

NYC OpenData



Open Data Policy and Technical Standards Manual

November 2016

**City of New York
Department of Information Technology and Telecommunications**

TABLE OF CONTENTS

1.0	Executive Summary.....	2
2.0	Introduction.....	2
2.1	Background, Authority, and Governance.....	2
2.2	Definitions.....	3
3.0	City Policies.....	6
3.1	Governance and Oversight.....	6
3.2	Technology Projects.....	8
3.3	Tools.....	8
3.4	Data Sets.....	8
4.0	City Standards.....	9
4.1	Data Sets.....	9
4.2	Data Set Publishing.....	12
4.3	Maintenance.....	13
4.4	Ownership, Responsibility, and Retention.....	14
4.5	Exemption from Public Access.....	15
5.0	City Guidelines.....	17
5.1	Data Set Selection.....	17
5.2	Guidelines for the Division of Large Datasets.....	18
6.0	Public Policies.....	19
6.1	Terms of Use.....	19
6.2	Ensuring Continued Public Access to Data Sets.....	19
6.3	Protecting the NYC Open Data Portal.....	19
6.4	Improving Service Delivery.....	19
7.0	Appendix A: NYC Administrative Code (Title 23, Chapter 5).....	21
8.0	Appendix B: End Notes.....	27

1.0 Executive Summary

On March 7, 2012, Mayor Bloomberg signed Local Law 11 of 2012, which amends the administrative code to mandate that qualifying City-managed digital data be made available to the public-at-large through a single web portal. By September 2012, the City's Department of Information Technology and Telecommunications (DoITT) is to prepare and publish a technical standards manual for City agencies, setting forth policies and standards for open data.

This document is the Technical Standards Manual (TSM) that defines the Citywide policies, standards, and guidelines required to support the legislation.

2.0 Introduction

2.1 Background, Authority, and Governance

Local Law 24 of 1995 established DoITT as "New York City's information technology and telecommunications agency."

Chapter 48 §1072(a) of the New York City Charter established the authority of DoITT by assigning powers and duties "to plan, formulate, coordinate and advance information technology and telecommunications policies for the city."¹

Under the administration of Mayor Bloomberg, DoITT published "PlanIT" in 2007 to drive IT development, improve and transform service delivery, and make City government more open by expanding the accessibility, transparency, and accountability of City government to its residents, businesses, visitors, and employees.

In late 2009, DoITT launched the NYC DataMine, which served as the City's first effort to identify data sets and make them available to the public for download. In February 2010, DoITT published its "30-Day Report: Enabling the Connected City,"² which committed to establishing Citywide open data policies to improve access to public information.

Executive Order No. 140 of 2010 directed DoITT to "be responsible for establishing and enforcing Citywide IT policies and for ensuring that such policies are aligned with the City's business needs and investments, as well as the individual business needs of each agency."³

In October 2011, DoITT launched the [NYC Open Data Portal](#), advancing beyond simple downloads to fully interactive viewing, searching, and visualizing of City data. Additionally, the NYC Open Data Portal provides sophisticated feedback mechanisms, metrics, and Representational State Transfer (REST) Application Programming Interfaces (APIs).

In 2012, the New York City Council passed Local Law 11, which was subsequently signed by Mayor Bloomberg, to make City data available online using open standards to ensure that City government is transparent, effective, and accountable to the public

Local Law 11 streamlines intra-governmental and inter-governmental communication and interoperability, permits the public to assist in identifying efficient solutions for government, promotes innovative strategies for social progress, and creates economic opportunities. Local

Law 11 directs DoITT to produce the TSM for the publishing of public data sets in raw or unprocessed form. These public data sets are to be made available on the Internet, accessible through a single web portal that is linked to *NYC.gov* (or any successor website maintained by, or on behalf of, the City of New York).

The TSM identifies the reasons why each technical standard was selected and the types of data for which it is applicable, and may recommend or require that data be published in more than one technical format. The TSM includes a plan to adopt or use a web application programming interface that permits application programs to request and receive public data sets directly from the single web portal. DoITT will update the TSM as necessary, as required by the law.

2.2 Definitions

Several terms in the TSM have specific meanings as defined in the Local Law 11 and these are included in the list below. When a term is defined in the law, it is so indicated.

Term	Definition
Agency	<p>An Agency, when the term is capitalized, 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. (Per Local Law 11 definition)</p> <p>When not capitalized, the term agency means either an agency in another jurisdiction, such as federal or state, or an exact quotation from another published City document, such as a City Local Law or Executive Order.</p>
City	<p>The City of New York is referred to interchangeably as New York City, NYC, or the City.</p>
Citywide Standard	<p>An industry standard or <i>de facto</i> standard which is adopted and/or mandated by DoITT to be used in relation to the City’s information technology systems and services. DoITT is directed by the City Charter and subsequent Executive Order to establish Citywide standards that apply to all New York City Agencies.</p>

