



OPEN DATA POLICY

1. Background

On March 11, 2013, Governor Andrew Cuomo launched the Open NY initiative which included the launch of New York State's Open Data Website, data.ny.gov.

Concurrent with the launch of Open NY, Governor Cuomo issued Executive Order No. 95, "Using Technology to Promote Transparency, Improve Government Performance and Enhance Citizen Engagement." This Executive Order directed covered state entities to identify and catalogue their data, and make publishable State data available on the new transparency website.

2. Purpose of Policy

The New York State Olympic Regional Development Authority ("ORDA") is committed to the principles of open, accessible, efficient, and transparent government, and the use of technology to help put those principles into practice.

This policy creates a process for making ORDA data available to the public on the Open Data Website. ORDA will make its "data", as that term is defined herein, publicly available online to promote civic engagement, improve service delivery, allow for more effective communication with the public, and increase opportunities for economic development. This policy is also intended to make the operation of ORDA more transparent, effective, and accountable to the public.

3. Definitions

Covered State Entity

- (i) any State agency or department, or any office, division, bureau, or board of such State agency or department, except where the head of such agency or department is not appointed by the Governor,
- (ii) any State board, committee, or commission, at least one of whose members is appointed by the Governor, and
- (iii) all public-benefit corporations, public authorities and commissions, for which the Governor appoints the Chair, the Chief Executive, or the majority of Board Members, except for the Port Authority of New York and New Jersey.

Chief Data Officer (CDO) The New York State Chief Data Officer in the Office of Information Technology Services or a designee thereof.

Data Final versions of statistical or factual information that:

- (i) are in alphanumeric form reflected in a list, table, graph, chart or other nonnarrative form, that can be digitally transmitted or processed;
- (ii) are regularly created or maintained by or on behalf of a covered State entity and are controlled by such entity; and
- (iii) record a measurement, transaction or determination related to the mission of the covered State entity.

The term “data” shall not include image files, such as designs, drawings, photos or scanned copies of original documents; provided, however, that the term “data” shall include statistical or factual information about image files and geographic information system data.

Dataset A named collection of related records maintained on a storage device, with the collection containing data organized or formatted in a specific or prescribed way, often in tabular form.

ITS The New York State Office of Information Technology Services.

Publishable State Data Data that is collected by a covered State entity where the entity is permitted, required, or able to make the data available to the public, consistent with any and all applicable laws, rules, regulations, ordinances, resolutions, policies or other restrictions, requirements or rights associated with the State data, including but not limited to contractual or other legal orders, restrictions or requirements. Data shall not be Publishable State data if making such data available on the Open Data Website would violate statute or regulation (e.g., disclosure that would constitute an unwarranted invasion of personal privacy), endanger the public health, safety or welfare, hinder the operation of government, including criminal and civil investigations, or impose an undue financial, operational or administrative burden on the covered State entity or the State.

Protected Data Any data set, or portion thereof, exempt from disclosure pursuant to the New York State Freedom of Information Law (“FOIL”), or any other law or regulation.

4. Data Coordinator

ORDA's President and CEO will designate a Data Coordinator. The Data Coordinator will:

- have authority equivalent to the head of a division or department within ORDA;
- have knowledge of data and resources in use by ORDA; and
- be responsible for ORDA's compliance with the Executive Order, this policy, and future directives which may be needed to support the open data program.

The Data Coordinator will serve as the liaison between the ITS Open Data Website team and ORDA. In that position, the Data Coordinator will convey to ORDA's Data Owners, as identified herein, in-house counsel, and President/CEO any specific needs of the ITS team who maintain the Open Data Website, to enable them to format or define such data in an optimal manner for publication (such as formatting the data or defining a structure that is optimal for publication). These efforts provide an additional internal control to ensure the dataset is properly evaluated before being provided to the Open Data platform.

