



# Open Data Policy and Technical Standards Manual

**September 2012**

**Michael R. Bloomberg**

Mayor

**Rahul N. Merchant**

Commissioner


## TABLE OF CONTENTS

1.0	Executive Summary.....	1
2.0	Introduction.....	1
2.1	Background, Authority, and Governance.....	1
2.2	Definitions.....	2
3.0	City Policies.....	5
3.1	Governance and Oversight.....	5
3.2	Technology Projects.....	7
3.3	Tools.....	8
3.4	Data Sets.....	8
4.0	City Standards.....	9
4.1	Integration Architecture.....	9
4.2	Data Sets.....	11
4.3	Data Set Publishing.....	14
4.4	Maintenance.....	15
4.5	Ownership, Responsibility, and Retention.....	16
4.6	Exemption from Public Access.....	17
5.0	City Guidelines.....	19
5.1	Data Set Selection.....	19
6.0	Public Policies.....	20
6.1	Explicit Identification of Public Use by Third Parties.....	20
6.2	Ensuring Continued Public Access to Data Sets.....	21
6.3	Protecting the NYC OpenData Portal.....	21
6.4	Improving Service Delivery.....	21
7.0	Public Standards.....	21
7.1	Web Application.....	21
7.2	Application Programming Interface (API).....	21
7.3	Download Formats.....	22
8.0	Appendix A: Local Law 11 of 2012.....	23
9.0	Appendix B: End Notes.....	29

## 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.

The legislation additionally sets forth the following timeline:

- On or before March 7, 2013, all qualifying data currently on *NYC.gov* must be made available via a single web portal (the NYC OpenData portal).
  - On or before September 9, 2013, DoITT, on behalf of the City, must publish a compliance plan detailing qualifying data and the timeline to make the data publicly available. Following this, compliance plans must be updated annually and published no later than July 15th of each successive year.
  - On or before December 31, 2018, the City must complete its efforts to publish all qualifying data via the NYC OpenData portal.
- 

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.”<sup>1</sup>

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,”<sup>2</sup> 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.”<sup>3</sup>

In October 2011, DoITT launched the [NYC OpenData portal](#), advancing beyond simple downloads to fully interactive viewing, searching, and visualizing of City data. Additionally, the NYC OpenData 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
<b>Agency</b>	<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>
<b>City</b>	The City of New York is referred to interchangeably as New York City, NYC, or the City.
<b>Citywide Standard</b>	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.
<b>Data</b>	<p>(Per Local Law 11 definition)</p> <p>Final versions of statistical or factual information</p> <ol style="list-style-type: none"> <li>1. In alphanumeric form reflected in a list, table, graph, chart or other non-narrative form, that can be digitally transmitted or processed; and</li> <li>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.</li> </ol> <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>
<b>Data Set</b>	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.

Term	Definition
<b>DoITT</b>	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.”
<b>Metadata</b>	Contextual information that makes the public data sets easier to understand and use.
<b>ODC</b>	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.
<b>Open Standard</b>	(Per Local Law 11)  A technical standard developed and maintained by a voluntary consensus standards body that is available to the public without royalty or fee.
<b>Public Data Set</b>	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.  The entry above is only a summary. Please consult Local Law 11 for the full authoritative definition.
<b>Requirement</b>	<ol style="list-style-type: none"> <li>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.</li> <li>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.</li> <li>3. A singular documented need of what a particular product or service should be or perform.</li> </ol>

Term	Definition
<b>Technical Standard</b>	<p>(Per Local Law 11 definition)</p> <p>A standard that includes:</p> <ol style="list-style-type: none"> <li>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</li> <li>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.</li> </ol>
<b>Voluntary Consensus Standards Body</b>	<p>(Per Local Law 11 definition)</p> <p>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.</p>

### 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 Compliance

Local Law 11 establishes the following compliance requirements:



### 3.1.3.1 Initial Compliance (March 7, 2013)

Local Law 11 requires that on or before March 7, 2013, all qualifying data sets currently on NYC.gov must be made available via a single web portal.

If an Agency cannot make available all of its public data sets on the NYC OpenData portal by March 7, 2013, that Agency shall clearly and concisely state the reasons why any applicable data set cannot be made available, and the date by which the Agency believes that it will be available on the NYC OpenData portal.