Term	Definition
Data	<p>(Per Local Law 11 definition)</p> <p>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 Set	<p>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, often in tabular form, and accessed by a specific access method that is based on the data set organization.</p>
DoITT	<p>The New York City Department of Information Technology and Telecommunications. The term “department,” as used in Local Law 11, is defined as “the department of information technology and telecommunications or any successor agency.”</p>
Metadata	<p>Contextual information that makes the public data sets easier to understand and use.</p>
ODC	<p>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.</p>
Open Standard	<p>(Per Local Law 11)</p> <p>A technical standard developed and maintained by a voluntary consensus standards body that is available to the public without royalty or fee.</p>

Term	Definition
Public Data Set	<p>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.</p> <p>The entry above is only a summary. Please consult Local Law 11 for the full authoritative definition.</p>
Requirement	<ol style="list-style-type: none"> 1. A condition or capability needed by a stakeholder (such as an entity, member of the public, or the City) to solve a problem or achieve an objective. 2. A condition or capability that must be met or possessed by a solution or solution component to satisfy a contract, standard, specification, or other formally imposed document. 3. A singular documented need of what a particular product or service should be or perform.
Technical Standard	<p>(Per Local Law 11 definition)</p> <p>A standard that includes:</p> <ol style="list-style-type: none"> 1. The common and repeated use of rules, conditions, guidelines, or characteristics for products or related processes and production methods, and related management systems practices; and 2. (i) The definition of terms; (ii) classification of components; (iii) delineation of procedures; (iv) specifications of dimensions, materials, performance, designs, or operations; (v) measurement of quality and quantity in describing materials, processes, products, systems, services, or practices; (vi) test methods and sampling procedures; or (vii) descriptions of fit and measurements of size or strength.

Term	Definition
Voluntary Consensus Standards Body	(Per Local Law 11 definition) A domestic or international organization that develops and maintains a technical standard. A Voluntary Consensus Standards Body utilizes a transparent deliberative process, permits the participation of any party, and achieves general consensus, although not necessarily unanimity, of the participating parties, including a process for attempting to resolve any differences in viewpoint.

3.0 City Policies

In accordance with Local Law 11, DoITT has established the following policies. Citywide open data policies apply to all City Agencies as defined by this law.

3.1 Governance and Oversight

3.1.1 Chief Open Platform Officer

DoITT must designate a Chief Open Platform Officer (COPO). The COPO will act as the overall manager for information technology initiatives as they pertain to open government, and making machine-readable data and transactional services available for public use. The COPO will have the following responsibilities:

- Oversee the operation and ongoing evolution of the City's Open Data program, and manage DoITT's Open Data team;
- Engage with Agency Open Data Coordinators (as described below) and with executive leadership to strategically plan, and provide support for, the publishing of data sets;
- Coordinate the delivery of transactional services and the development of related developer ecosystems;
- Publish the annual Compliance Plan (as mandated by Local Law 11);
- Determine, operationalize, and support the public engagement strategy, including outreach and public inquiries;
- Collaborate with other local, state, and federal government entities on open government strategies and data specifications; and
- Ensure the appropriate governance and technical controls exist as required by the City's Chief Information Security Officer.

Pre-existing processes such as Freedom of Information Law requests or public hearing governance will not be part of the COPO's responsibilities.

3.1.2 Open Data Coordinators

Each Agency must designate one of its employees as the Agency’s Open Data Coordinator (ODC). Each ODC will act as the coordinator of open data efforts for that Agency. Each ODC will have the following responsibilities:

- Identify and enable the delivery of data sets to the single web portal;
- Develop the Agency portions of the citywide compliance report;
- Liaise with the Chief Open Platform Officer and the DoITT team supporting the Open Data program; and
- Address feedback from the public for the Agency’s public data sets.
- Perform public outreach and presentations with the aim of increasing strategic use of the City’s data sets.

The ODC should be well-connected to or a part of the following Agency teams, in order of preference:

- Analytics / Strategic / Management Planning;
- Web Content / Communications;
- Legal; and
- Information Technology.

3.1.3 Agency Compliance Plan

DoITT submitted a compliance plan to the Mayor and the City Council on September 9, 2013. Updated versions must be released annually, on or before July 15 of each year. DoITT will supply a template to each Agency to facilitate development of the compliance plan and updates. City agencies must cooperate with DoITT in the preparation of their plans. DoITT must make the final version of the compliance plan available to the public on the single web portal.

The compliance plan submitted to the Mayor and the City Council shall include a summary description of public data sets under the control of each Agency, prioritization of public data sets for inclusion on the single web portal, and a timeline for their publication or update. If a public data set cannot be made available on the single web portal by December 31, 2018, the plan shall clearly state the reasons why that set cannot be made available, and, to the extent practicable, the date by which the data will be available on the single web portal.

3.1.4 Open Data Dashboard

DoITT maintains a public Open Data dashboard to promote public and internal accountability. The dashboard tracks agency compliance plans and interim progress – specifically which data sets have been published through the NYC Open Data Portal or by direct public access. The dashboard also displays by Agency:

- Scheduled data set releases;

- Inventory of data sets released; and
- Data set releases behind schedule.

3.2 Technology Projects

3.2.1 New Technology Systems

All new City projects to implement information technology systems, excluding infrastructure, must include funding, staffing, and tasks to comply with Open Data legislation, policies, and technical standards.