5. Publication of Data

Within thirty (30) days of the effective date of this policy, ORDA will create a catalogue of its Publishable State Data and propose a schedule to ITS and the CDO for making its Publishable State Data publicly available. Such schedules will be made publicly available and will further include a schedule for updating the data catalogue as appropriate. ORDA will prioritize data publication in accordance with guidelines as set forth herein.

6. Prioritization, Publication, Alteration, and Removal Process

ORDA shall engage in an internal review process established by the Data Coordinator in accordance with this policy, and shall obtain appropriate internal approvals for the datasets from the President & CEO, as recommended by the Data Owners responsible for the relevant datasets, which it wishes to commit to the Open Data Website. The internal review process will include a means by which to assure increasing data content quality and accuracy, and compliance with all security, privacy, confidentiality laws, rules, and regulations, as well as any Intellectual Property Rights requirements, and status under FOIL (including whether data may lawfully be withheld under the FOIL exemptions).

Prioritization:

In creating a data catalogue, ORDA's Data Coordinator will identify those datasets that are high value, high quality, complete, and in accordance with the definition of "Publishable State Data".

“High value” data, as defined within Executive Order 95, is that which can be used to increase ORDA’s accountability and responsiveness, improve public knowledge of ORDA and its operations, further the mission of ORDA, create economic opportunity, or respond to a need or demand identified after public consultation.

When creating a schedule for publication of datasets, a number of factors must be assessed. Prioritizing initial and ongoing publication will entail balancing high value with data quality, data availability, and data readiness.

The amount of time to prepare high quality data can vary significantly, as datasets vary in complexity. Therefore, in prioritizing data for release, ORDA’s Data Coordinator must account for sufficient time to: identify data, assess the data (i.e., ensure consistency, timeliness, relevance, completeness, and accuracy of the data), ensure completeness of the metadata and data dictionary, review and obtain all necessary approvals to publish the data, and prepare data, metadata and requisite accompanying documentation for publication.

Publication:

To publish a dataset, at a minimum, the Data Coordinator will obtain explicit approval from the individuals listed below, which will be memorialized on an ITS Standardized Approval Form, a copy of which is attached as Exhibit B, to be completed and signed prior to dataset publication. The Data Coordinator may determine whether additional internal approvals and signatures are required and shall include such additional persons in its review and approval for publication process.

The Data Coordinator is responsible for obtaining the following approvals:

- Data Owner: ORDA Department Heads and Directors will be the designated Data Owners.

The Data Owners will have the greatest familiarity with and knowledge of the dataset and the data it contains, and the purpose for the collection of the data. The Data Owners will know the accuracy and currency of the data and be best able to describe and fill in the metadata elements describing the data. Approval by the Data Owners also validates that ORDA has secured permission for publication and knowledge from the department which is most responsible for the specific data. Each Data Owner may identify individuals within their respective departments and/or units to assist them in the collection of data.

- Legal Department: ORDA’s in-house legal counsel will confirm that ORDA has sufficiently reviewed the dataset to ensure its publication complies with privacy and security requirements, intellectual property rights, FOIL responsibilities, this policy, and any other applicable laws, rules and regulations.

- **President/CEO:** Approval by ORDA's President/CEO demonstrates knowledge within ORDA's leadership that it is providing a dataset to data.ny.gov under full authority. It also serves as the ultimate internal control within ORDA to ensure personnel completed proper evaluations of the datasets.

Removal:

To remove a dataset, at a minimum, the Data Coordinator must receive explicit approval from the Data Owners identified hereinabove, ORDA's in-house counsel, and ORDA's President/CEO, which will be memorialized on an ITS Standardized Removal Form, a copy of which is attached as Exhibit C, to be completed and signed prior to dataset removal. The Data Coordinator may determine additional internal approvals and signatures are required and should include such additional persons in their review and approval process.

Alteration:

To alter a dataset, at a minimum, the Data Coordinator will obtain approval from the Data Owners identified hereinabove, which will be memorialized on an ITS Standardized Alteration Form, a copy of which is attached as Exhibit D, to be completed and signed prior to publication of the altered dataset. The Data Coordinator may determine that additional internal approvals and signatures are required and should include such additional persons in their review and approval process.