### 3.1.3.2 Agency Compliance Plan (September 9, 2013)

DoITT must submit an initial compliance plan to the Mayor and the City Council on September 9, 2013. Following the initial plan, 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 will create a public Open Data dashboard to promote public and internal accountability. The dashboard will track agency compliance plans and interim progress – specifically which data sets have been published through the NYC OpenData portal or by direct public access. The dashboard will also display 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 will provide and manage the NYC OpenData portal for the City's public data set catalog. All public data sets must be identified and referenced on this site. The single web portal will include 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 OpenData 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 must 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 OpenData 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 OpenData 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 OpenData portal itself, the NYC OpenData 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 OpenData 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 Integration Architecture

Delivery of data from back-end systems to the NYC OpenData portal will take place through an architecture which permits de-coupling and enables a layer of abstraction. This architecture leverages infrastructural investments and technical capabilities that either already exist within Agencies or would require minimal effort to implement.

A de-coupled architecture provides the following benefits:

- It supports effective data governance including the ability to avoid the public release of information not otherwise defined as public data, including personally identifiable information.
- It supports the ability to monitor update cycles and identify issues for rapid resolution.
- It permits the use of centralized services to help in data transformation, such as the conversion of geographic coordinates from the internally-used NY State Plane (FIPS

Zone 3104) projection to the widely-Internet-used Web Mercator (WGS 84/EPSG:3857) projection.

- It permits the City to migrate to a different NYC OpenData portal solution without requiring significant re-engineering at every data source.

The diagram below represents a logical architecture for the integration capabilities.