"Infrastructure" refers to those projects that solely involve the procurement and deployment of desktop computers, servers, operating systems, telephones, routers, firewalls, networks, and similar hardware capabilities.

3.2.2 Existing Technology Systems

All operating information technology systems, excluding infrastructure, must be evaluated for data that is eligible for publication. For each system, the Agency that owns or prepares the data must create a plan to publish data to comply with the Open Data legislation, policies, and technical standards. An updated summary of the plan must be provided to DoITT for inclusion in each Compliance Plan.

3.3 Tools

3.3.1 Enterprise Open Data Platform

DoITT manages the NYC Open Data Portal, the City's public data set catalog. All public data sets must be identified and referenced on this site. The single web portal includes interactive features to solicit public feedback. Public feedback includes discussion of Open Data policies, standards, guidelines, data sets, and prioritization of data set releases. Whenever practicable, DoITT will ensure the platform uses open Internet standards and specifications.

3.3.2 Enterprise Data Publishing Tools

DoITT will provide and manage automation tools to extract data from City technology systems, applying transformations, and loading data into the NYC Open Data platform. Agencies should leverage these tools as much as possible.

3.4 Data Sets

3.4.1 Open By Default

All public data sets must be considered open unless they contain information designated as sensitive, private, or confidential as defined by the [Citywide Data Classification Policy](#), or 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. The Agency submitting the data set is required to review its status and maintain currency with public disclosure requirements.

3.4.2 Alignment to Open Specifications

The structure of published data sets should align to New York State, national, or international specifications, where they exist. Agencies may consult with DoITT regarding current specifications.

3.4.3 Metadata

Every Agency must create, publish, and maintain on the NYC Open Data Portal accurate metadata for each public data set as set forth in the City Standards for metadata in this document.

3.4.4 Maintenance

Every Agency must ensure that each public data set and associated metadata is kept current to the extent that the agency regularly maintains or operationally updates the public data set.

3.4.5 Third-Party Search Engines

All public data sets must be accessible to the public through the NYC Open Data Portal and accessible to third-party (Internet-based) search engines. In any exceptional case in which for technical reasons it is not feasible to host a public data set on the NYC Open Data Portal itself, the NYC Open Data Portal must provide a direct link to the public data set that is hosted elsewhere. An Agency may self-host the public data set, provided that the public data set is accessible through the link on the NYC Open Data Portal to the public and to third-party (Internet-based) search engines.

3.4.6 Language

Agency data sets, including metadata, are not required to be published in additional languages beyond those used at their sources. The City may opt to implement automated language translation of data sets in the future.

4.0 City Standards

4.1 Data Sets

4.1.1 Data Content

4.1.1.1 Data Types

Data sets may contain the following data types:

- Numbers, money, and percentages;
- Dates/times (ISO-8601 preferred);

- Booleans;
- E-mail addresses;
- URLs; and
- Location columns.

4.1.1.2 Address Data

For any dataset on the NYC Open Data Portal that includes row-level address fields, agencies must separate locational information into “core address” and “core geospatial reference” attributes. These attributes will appear on the Portal according to a standard column naming convention.

Agencies will be responsible for separating **core address** information into five standard column fields:

- “NUMBER”
- “STREET”
- “UNIT”
- “POSTCODE”
- “BOROUGH”

Agencies will also be required, with technical guidance from the Open Data team, to include eight standard column fields of **core geospatial reference** information:

- “LATITUDE”
- “LONGITUDE”
- “COMMUNITY BOARD”
- “COUNCIL DISTRICT”
- “CENSUS TRACT”
- “BIN” (Building Identification Number)
- “BBL” (Borough-Block-Lot)
- “NTA” (Neighborhood Tabulation Area)

If an agency cannot comply with these standards, Local Law 110 of 2015 requires it to report why it has failed to comply and how it plans to comply in the future. The Open Data team will work with noncompliant agencies to identify when an exemption is appropriate, how to report the reason for exemption, and what can be changed to ensure future compliance.

As per current Open Data policy, agencies will be alerted but are not required to change their database of record when records on the Portal are incorrect as a result of a data entry error.

Similarly, the Open Data team does not warranty the accuracy of data that is incomplete or in error as the result of Geosupport. However, the Open Data team will work with agencies to mitigate known issues, and regularly considers internal and public feedback for future updates to standards and technical tools.

4.1.1.3 Geocoding

We recommend agencies whose datasets do not already contain the eight *core geospatial reference* fields to use Geosupport, a publicly available tool that also serves as the City of New York's geocoder of record maintained by the Department of City Planning. *Core address* data entered into Geosupport can return all required *core geospatial reference* data. Agencies may geocode their locational data at the database level or the extraction level. Alternatively, agencies may elect to have the Open Data team establish an automated feed, in which datasets are passed through an ETL (Extract-Transform-Load process) where they are geocoded and uploaded directly to the Portal.

When a dataset is geocoded, its data dictionary must designate which attribute fields were reported directly by the agency, and which attribute fields were created by geocoding in order to meet these standards. Finally, 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.

4.1.2 Metadata

4.1.2.1 Data Dictionaries

As mandated by Local Law 107 of 2015, all datasets on the Open Data Portal must be accompanied by a plain language data dictionary, with the goal of making the data more understandable.