7. Standardization

The Data Coordinator will ensure that the following requirements are met when publishing datasets to the Open Data Website:

- **Metadata** – The Open Data Website adheres to core components of the Dublin Core standard for metadata and uses the current recommended set of elements, which are required to accompany each dataset (Refer to Exhibit A: Metadata Elements for additional details).
- **Descriptive Information** – The Open Data Website requires covered state entities to submit metadata and supplemental documentation with each dataset (e.g., data dictionaries, overview documents, etc.) in order to fully describe and maximize public understanding and interpretation of the data.
- **Domain Categories** – The Open Data Website supports common domain models and allows the Data Coordinator to transform and anchor datasets in a particular domain (Refer to Exhibit A for examples of categories).
- **Catalogue Sharing** – The Open Data Website combines with several data catalogs and explores common, open formats such as Data Catalog Vocabulary (DCAT) or Project Open Data (see <https://resources.data.gov/schemas/dcat-us/v1.1/>)

- **Datasets** – ORDA must provide standardized open data file formats to facilitate automatic processing of the data, making it easily accessible and available in machine-readable format.
- **Open Specifications** – When possible, published datasets must be compatible with open specifications (e.g. KML/KMZ and GeoJSON).
- **Content Formats** – Datasets must be machine-readable and have a clear separation of metadata from the original source data.
- **Tabular Data** – The Open Data Website currently supports the format CSV & TSV: Comma/Tab Separated Values.
- **Geographic Data** – The Open Data Website supports two data formats for geospatial information. The appropriate format is dependent on the specific characteristic of the underlying geographic data.
 - i. Points: All Tabular File Formats or Shapefile
 - ii. Lines: Shapfile
 - iii. Polygons: Shapefile

Point data can be stored in either tabular or Shapefile format. Tabular formatting of points requires either columns for latitude and longitude, or complete address information (house number, street, village/town/city, state, and zip code) that can be geocoded. In contrast, lines and polygons define complex geometric structures that are not easily defined as column attributes. Therefore, Shapefile format is a preferred format for these complex geographic structures.

Each shapefile (at a minimum) should contain the following files:

- i. .shp: Defines the geometry (shapes)
- ii. .dbf: Defines the attribute table
- iii. .prj: Projection, ensures the feature locations are accurately rendered on the map
- iv. .shx: Shape indexing file, for efficient processing

Note: Shapefiles that use projections other than WGS-1984/Web Mercator will not be transformed by the platform and may result in inaccurate representations of location.

Other supported geospatial formats may include Keyhole Markup Language (KML/KMZ).

- **Geocoding** – The Open Data Website supports geocoding services that converts address information into mappable coordinates (Latitude/Longitude)

8. Updates to Published Data Sets

ORDA will create an update schedule to maintain data freshness on the Open Data Website. Individual data sets will be evaluated for staticity and schedules and value of data to be updated to the public. All data update schedules will be maintained in the metadata

uploaded to the Open Data Website with a minimum update schedule of annually and increased frequency, up to monthly, for more publicly valuable or volatile data sets.

- **Replace:** All existing records are removed, and new records are inserted.
- **Append:** New dataset records are inserted to existing records.

9. Protected Data

- **Security, Privacy, Regulatory, & Aggregate Data.**

The public release of some data might result in the violation of laws, rules, or regulations. Some data may not be appropriate to release because it can compromise internal ORDA processes, such as procurement. Other data may contain personally identifiable information. Finally, even if detailed data appears innocuous, it may be possible to easily combine it with other public information to reveal sensitive details. Even if there are no legal impediments to publishing the data, releasing the data may have unintended or undesirable effects. The Data Coordinator will confer with ORDA's legal counsel prior to any publication of data on the Open Data Website and exclude any datasets whose publication would cause harm by disclosing ORDA's internal processes, as such data does not constitute "Publishable State Data".

