

Agency Self-Publishing Guidelines

Why self-publish data?

- Create and manage your own data pipelines
- Tweak metadata or data dictionaries for already-public datasets yourself without additional approval
- Manually update already-public datasets yourself without additional approval
- Stage datasets yourself to share with City staff so they can see exactly what the data will look like when its published on NYC Open Data

Requesting self-publishing access

- Email the Open Data Team at opendata@oti.nyc.gov
 - Requests for self-publishing access will be evaluated by the Open Data Team based on the team's capacity to support the requesting agency, the agency's publishing volume, and the agency's record of sharing high-quality data
- If a self-publishing request is approved, the Open Data team will arrange for a tutorial on the publishing process with the approved agency

Guidelines for Publishing on NYC Open Data

- Data will not be made public unless it has been reviewed against and verified to meet [NYC's Open Data Quality Standards](#) by **both** the agency's Open Data Coordinator and the NYC Open Data Team
 - This includes complete documentation, with a data dictionary that follows [NYC's data dictionary template](#)
 - Any agency self-publishing data must ensure that all metadata on NYC Open Data is completely and accurately filled out, including:
 - The dataset row label "Each row is..."
 - Descriptions for the dataset and each dataset field
 - The dataset update and data change frequency
 - Dataset tags
 - Data needs to be in a flattened data format (for example, CSVs, relational database tables/views, etc.)
 - Data cannot be in a nested structure (for example, complex JSON objects)
 - Datasets containing geospatial data will not be made public without an accompanying map unless the agency and Open Data Team agree that a map will reduce useability and/or increase confusion
- Until the Open Data Team's review and approval, published datasets will only be privately visible to NYC users who have [explicitly been granted access](#) by the agency ODC or NYC Open Data Team
- When a dataset is ready to be shared publicly, **email** opendata@oti.nyc.gov to request that the Open Data Team review any new datasets and approve them for publication. Submitting the dataset for publication approval through the Tyler publishing interface is not sufficient.
- For assistance with self-publishing please contact the Open Data Team at opendata@oti.nyc.gov. If necessary, the Open Data Team will escalate to the platform vendor, Tyler (formerly known as Socrata), for additional assistance.

Self-Publishing Technical Reference Materials

Depending on agency setup, select the best method to publish data onto NYC Open Data:

Name	Link	Additional Notes
Socrata Publisher Documentation (Getting Started Guide)	https://dev.socrata.com/publishers/	This gives an overview on how to publish data onto NYC Open Data.
Socrata Gateway	https://support.socrata.com/hc/en-us/sections/360005092553-Socrata-Gateway https://support.socrata.com/hc/en-us/articles/360033426994-Data-Insights-Gateway-for-Data-Publishers	This is the newest way to publish data onto the NYC Open Data. It requires a "user agent" to be set up on a server with access to the source system hosting your data. Various plug-ins are available to the agent to connect to different source systems, including flat files, databases both on-premise and in the cloud, and geospatial services. Tyler recommended this to OTI. While this works fine for simpler use cases, we are currently not using this approach due to our custom requirements.
Socrata Publishing API (Socrata Data Management API)	https://socratapublishing.docs.apiary.io/	This is the underlying API used by the Socrata Gateway. This is especially useful if custom logic is required that the Socrata Gateway does not handle. This is Tyler's recommended approach to data publishing if customers need more features beyond what is offered by Socrata Gateway.
Socrata DataSync	https://socrata.github.io/datasync/	This is an executable Java application which has a graphical and command-line interface for manual or scheduled jobs where the source data is a <u>flat file</u>. It is also a Java SDK which can be used in a Java application. This is currently how OTI publishes data onto NYC Open Data.
SODA Producer API	https://dev.socrata.com/publishers/soda-producer/soda-producer-basics.html	This is a RESTful API to publish onto NYC Open Data. There are <u>libraries</u> written in various languages that use this API.
Libraries & SDKs	https://dev.socrata.com/libraries/	This is a link to the various libraries and SDKs developed by Socrata for a number of different languages.

Change Log

- 11/07/2022 – Initial release