wide Updates

OTI is responsible for maintaining and protecting NYC Open Data from unlawful abuse or attempts to damage or impair use of the website. To that end, OTI reserves the right to implement any necessary security measure to ensure the health of NYC Open Data.

Additionally, the Open Data Team reserves the right to perform platform-wide metadata updates to NYC Open Data (at times without advance notice) to adhere to legislative and reporting requirements, or to improve user experience and metadata documentation.

5.4 – Data Interoperability

Many agencies maintain similar datasets referring to the same subject matter (for example, Permits, Inspections, Requests for Service, etc.). To the extent possible, agencies should strive to make these similar datasets interoperable by publishing similar fields and making reference to related content and context. Additionally, the Open Data Team can provide guidance around the potential use of shared data specifications.

5.5 - Dataset Retention and Archiving

The Open Data Law mandates the preservation of New York City historical record as represented in NYC Open Data. This standard establishes requirements and guidelines for retaining row data and archiving datasets on NYC Open Data.

The Open Data Team must evaluate each dataset on NYC Open Data with its agency owners to determine a retention and archiving solution. This solution should both maximize usability of the data for all users and also allow users to perform analyses most suited to the dataset type and content. The archiving process for each dataset will be specified in its metadata.

Datasets with an exceptionally high rate of growth may be subject to archival. In these instances, older data may be archived outside of NYC Open Data in a publicly accessible archiving directory.

For more information and assistance on retention and archiving requirements, ODCs should contact the Open Data Team.

6. Open Data and the Public

6.1 - Open Data Help Desk

NYC Open Data includes an online help desk to solicit feedback from the public. The Open Data Team and ODCs monitor this feedback channel during standard business hours to ensure inquiries are appropriately responded to. Through this forum, users can request new datasets, ask a question about an existing dataset, report errors, and suggest a partnership – among other options. ODCs should reach out to the Open Data Team to ensure they are making optimal use of the help desk. See [section 6.4](#) for further guidance and resources on responding to public inquiries. The help desk can be accessed via <https://opendata.cityofnewyork.us/engage/>.

6.2 - Responding to Public Dataset Requests

Users are encouraged to make requests for data they are unable to find on NYC Open Data. Per the Open Data Law, OTI must provide an initial response to each such request within two weeks of receipt, and refer the inquiry to the appropriate agency, who must respond within two months. The [Public Dataset Requests dataset](#) contains the complete list of all such requests and tracks their current status. Additional guidance on responding to public inquiries can be found in [section 6.4](#).

The Open Data Team requires that responses to dataset requests are clear and conclusive and include the following information at a minimum:

- *If the requested data qualifies as a public dataset:*
 - Confirm this fact
 - Provide the user a date by which the data will be released on NYC Open Data
 - Add the dataset to the agency's Annual Compliance Plan
- *If the data requested does not qualify as a public dataset:*
 - Confirm this fact
 - Provide the user with a reason why the data does not qualify as a public dataset

Agencies should not direct users to submit Freedom of Information Law (FOIL) requests in response to a dataset request.

6.3 - Dataset Questions and Errors

If a public inquiry asks a question, or identifies a potential error within a dataset, the appropriate agency should provide the user with a response within two weeks of the submission of the inquiry. Any data quality issues that are not able to be resolved with a change to the dataset should be directly addressed in the dataset's data dictionary.

6.4 - Guidance for Responding to Public Inquiries

- Agencies are expected to confirm that an inquiry is correctly routed to them within 1 week.
- Agencies are expected to respond to inquiries within 2 weeks.
- Agencies are expected to include a greeting in their response.

- Agencies are expected to utilize an email signature that indicates the name of their agency.
- If a user submits a follow-up questions to an inquiry, the request should remain open until all subsequent questions are addressed.

More information can be found in the [NYC Open Data Help Desk Policies](#) for Open Data Coordinators.

Appendix A - Definitions

Term	Definition
Agency	An agency, per NYC’s Open Data Law, means an office, administration, department, division, bureau, board, commission, advisory committee or other governmental entity performing a governmental function of the City of New York.
Agency Privacy Officer (APO)	An APO is a City employee appointed by their agency head to serve as a privacy resource for the agency, and has specific duties related to the agency’s Identifying Information Law compliance. To ensure that information published on NYC Open Data is in accordance with the IIL’s requirements, an ODC must obtain APO approval before including any identifying information in a public dataset. In addition, ODCs should have their APOs review all new datasets before publication to help determine whether it contains any identifying information.
City	The City of New York is referred to interchangeably as New York City, NYC, or the City.
Data	<p>Per the Open Data Law, data are final versions of statistical or factual information</p> <ol style="list-style-type: none"> 1) In alphanumeric form reflected in a list, table, graph, chart or other non-narrative form, that can be digitally transmitted or processed; and 2) Regularly created or maintained by or on behalf of and owned by an agency that records a measurement, transaction, or determination related to the mission of an agency <p>Data shall not include information provided to an agency by other governmental entities, nor shall it include image files, such as designs, drawings, maps, photos, or scanned copies of original documents, provided that it shall include statistical or factual information about such image files and shall include geographic information system (GIS) data.</p>
Data Dictionary	Required plain-language documentation that is shared with each NYC Open Data dataset. See section 4.4.2 .
Dataset	A named collection of related records on a storage device, with the collection containing individual data units organized or formatted in a specific and prescribed way.
Metadata	Contextual information that makes the public datasets easier to understand and use. See section 4.4 .
Office of Technology and Innovation (OTI)	<p>The Office of Technology and Innovation consolidates the previous Department of Information Technology and Telecommunications (DoITT), Mayor’s Office of Data Analytics (MODA), Mayor’s Office of Information Privacy (MOIP), and NYC Cyber Command into a single agency.</p> <p>The NYC Open Data Team is located within OTI’s Office of Data Analytics.</p>

Open Data Coordinator (ODC)	An Open Data Coordinator. An ODC is a City employee designated by an agency to act as the coordinator of open data efforts for that agency. See section 2.3 for more information.
Open Data Law	New York City's first law on Open Data was passed in 2012. This law has since been amended and bolstered. NYC's laws around Open Data are consolidated in Title 23, Chapter 5 of New York City's Administrative Code , also referred to as the Open Data Law.
Public dataset	A comprehensive collection of interrelated data that is available for inspection by the public in accordance with any provision of law and is maintained on a computer system by, or on behalf of, an Agency, excluding any data to which an Agency may deny access pursuant to the Public Officers Law or any other provision of law or any federal or state rule or regulation. See the Open Data Law for a full definition.

Appendix B - Contacting the Open Data Team

Members of the public with questions about this document, a particular dataset, or the NYC Open Data program should contact the [Open Data help desk](#) for assistance.

City Open Data Coordinators can [submit a request](#) for the Open Data Team or email the team at opendata@oti.nyc.gov.