- **Thresholds**

The Health Insurance Portability and Accountability Act ("HIPAA") and its privacy regulations have very exacting requirements for determining whether data have been sufficiently de-identified so as not to compromise individual privacy. The Data Coordinator will confer with ORDA's legal counsel prior to the publication of data on the Open Data Website and exclude any datasets whose publication would cause harm as described in HIPAA's privacy regulations, as such data does not constitute "Publishable State Data".

- **FOIL Applicability**

Under the NYS Public Officers Law, Article 6 (the NYS Freedom of Information Law, or "FOIL"), the presumption is that government records shall be open to the public, unless excludable under a narrow set of specific exemptions including such concerns as invasion of personal privacy, impairment of contractual or collective bargaining negotiations, exposure of protected trade secrets, interference with law enforcement or judicial proceedings, endangering life or safety, jeopardizing the security of state information technology assets, systems and infrastructures, and others. The Data Coordinator will confer with ORDA's FOIL officers and/or legal counsel prior to the publication of data on the Open Data Website and exclude any datasets whose publication would cause the harms described in the FOIL law, as such data does not constitute "Publishable State Data."

10. Enforcement Standards and Compliance

ORDA's President and CEO, in accordance with the requirements set forth in Executive Order No. 95, shall have the authority to promulgate reasonable rules and regulations to implement the requirements of this policy.

EXHIBIT A: METADATA ELEMENTS

Metadata Element	Description	Dublin Core Value
Dataset Name / Title	The name of the dataset as it will appear on the platform.	Title
Dataset Description	Short description that explains the purpose of the Dataset and the data within.	Description
Category	The general category that the dataset is included in on the site (The categories include: Economic Development, Education, Energy & Environment, Government & Finance, Health, Human Services, Public Safety, Recreation, Transparency, and Transportation).	Type
Tags / Keywords	Keywords about the dataset used for searching purposes.	Subject
Data Provided By	The Agency that provided the data.	Contributor
URL to Dataset Program Web Page	The URL to the program area web pages.	n/a
Responsible Organization Within Agency	The organization that the dataset owner is a part of.	Creator
Time Period	The time frame of data available in the associated data file (e.g., Beginning 2005).	Coverage (temporal)
Create Date	The date the resource was made available in its present form – auto generated.	Date
Posting Frequency	How often the Dataset will be refreshed (valid values are: Current, Daily, Twice weekly, Weekly, Monthly, Quarterly, Semi-annually, Annually, Biennially, Decennially, As Needed, Static - Not Updated).	n/a
Contact e-mail information	The email address the viewers of the data can use to ask questions about the dataset. This must be provided as an email address; email forms are not acceptable. General mailboxes are recommended, as the agency will be responsible for maintaining a valid email address.	n/a
Coverage	The coverage area included in the dataset (e.g., Statewide).	Coverage (spatial)
Granularity	The lowest levels of granularity available within the data file (ex. County).	n/a
Define any limitations	Description of any limitations of the Dataset or exclusions.	Rights
URL(s) to additional resources (optional)	URLs to additional resources that may be useful to an end-user.	Relation

Metadata Element	Description	Dublin Core Value
Narrative Information Overview Document	One to two-page document that explains the dataset in greater detail, explains the data collection process, and any limitations in the data use.	n/a
Data Collection Tool/ Data Input	Explanation of the data collection methodology.	n/a
Data Dictionary and/ or Data file layout	Data dictionary should explain the fields within the dataset in terms of their definition, type, size, and any other pertinent information that describes the dataset.	n/a
Benefit of Utilizing Dataset (optional)	Additional supporting documentation can include a data collection/ input tool, a benefits document that describes what can be gained from analyzing the data.	n/a



OPEN NY – OPEN DATA APPROVAL FORM

The State of New York has created OPEN-NY, an open data portal at <http://www.data.ny.gov>. It brings datasets from local, state and federal government entities ("agencies") into one location searchable by the public.