Outlined below are the minimum standards that must be adhered to:

- Agency name, dataset name, dataset description, and update frequency must all be provided
- Each column name should be listed and defined
- Where applicable and reasonable ,terms, acronyms, codes, and units of measure should be defined
- To the extent practical, a range of possible values should be included
- History of modifications to dataset format should be documented

Agencies may choose to provide additional information deemed relevant, including but not limited to, method of collection, relationship with or between other datasets, system of record, field lengths, etc.

Data dictionaries can be provided in a file format of an agencies choosing, but must include the above minimum requirements.

4.1.2.2 Standards for Data Citation

An Agency should include any preferred citation for a data set in the data set's metadata or supporting documentation.

4.2 Data Set Publishing

The Agency ODC should work closely with DoITT during the initial data set publishing process to identify the best technical approach to automate delivery to the public. The following mechanisms are supported:

4.2.1 Database Management System (DBMS) Access

For Agencies that require DoITT assistance to extract data from back-office systems, the Agency should provide read-only DBMS credentials for the necessary databases, tables, stored procedures, and/or views. The credentials should not permit access to tables, columns, or other entities that contain information that is not included within the definition of public data set because it is exempt from disclosure.

If the Agency operates a data warehouse, it should provide access to extract public data sets from the warehouse rather than the source operational system.

4.2.2 File Transfers

Agencies may choose to publish files to a location on the City intranet that DoITT staff or DoITT-managed automation tools can access. Specific details, such as location, formats, naming conventions, and sizing, should be discussed with DoITT.

4.2.3 Enterprise Service Bus

Agencies may leverage DoITT's Enterprise Service Bus (DataShare) to publish public data sets. This option may be especially desirable if DataShare already automatically transfers the data set.

4.2.4 Self-Hosting

In any exceptional case in which transaction volumes, data structure, technical barriers, or resource limits prevent hosting a public data set on the NYC Open Data portal itself, the NYC Open Data Portal must provide a direct link to the public data set that is hosted elsewhere so that the data set is accessible to the public through the NYC Open Data Portal. In such an exceptional case, an Agency may self-host the relevant public data set, provided that the public data set is accessible to the public through the link on the NYC Open Data Portal according to following standards:

- The agency must provide a single, unique, publicly accessible URL for each data set along with the data set-level metadata. This information will be made available in the NYC Open Data Portal’s catalog of data sets. It is strongly preferred that the URL be the location of the data set, and not an intermediate web page.
- The data set must be machine-readable and in one of the formats listed in the Public Standards section of this document. The Portable Document Format (PDF) is not permitted.
- The data set may be encapsulated in a single archive file (and optionally compressed) if it consists of multiple related files.

4.3 Maintenance

Data sets published on the NYC Open Data Portal must be maintained for accuracy, timeliness, and accessibility, as set forth below.

4.3.1 Data Set Content Updates

Agency ODCs are responsible for identifying an update frequency for each public data set as an element in its data set metadata, and for ensuring that their data set content updates are maintained and published according to the data set’s identified schedule or to the extent that the agency regularly maintains or updates the public data set.

4.3.2 Structural Changes (Fields, Data Types)

The ODC or Agency liaison must not modify existing data structure during normal updates to the data set. The number of data elements per record, name, format, and order of the data elements must be consistent with the originally-published version. The Agency ODC should notify DoITT prior to any structural changes to data sets.

4.3.3 Content Support

DoITT will contact the Agency ODC to obtain feedback or a direct answer to comments or inquiries from the public that relate to data set contents or supporting documentation. The Agency will provide DoITT with an expected timeframe to resolve the support inquiry as soon as possible. The Agency must then notify DoITT when the updates or corrections are ready for publication.

An Agency that proactively identifies defects or improvements related to its data set content or supporting documentation must notify DoITT prior to publication of any changes.

4.4 Ownership, Responsibility, and Retention

4.4.1 Ownership

Agencies retain ownership over the data sets that they submit. All data and data sets remain the property of the originating Agency and public users acquire no ownership rights to Agency data or data sets.

The data sets published on *NYC.gov* or the NYC Open Data Portal become a public resource available to anyone with access to the Internet. The public use of the data sets may include development of applications. In this case, the developers retain all intellectual property ownership in their applications, excluding the Agency data itself, whose ownership continues to reside with the Agency.

4.4.2 Responsibility

The Agency that owns the data set is responsible for all aspects of the quality, integrity, and security of the data set contents, as detailed below, and as subject to limitations on liability contained in Local Law 11. Agencies do not relinquish control of their data to DoITT when the data set is submitted for publication on the NYC Open Data Portal.

Agencies are responsible for ensuring that all of their submitted data has been reviewed by appropriate Agency management for confidentiality, privacy, security, and all other content limitation issues consistent with Local Law 11 before the data is submitted for publication. The Agency supplying the data is also responsible for maintaining records of information privacy status and public-disclosure requirements.

The Agency is responsible for updating its data according to the frequency identified in the data set metadata or to the extent that the agency regularly maintains or updates the public data set.