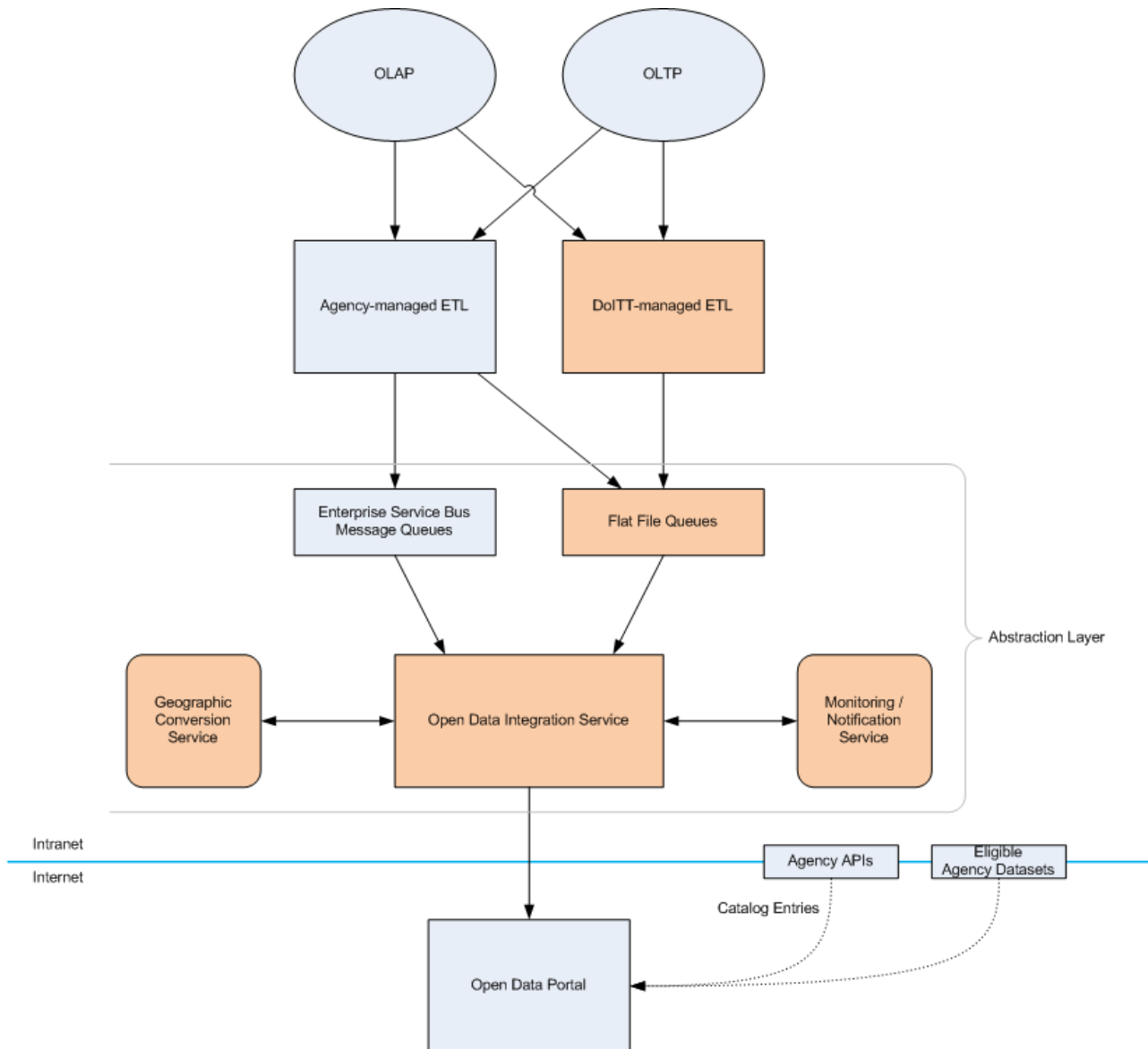


Figure 1 – Logical Integration Architecture

**ETL:** Extract-Transform-Load – a process by which data is extracted from a source system, manipulated as required by business rules, and then loaded into another system.

**OLAP:** Online Analytical Processing data storage – typically used for reporting and data mining capabilities. Data sets are typically de-normalized, and information therein may be routinely extracted from OLTP systems.

**OLTP:** Online Transactional Processing data storage – typically used for data entry and retrieval. Data sets are typically highly normalized, and information therein may be routinely extracted, transformed, and loaded into OLAP systems.

## 4.2 Data Sets

### 4.2.1 Data Content

#### 4.2.1.1 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.

Please refer to the NYC OpenData portal's import specifications for details on formatting and parsing of the above data types: <http://dev.socrata.com/publisher/import-data-types>.

#### 4.2.1.2 Format and Code Elements

Data values must not contain elements or markup used for presentation, nor should they contain interpreted or raw application source code. For example, HTML formatting tags such as <script>, <table>, <tr>, <td>, or <br> are not permitted.

#### 4.2.1.3 Geospatial Data

Geospatial data must be published in the Web Mercator coordinate system (WGS 84/EPWG:3857) to make the data easy to use with popular online mapping services. Although this is the most useful coordinate system for web-based mapping, Web Mercator as a geographic coordinate system is not a projection, and therefore the measurement of distance and area on such data will not be as accurate as a projected coordinate system.

Agencies may also make their data available in the New York State Plane, Long Island Zone, coordinate system (FIPS Zone 3104/EPWG:2263). If the data is hosted directly through the OpenData platform, it will be automatically converted to Web Mercator.

#### 4.2.1.4 Geocoding

Data sets providing information on location in tabular format can be automatically geocoded by the OpenData platform.

Address – single column that includes the building number, street name, city, state and zip code.

Intersection – single column with the cross streets concatenated by an ampersand (&) or two separate columns (e.g., cross street 1 and cross street 2).

#### 4.2.2 Metadata

##### 4.2.2.1 Data Set Metadata

For each data set published, the providing Agency must, at a minimum, provide values for all of the metadata elements as defined in the latest version of the [DublinCore Metadata Element Set](#). In addition, the Agency must provide the metadata element “*frequency*” which must correspond to a value contained in the [DublinCore Collection Description Frequency Vocabulary](#).