While the site is administered by the New York State Office of Information Technology Services (ITS), the agencies submitting datasets to OPEN-NY are the authoritative sources for their content. These agencies are responsible for certifying that their content has been reviewed and approved for publication by appropriate agency personnel, including confirming compliance with all laws, rules, and regulations related to confidentiality, privacy, security, intellectual property rights, and the Freedom of Information Law (FOIL).

Agencies are also responsible for version control of their content and retaining record copies of the content in compliance with record retention requirements of the NYS Arts and Cultural Affairs Law.

1. Governmental Entity: Your governmental entity is:

(Name of specific State entity or specific Local Government)

Datasets prepared for publication on the OpenNY open data platform require one-time approval – and the completion of Attachment B - prior to publication to ensure they have been properly vetted. Your agency may require additional internal approvals.

For every dataset, agencies must receive explicit approval and sign-off from the individuals listed in Attachment B. Approval forms must be completed and signed prior to dataset publication. Agencies may determine additional internal approvals and signatures are required, and should include such additional persons in their review and sign off process

At a minimum, the four approval signatures in Attachment B are required for each dataset.

- For simultaneous publication of multiple datasets, a single form may be used where the Data Owner and all other signatories are the same. (And, the names of each dataset must be individually identified).

For the refreshing of a dataset by an agency, agencies have an ongoing responsibility to ensure the new data won't raise any concerns such as security or privacy issues; that the data continues to be in compliance with all laws, rules, and regulations related to confidentiality, privacy, security, intellectual property rights, and the Freedom of Information Law (FOIL).

Please forward a copy of this document, including Attachment B, with each dataset submission

Attachment B

We approve the uploading and publication of the following datasets onto OPEN-NY, having checked that they meet our standards for quality, accuracy, and compliance with laws protecting security, privacy, or intellectual property, and we are retaining original record copies to the extent required by Articles 57 or 57-A of the NYS Arts and Cultural Affairs Law.

I. Lists of Datasets for Publication:

1. _____
2. _____
3. _____
4. _____
5. _____

II. Approved by:

1. Government entity's Data Owner of the above noted datasets (e.g. dep't head, bureau director, etc.):

Printed Name: _____ Title: _____

Signature: _____ Date: _____

2. Government entity's legal counsel (e.g. in-house or outside attorney):

Printed Name: _____ Title: _____

Signature: _____ Date: _____

3. Government entity's Point of Contact: (e.g, Data Coordinator)

Printed Name: _____ Title: _____

Signature: _____ Date: _____

4. Government entity's Chief Executive (or his/her designee):

Printed Name: _____ Title: _____

Signature: _____ Date: _____



OPEN NY – Approval for Dataset Automated Pull or Transform

Governor Cuomo, through Executive Order No. 95 ("EO-95"), created OPEN-NY, the State's open data portal that brings datasets from federal, local, and state government entities into one location searchable by the public (<http://www.data.ny.gov>). Under EO-95, the NYS Office of Information Technology Services (ITS) administers the site. ITS publishes an Open Data Handbook describing how agencies can structure conformant datasets.

Dataset Automated Pull: As the authoritative sources for their content, agencies submit datasets to OPEN-NY only after certifying that such "Publishable" data has been reviewed and approved for publication by appropriate agency personnel. This ongoing obligation, applicable to either initial manual uploads or to any updates of that data by the agency, includes confirming compliance with laws related to confidentiality, privacy, security, intellectual property rights, and the Freedom of Information Law (FOIL).

Agencies are also responsible for version control of their content, and for retaining record copies of the content in compliance with record retention requirements of the NYS Arts and Cultural Affairs Law.

By mutual agreement between the agency and ITS, OPEN-NY may use technical means, such as application programming interfaces (APIs), to automate dataset updates on behalf of the agency.

Dataset Transforms: "Transforms" are enhancements of the usability of the data beyond that which is required to be conformant with Open Data Handbook requirements. Agencies may request transforms at either the time of upload, or thereafter.