4.4.3 Retention and Archiving

On November 30, 2015, Local Law 106 of 2015 amended the Open Data Law of 2012 to mandate the preservation of the New York City historical record as represented by the City's official data. This standard establishes requirements and guidelines for retaining row data and archiving datasets on the Open Data Portal.

The Open Data team must evaluate each dataset on the Portal with the Agency that maintains the data 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. The Open Data team will periodically review public and internal feedback to determine when the approved retention and archiving process can be improved.

The Open Data team will modify inaccurate or non-public data that is represented in error in order to ensure accuracy of all records on the Open Data Portal.

4.4.3.1 Tabular Datasets

Each unique identified record at the row-level must never be removed from the Portal.

Examples of data retention protocol for tabular datasets include:

- Some datasets containing only “active” records that are updated frequently should have a complementary dataset for records that become “inactive,” which includes the date on which the records are archived.
- Some datasets refer to a roughly fixed set of data points (e.g., restaurants in the City) for which updated data is routinely collected. As new data is collected, new records should be added alongside existing records and include their date of creation.
- Some datasets on the Portal are not subject to regular updates and bear no risk of removed records, and therefore require no separate archiving protocol.

4.4.3.2 Geospatial Files

When geospatial files are updated, the most recent copy must be moved to an archive folder and indexed according to date of archive. The archive folder will be accessible through the up-to-date dataset on the Portal.

4.4.3.1 Non-data Files

When non-data files are updated, the most recent copy must be moved to an archive folder and indexed according to date of archive. The archive folder will be accessible through the up-to-date dataset on the Portal.

4.5 Exemption from Public Access

Public data to be made available per Local Law 11 does not include any data set to which an Agency may deny access pursuant to the [Freedom of Information Law \(FOIL\)](#) or any other provision of a federal or state law, rule or regulation or local law.⁴ (That notwithstanding, by itself, Local Law 11 *does not prohibit* Agencies from releasing such FOIL-deniable data.)

Records deniable under FOIL are those that:

- (a) are specifically exempted from disclosure by state or federal statute;
 - (b) if disclosed would result in an unwarranted invasion of personal privacy;
 - (c) if disclosed would impair present or imminent contract awards or collective bargaining negotiations;
 - (d) are trade secrets or are submitted to an agency by a commercial enterprise or derived from information obtained from a commercial enterprise and which if disclosed would cause substantial injury to the competitive position of the subject enterprise;
 - (e) are compiled for law enforcement purposes and which if disclosed would:
 - i. interfere with law enforcement investigations or judicial proceedings;
 - ii. deprive a person of a right to a fair trial or impartial adjudication;
 - iii. identify a confidential source or disclose confidential information relative to a criminal investigation; or
 - iv. reveal criminal investigative techniques or procedures, except routine techniques and procedures;
 - (f) could if disclosed endanger the life or safety of any person;
 - (g) are inter-agency or intra-agency communications, except to the extent that such materials consist of:
 - i. statistical or factual tabulations or data;
 - ii. instructions to staff that affect the public;
 - iii. final agency policy or determinations; or
 - iv. external audits, including but not limited to audits performed by the comptroller and the federal government;
 - (h) are examination questions or answers that are requested prior to the final administration of such questions;
 - (i) if disclosed, would jeopardize an agency's capacity to guarantee the security of its information technology assets, such assets encompassing both electronic information systems and infrastructures;
 - (j) are photographs, microphotographs, videotape or other recorded images prepared under authority of section eleven hundred eleven-a of the vehicle and traffic law (this exemption will be repealed effective December 1, 2014);
 - (k) are photographs, microphotographs, videotape or other recorded images prepared under authority of section eleven hundred eleven-b of the vehicle and traffic law (this exemption will be repealed effective December 1, 2014); or
 - (l) are photographs, microphotographs, videotape or other recorded images produced by a bus lane photo device prepared under authority of section eleven hundred eleven-c of the vehicle and traffic law (this exemption will be repealed effective September 20, 2015).
- For subparagraphs (j) through (l) above, such information must be included on the date such subparagraphs will be repealed.

Local Law 11 specifies the following additional exemptions:

- any portion of such data set to which an agency may deny access pursuant to the public officers law or any other provision of a federal or state law, rule or regulation or local law;
- any data set that contains a significant amount of data to which an agency may deny access pursuant to the public officers law or any other provision of a federal or state law, rule or regulation or local law and where removing such data would impose undue financial or administrative burden;
- data that reflects the internal deliberative process of an agency or agencies, including but not limited to negotiating positions, future procurements, or pending or reasonably anticipated legal or administrative proceedings;
- data stored on an agency-owned personal computing device, or data stored on a portion of a network that has been exclusively assigned to a single agency employee or a single agency owned or controlled computing device;
- materials subject to copyright, patent, trademark, confidentiality agreements or trade secret protection;
- proprietary applications, computer code, software, operating systems or similar materials; or
- employment records, internal employee-related directories or lists, and facilities data, information technology, internal service-desk and other data related to internal agency administration.

Nothing in the legislation, policies, or standards shall be deemed to prohibit an Agency from voluntarily disclosing information not otherwise defined as a public data set, nor shall it be deemed to prohibit an agency from making such voluntarily disclosed information accessible through the NYC Open Data Portal.