The following table represents a list of required metadata elements for data sets as of the publication of this technical standard:

Label	Description	Permitted Values (if applicable)
Contributor	Indicates the agency that supplied the data	
Coverage	Indicates the range of data from either a temporal or spatial perspective.	
Creator	Indicates the agency that supplied the data	
Date	Auto-generated by Socrata when data set (or metadata) is modified.	
Description	A brief description of the data set	
Format	Dependent upon export methodology. Refer to Section 7.3 for specific details.	
Frequency	Indicates the rate at which the information in the data set is updated.	Not updated (historical only) Annual Quarterly Bi-monthly Monthly Bi-weekly Weekly

Label	Description	Permitted Values (if applicable)
		Daily
		Hourly
		Continuous
Identifier	Socrata uses a 9-character identifier (usually xxxx-xxxx); may have the option for better permlinking under the "resource name" field.	
Language	Language of the data set. Assumed to be en-US for all data sets. Exceptions must be noted.	en-US
Publisher	Entity that is responsible for publishing the data; this will always be the City of New York.	City of New York
Relation	Not used	
Rights	NYC data sets should be attributed to the City.	
Source	Identifies the name of the source system within the City	
Subject	Comma-separated list of nouns describing the content of the data set	
Title	The brief descriptive name of the data set	
Type	The category of the data set identified by the list of possible values. If a data set can fall into multiple categories, select the one which is most significant. This list will be subject to change on an ongoing basis.	Business and Economic
		Community Service
		Construction and Housing
		Cultural Affairs
		Education
		Environmental Sustainability
		Events
		Facilities and Structures
		Finances
		Government
		Health
		Library
		Media

Label	Description	Permitted Values (if applicable)
		Organizations
		Other
		Property
		Public Safety
		Social Services
		Statistics
		Transportation
		Women's Issues

#### 4.2.2.2 Column Metadata

Although metadata for columns within a data set is not required, it should be provided when the column identifiers do not provide a user with enough information to use it effectively. For example, the metadata for a column containing restaurant inspection letter grades should indicate the possible values and their meanings.

#### 4.2.2.3 Standards for Data Citation

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

### 4.3 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:

#### 4.3.1 Database Management System (DBMS) Access

For Agencies that require DoITT assistance to extract data from back-office systems, the Agency must 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.



### 4.3.2 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.

### 4.3.3 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.

### 4.3.4 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 OpenData 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 OpenData 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 OpenData 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 OpenData 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.

## 4.4 Maintenance

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

### 4.4.1 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.

#### 4.4.2 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.

#### 4.4.3 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.

### 4.5 Ownership, Responsibility, and Retention

#### 4.5.1 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 OpenData 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.

#### 4.5.2 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 OpenData 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.

#### 4.5.3 Retention

As the authoritative source of the information, submitting Agencies retain version control of public data sets and must comply with record retention schedules and requirements outlined by the New York City Department of Records and Information Services.<sup>4</sup>

#### 4.6 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.<sup>5</sup> (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 OpenData 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 OpenData 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 Public Review

Public review creates a means for Agencies to establish and maintain communications with members of the public and governments inside and outside the City's jurisdiction. Such review ensures that Agencies create valuable information products, along with ways of disseminating information.

#### 5.1.4 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.

## 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 Explicit Identification of Public Use by Third Parties

Notwithstanding Section 6.0 above, DoITT may require third party entities such as application developers to explicitly identify the source, version, and modifications made to a public data set, should they publicly re-publish it elsewhere or incorporate the public data set into an application. This is addressed through updated Terms of Use for the NYC OpenData portal.

#### 6.1.1 Terms of Use

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

By accessing data sets and feeds available through the NYC OpenData 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 OpenData 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 OpenData 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 OpenData. 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 OpenData portal's system resources are shared fairly between all visitors.

## 6.3 Protecting the NYC OpenData Portal

DoITT may implement rate-limiting and other security mechanisms to ensure that the NYC OpenData 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 OpenData 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 OpenData 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 Public Standards

### 7.1 Web Application

The NYC OpenData portal supports most modern web browsers. The web application permits the listing, viewing, exporting, embedding, filtering, visualizing, personalizing, commenting on, and rating of public data sets.

### 7.2 Application Programming Interface (API)

The NYC OpenData portal supports a generic API that permits access to all published data sets in a similar manner. The API is based upon the Socrata Open Data API ([SODA](#)).