As transforms are bounded by platform capabilities, not all requested transforms may be possible. Agencies are encouraged to contact ITS (its.dl.open.solution.team@its.ny.gov) to discuss the desired dataset transformation. As with dataset uploads, agencies must receive explicit internal reviews and approvals prior to any transforms being designed or developed.

Prior to dataset publication, for every dataset initial upload or transform, agencies must complete and provide to ITS signed copies of either the attached Attachment C, or Attachment D, or both, indicating explicit approval and sign-off from the individuals listed in those attachments. Agencies are welcome to require additional internal approvals and signatures, and include such additional persons in their review and sign off process.

For simultaneous publication or transform of multiple datasets, a single form may be used where the agency's Data Owner and all other agency signatories are the same. (And, the names of each dataset must be individually identified).

Please forward a copy of this document, including either Attachment C, Attachment D, or both, with each dataset submission

OPEN NY – Approval for Dataset Automated Pull - Attachment C

Our governmental entity: _____ [*Enter name of specific State or Local Government*], confirming they meet our agency's standards for quality, accuracy, and compliance with EO-95 and laws protecting security, privacy, or intellectual property, and that we are retaining original record copies to the extent we are required to do so by Articles 57 or 57-A of the NYS Arts and Cultural Affairs Law, approves the automated pull and publication of the following datasets onto OPEN-NY:

I. Dataset Name and Identifier (4x4):

II. Approved by:

1. Government entity's Data Owner of the dataset (e.g. dep't head, bureau director, etc.):

Printed Name: _____ Title: _____

Signature: _____ Date: _____

2. Government entity's legal counsel (e.g. in-house or outside attorney):

Printed Name: _____ Title: _____

Signature: _____ Date: _____

3. Government entity's Point of Contact: (e.g, Data Coordinator)

Printed Name: _____ Title: _____

Signature: _____ Date: _____

4. Government entity's Chief Executive (or his/her designee):

Printed Name: _____ Title: _____

Signature: _____ Date: _____

OPEN NY – Approval for Dataset Transform - Attachment D

Our agency: _____ [Enter name of specific State or Local Government entity], having previously approved the uploading and publication of the following dataset onto OPEN-NY; having checked that it meets our standards for quality, accuracy, and compliance with laws protecting security, privacy, or intellectual property; and retaining original record copies to the extent required by Articles 57 or 57-A of the NYS Arts and Cultural Affairs Law, approves the following transform/changes to the specified item(s):

I. OPEN-NY Dataset name and Identifier (4x4):

II. Change Type(s): (List each transform separately, and if needed attach lookup tables, sample code, or other related documents as supplemental materials) **Supplemental materials attached?** [] Yes [] No

[] **Automated Data Pull** (The agency must report to OPEN-NY prior to any data source changes).

Location of data source: _____

Supplied format: _____

Authentication credentials required? _____

If 'Yes', do credentials expire? _____

If 'Yes', renewal timeframe and contact information: _____

[] **Descriptions of Transform(s)**

1. Field Name: _____

Transform description: _____

Example: _____

Transform Error Handling: _____

2. Field Name: _____

Transform description: _____

Example: _____

Transform Error Handling: _____

3. Field Name: _____

Transform description: _____

Example: _____

Transform Error Handling: _____

4. Field Name: _____

Transform description: _____

Example: _____

Transform Error Handling: _____

5. Field Name: _____

Transform description: _____

Example: _____

Transform Error Handling: _____

6. Field Name: _____

Transform description: _____

Example: _____

Transform Error Handling: _____

7. Field Name: _____

Transform description: _____

Example: _____

Transform Error Handling: _____

_____ Mark here if more than 7 transforms are needed and attach additional sheet(s).

III. Approved by:

1. Government entity's Data Owner of the above noted datasets (e.g. dep't head, bureau director, etc.):

Printed Name: _____ Title: _____

Signature: _____ Date: _____

2. Government entity's Point of Contact: (e.g, Data Coordinator)

Printed Name: _____ Title: _____

Signature: _____ Date: _____