5.0 City Guidelines

5.1 Data Set Selection

Agencies should use the following guidelines to select and prioritize their data sets for publication.

5.1.1 Prioritization Criteria

For purposes of prioritizing public data sets, Agencies should consider whether information embodied in the public data set:

- Increases Agency accountability and responsiveness;
- Improves public knowledge of the Agency and its operations;
- Responds to a need or demand identified by the public;
- Furthers the mission of the Agency;

- Reduces the impact of automated tools which scan the City’s website for data;
- Fosters agency/interagency efficiency; or
- Creates economic opportunity.

5.1.2 Public Input and Participation

Public input is essential to selecting and disseminating information. The NYC Open Data Portal includes an online forum to solicit feedback from the public and to encourage public discussion on open data policies and public data set availability. Agencies should use this forum to solicit recommendations regarding the presentation of data, data types, and metadata from individuals, groups, and organizations.

5.1.3 Inter-agency Coordination

Interoperability makes data more valuable by making it easier to derive new uses from combinations of data. If data sets from two or more Agencies refer to the same kinds of things, the Agencies that create the data sets should strive to make them interoperable. This may mean developing a shared data specification or adopting an existing specification, through coordination between Agencies. The use of open data formats often, but not always, entails interoperability. However, Agencies must weigh the advantages of distributing non-interoperable data quickly against the net gain of investing in interoperability and delaying a release of the data.

5.2 Guidelines for the Division of Large Datasets

Local Law 106 of 2015 amended the Open Data Law of 2012 to mandate the creation of a set of guidelines for users who may experience technical difficulties manipulating and downloading the larger datasets on the Portal.

Users may specify multiple conditions restricting the number of records that appear on the dataset using the “Filter Panel.” In this panel, users may “Add a New Filter Condition” and select attribute values that are exact matches with the “is” condition, or a fall within a range of values with the “contains” condition. A narrower selection of results will require less computing power to view and manipulate on the Portal and will create smaller data files that are quicker to download onto a user’s local device.

For more information on filtering tabular datasets, including video tutorials, see the “Filters” topic in the Socrata Knowledge Base: <https://support.socrata.com/hc/en-us/sections/200469748-Filters>

6.0 Public Policies

Local Law 11, specifically § 23-502 d, requires that data sets must be available without registration requirement, license requirement, or usage restrictions, except as provided in Sections 6.1 through 6.4 below.

6.1 Terms of Use

The following Terms of Use apply to visitors to the NYC Open Data Portal and application developers who obtain City data through this single web portal:

By accessing data sets and feeds available through the NYC Open Data Portal (or the "Site"), the user agrees to all of the [Terms of Use of NYC.gov](#) as well as the [NYC.gov Privacy Policy](#). The user also agrees to any additional terms of use defined by entities providing data or feeds through the Site. Entities providing data include, without limitation, agencies, bureaus, offices, departments and other discrete entities of the City of New York ("City"). Public data sets made available on the NYC Open Data Portal are provided for informational purposes. The City does not warranty the completeness, accuracy, content, or fitness for any particular purpose or use of any public data set made available on the NYC Open Data Portal, nor are any such warranties to be implied or inferred with respect to the public data sets furnished therein.

The City is not liable for any deficiencies in the completeness, accuracy, content, or fitness for any particular purpose or use of any public data set, or application utilizing such data set, provided by any third party.

Submitting City Agencies are the authoritative source of data available on NYC Open Data. These entities are responsible for data quality and retain version control of data sets and feeds accessed on the Site. Data may be updated, corrected, overwritten and/or refreshed at any time. The anticipated update frequency is indicated for each data set on the Site. Older versions of data sets will not be retained.

6.2 Ensuring Continued Public Access to Data Sets

DoITT may implement rate-limiting on a per-visitor basis to ensure that the NYC Open Data Portal's system resources are shared fairly between all visitors.

6.3 Protecting the NYC Open Data Portal

DoITT may implement rate-limiting and other security mechanisms to ensure that the NYC Open Data Portal is not abused, damaged, or impaired.

6.4 Improving Service Delivery

DoITT may implement registration capabilities in order to gather usage metrics, particularly for third-party applications which routinely access the NYC Open Data Portal. In addition, the

developers of third-party applications will benefit from receiving updates from DoITT as enhancements to the site and the data it contains are applied over time.

The goal is to improve the way the NYC Open Data Portal functions and the data available through it, by using registration only to ensure access to public data sets, to protect the single web site housing public data sets from unlawful abuse or attempts to damage or impair use of the web site, or to analyze the types of data being used to improve service delivery.

7.0 Appendix A: NYC Administrative Code (Title 23, Chapter 5)

Chapter 5: Accessibility to Public Data Sets

§ 23-501 Definitions

As used in this chapter:

- a. "Agency" 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.
- b. "Data" means 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. Such term 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 data. Nothing in this chapter shall be deemed to prohibit an agency from voluntarily disclosing information not otherwise defined as "data" in this subdivision, nor shall it be deemed to prohibit an agency from making such voluntarily disclosed information accessible through the single web portal established pursuant to section [23-502](#).
- c. "Department" means the department of information technology and telecommunications or any successor agency.
- d. "Determination" means any final decision made by an agency with respect to a person, including, but not limited to:
 - (1) eligibility for services or benefits;
 - (2) issuing a permit;
 - (3) registration, certification and licensing; and
 - (4) liability for civil and criminal penalties.
- e. "Measurement" means to quantify by means of comparison to a reference standard any characteristic of an observable event, occurrence or object.
- f. "Open standard" means a technical standard developed and maintained by a voluntary consensus standards body that is available to the public without royalty or fee.