### 7.3 Download Formats

In 2009, the U.S. Government defined an open file format as “one that is platform independent, machine readable, and made available to the public without restrictions that would impede the re-use of that information.”<sup>6</sup>

The NYC OpenData portal supports the public downloading of data sets in the following formats:

Type	Type	Description
CSV	Text	Comma-separated values
JSON	Text	JavaScript Object Notation
PDF	Binary	Portable Document Format
RDF	Text	Resource Description Framework
RSS	Text	RDF Site Summary/Really Simple Syndication <i>(Note: each row is represented as a separate item, but each item’s description field contains an HTML table with the column names and row values)</i>
XLS	Binary	Microsoft Excel
XLSX	Binary	Microsoft Excel Open XML
XML	Text	Extensible Markup Language

In addition, some data sets may also have the following formats available:

Type	Type	Description
ZIP	Binary	Typically contains a shapefile set (SHP, SHX, DBF) (Geographic data sets only), or a collection of multiple files which are all part of the same data set. <i>(Note: this is usually provided within a compressed archive)</i>



## 8.0 Appendix A: Local Law 11 of 2012

New York City Council voted on and passed this legislation on February 29, 2012. Mayor Bloomberg signed it into law on March 7, 2012. The text below has been formatted to enhance readability.

**Local Law 11 of 2012** (previously [Introduction 0029A-2010](#))

*By Council Members Brewer, Fidler, Gentile, Gonzalez, James, Lander, Palma, Nelson, Lappin, Dromm, Garodnick, Reyna, Mealy, Vann, Jackson, Levin, Cabrera, Gennaro, Williams, Van Bramer, Rodriguez, Halloran and the Public Advocate (Mr. de Blasio)*

*A Local Law to amend the administrative code of the city of New York, in relation to publishing open data.*

Be it enacted by the Council as follows:

### ***Section 1. Declaration of legislative findings and intent.***

The council hereby finds and declares that it is in the best interest of New York city that its agencies and departments make their data available online using open standards. Making city data available online using open standards will make the operation of city government more transparent, effective and accountable to the public. It will streamline intra-governmental and inter-governmental communication and interoperability, permit the public to assist in identifying efficient solutions for government, promote innovative strategies for social progress, and create economic opportunities.

### ***Section 2. [Amendment to Administrative Code]***

Title 23 of the administrative code of the city of New York is amended by adding a new chapter 5 to read as follows:

#### **CHAPTER 5**

#### **ACCESSIBILITY TO PUBLIC DATA SETS**

§ 23-501 Definitions.

§ 23-502 Public data sets availability.

§ 23-503 Web portal administration.

§ 23-504 Open data legal policy.

§ 23-505 Internet data set policy and technical standards.

§ 23-506 Agency compliance plan.

#### **§ 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.

**§ 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 of this chapter.

**§ 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.

**§ 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.

***Section 3. [Effective date]***

This local law shall take effect immediately.

## 9.0 Appendix B: End Notes

- 
- <sup>1</sup> “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>.
- <sup>2</sup> 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](http://www.nyc.gov/html/doitt/downloads/pdf/30_day_report.pdf).
- <sup>3</sup> 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](http://www.nyc.gov/html/doitt/downloads/pdf/eo_140.pdf).
- <sup>4</sup> New York City Department of Records, “Managing New York City’s Records: Guidelines, Policies, and Procedures,” August, 2007, retrieved 02/25/2011 from [http://nyc.gov/html/records/pdf/records\\_retention\\_manual2007.pdf](http://nyc.gov/html/records/pdf/records_retention_manual2007.pdf).  
Section 6.0.1: “Pursuant to section 1133 of the New York City Charter, each City agency is responsible for compiling its own, agency-specific records retention schedule, which categorizes the records maintained by that agency and sets forth the time periods for which each category of records must be retained. The agency-specific records retention schedule should encompass all records maintained by the agency in any format.”
- <sup>5</sup> Freedom of Information Law (FOIL) is codified at [New York Department of State, Public Officers Law, Article 6, Sections 84-90](#).
- <sup>6</sup> Peter R. Orszag, Director, Open Government Initiative, “Open Government Directive: Memorandum for the Heads of Executive Departments and Agencies,” December 8, 2009, retrieved 02/04/2011 from <http://www.whitehouse.gov/open/documents/open-government-directive>.