- g. "Public data set" means 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. Such term shall not include:
- (1) any portion of such data set to which an agency may deny access pursuant to the public officers law or any other provision of a federal or state law, rule or regulation or local law;
 - (2) any data set that contains a significant amount of data to which an agency may deny access pursuant to the public officers law or any other provision of a federal or state law, rule or regulation or local law and where removing such data would impose undue financial or administrative burden;
 - (3) data that reflects the internal deliberative process of an agency or agencies, including but not limited to negotiating positions, future procurements, or pending or reasonably anticipated legal or administrative proceedings;
 - (4) data stored on an agency-owned personal computing device, or data stored on a portion of a network that has been exclusively assigned to a single agency employee or a single agency owned or controlled computing device;
 - (5) materials subject to copyright, patent, trademark, confidentiality agreements or trade secret protection;
 - (6) proprietary applications, computer code, software, operating systems or similar materials; or
 - (7) employment records, internal employee-related directories or lists, and facilities data, information technology, internal service-desk and other data related to internal agency administration.
- h. "Technical standard" means (1) the common and repeated use of rules, conditions, guidelines or characteristics for products or related processes and production methods, and related management systems practices; and (2) (i) the definition of terms; (ii) classification of components; (iii) delineation of procedures; (iv) specifications of dimensions, materials, performance, designs or operations; (v) measurement of quality and quantity in describing materials, processes, products, systems, services or practices; (vi) test methods and sampling procedures; or (vii) descriptions of fit and measurements of size or strength.
- i. "Transaction" means any interaction between an agency and any person related to the mission of an agency.
- j. "Voluntary consensus standards body" means a domestic or international organization that develops and maintains a technical standard that utilizes a transparent deliberative process, permits the participation of any party, and achieves general consensus, although not necessarily unanimity, of the participating parties, including a process for attempting to resolve any differences in viewpoint.

§ 23-502 Public data set availability

- a. Within one year of the effective date of this chapter and thereafter, the public data sets that agencies make available on the Internet shall be accessible through a single web portal that is linked to nyc.gov or any successor website maintained by, or on behalf of, the city of New York. If an agency cannot make all such public data sets available on the single web portal pursuant to this subdivision, the agency shall report to the department and to the council which public data set or sets that it is unable to make available, the reasons why it cannot do so and the date by which the agency expects that such public data set or sets will be available on the single web portal.
- b. Such public data sets shall be made available in accordance with technical standards published by the department pursuant to section [23-505](#) of this chapter and shall be in a format that permits automated processing and shall make use of appropriate technology to notify the public of all updates.
- c. Such public data sets shall be updated as often as is necessary to preserve the integrity and usefulness of the data sets to the extent that the agency regularly maintains or updates the public data set.
- d. Such public data sets shall be made available without any registration requirement, license requirement or restrictions on their use provided that the department may require a third party providing to the public any public data set, or application utilizing such data set, to explicitly identify the source and version of the public data set, and a description of any modifications made to such public data set. Registration requirements, license requirements or restrictions as used in this section shall not include measures required to ensure access to public data sets, to protect the single web site housing public data sets from unlawful abuse or attempts to damage or impair use of the web site, or to analyze the types of data being used to improve service delivery.
- e. Such public data sets shall be accessible to external search capabilities.
- f. Agencies shall review responses to freedom of information law requests that include the release of data to determine if such responses consist of or include public data sets that have not yet been included on the single web portal or the inclusion of which on the single web portal is not provided for in the compliance plan prepared pursuant to section [23-506](#). Each agency shall disclose in the update to such compliance plan the total number, since the last update, of such agency's freedom of information law responses that included the release of data, the total number of such responses determined to consist of or include a public data set that had not yet been included on

the single web portal and the total number of such responses that resulted in voluntarily disclosed information being made accessible through the single web portal.

§ 23-503 Web portal administration

- a. The department may take reasonable measures to maintain bandwidth availability of the web portal.
- b. The department shall conspicuously publish the open data legal policy, as provided in section [23-504](#), on the web portal.
- c. The department shall implement an on-line forum to solicit feedback from the public and to encourage public discussion on open data policies and public data set availability on the web portal.
- d. Requests received via the on-line forum for inclusion of particular public data sets shall be considered by agencies in making determinations as to priority for public data set inclusion on the single web portal pursuant to paragraph 5 of subdivision b of section [23-506](#). The department shall provide an initial response to each such request within two weeks of receipt. The agency to which the department refers the request shall post its final determination as to the request on the single web portal within two months of receipt.

§ 23-504 Open data legal policy

- a. Public data sets made available on the web portal are provided for informational purposes. The city does not warranty the completeness, accuracy, content or fitness for any particular purpose or use of any public data set made available on the web portal, nor are any such warranties to be implied or inferred with respect to the public data sets furnished therein.
- b. The city is not liable for any deficiencies in the completeness, accuracy, content or fitness for any particular purpose or use of any public data set, or application utilizing such data set, provided by any third party.
- c. This chapter shall not be construed to create a private right of action to enforce its provisions. Failure to comply with this chapter shall not result in liability to an agency.

§ 23-505 Internet data set policy and technical standards

- a. Within one hundred eighty days of the effective date of this chapter, the department shall prepare and publish a technical standards manual for the publishing of public data sets in raw or unprocessed form through a single web portal by city agencies for the purpose of making public data available to the greatest number of users and for the greatest number of applications and shall, whenever practicable, use open standards for web publishing and e-government. Such manual shall identify the reasons why each technical standard was selected and for which types of data it is applicable and may recommend or require that data be published in more than one technical standard. The manual shall include a plan to adopt or utilize a web application programming interface that permits application programs to request and receive public data sets directly from the web portal. Such manual shall be updated by the department as necessary.
- b. The department shall consult with voluntary consensus standards bodies and shall, when such participation is feasible, in the public interest and is compatible with agency and departmental missions, authorities and priorities, participate with such bodies in the development of technical and open standards.
- c. Each data set on the single web portal shall include a plain language data dictionary; provided that for any data set available on the single web portal before the effective date of the local law that added this subdivision, a data dictionary shall be added to such data set no later than December 31, 2017; and provided further that for any data set made available on the single web portal on or after the effective date of the local law that added this subdivision and before December 31, 2017, a data dictionary shall be added to such data set no later than thirty days after such data set is made available on the single web portal. Such data dictionary shall provide a description for each column heading used within the data set and shall include a description of any acronym, technical term, unit of measure, range of possible values, relationship between or among columns within the data set, frequency of updates to the data set, and other information or description that can provide context to the data, such as the method of collection, a history of modifications to the data set format, data or methods of collection, or any other contextual information that the agency providing the data deems relevant or the technical standards manual requires. Such data dictionary shall be in a format and layout to be determined by the technical standards manual. The data dictionary accompanying each data set, or a link to such data dictionary, shall be included directly on the single web portal. The most recent upload date and generation date for each data set shall be included directly on the single web portal.

§ 23-506 Agency compliance plan

- a. Within eighteen months of the effective date of this chapter, the department shall submit a compliance plan to the mayor and the council and shall make such plan available to the public on the web portal. Each agency shall cooperate with the

department in its preparation of such plan. The plan shall include a summary description of public data sets under the control of each agency on or after the effective date of this chapter, and shall prioritize such public data sets for inclusion on the single web portal on or before December 31, 2018 in accordance with the standards promulgated by the department pursuant to section [23-505](#) and shall create a timeline for their inclusion on the single web portal. If a public data set or sets cannot be made available on the single web portal on or before December 31, 2018, the plan shall state the reasons why such set or sets cannot be made available, and, to the extent practicable, the date by which the agency that owns the data believes that it will be available on the single web portal.

- b. For purposes of prioritizing public data sets, agencies shall consider whether information embodied in the public data set: (1) can be used to increase agency accountability and responsiveness; (2) improves public knowledge of the agency and its operations; (3) furthers the mission of the agency; (4) creates economic opportunity; or (5) responds to a need or demand identified by public consultation.
- c. No later than July fifteen, 2014 and every July fifteen thereafter, the department shall submit and post on the web portal an update of the compliance plan to the mayor and the council until all public data sets have been made available through a single web portal in compliance with this chapter. Such update shall include the specific measures undertaken to make public data sets available on the single web portal since the immediately preceding update, specific measures that will be undertaken prior to the next update, an update to the list of public data sets, if necessary, any changes to the prioritization of public data sets and an update to the timeline for the inclusion of data sets on the single web portal, if necessary. If a public data set cannot be made available on the single web portal on or before December 31, 2018, the update shall state the reasons why it cannot and, to the extent practicable, the date by which the agency believes that such public data set will be available on the single web portal.

8.0 Appendix B: End Notes

- ¹ “New York City Charter: As Amended Through December 18, 2009,” State of New York Legislative Bill Drafting Commission, February 2010, retrieved 02/28/2011 from <http://www.nyc.gov/html/charter/downloads/pdf/citycharter2009.pdf>.
- ² New York City Department of Information Technology & Telecommunications, “30-Day Report: Enabling the Connected City,” February 2010, retrieved 02/28/2011 from http://www.nyc.gov/html/doitt/downloads/pdf/30_day_report.pdf >.
- ³ Michael R. Bloomberg, Mayor, City of New York, “Executive Order No.140: Authorizing the Department of Information Technology and Telecommunications to Consolidate the Infrastructure of and Establish Coordinated Citywide Policies for Information Technology and Telecommunications for the City of New York,” issued October 20, 2010, retrieved 02/04/2011 from http://www.nyc.gov/html/doitt/downloads/pdf/eo_140.pdf.
- ⁴ Freedom of Information Law (FOIL) is codified at [New York Department of State, Public Officers Law, Article 6, Sections 84-90](#).