

Feasibly Automating Your Data

This document should provide guidance in filling out Tab A “Current Inventory” in the compliance plan workbooks regarding the feasible automation of data.

Local Law 251 of 2017 requires all agencies to report if their current and future inventory of datasets can be feasibly automated. Automation (with regard to NYC OpenData) is the automatic updating of a dataset without human intervention.

The goal of this legislation is to help identify (and eventually automate) data that can and should be automatically refreshed. Automation helps reduce the time and resources required to update datasets, helps provide more consistent and reliable data to the public, and makes your job as an Open Data Coordinator (ODC) easier.

The Open Data team’s goal is to automate as many datasets as possible, and our assumption is that the large majority of datasets can in fact be feasibly automated. If there are datasets you do not believe can be feasibly automated, please reach out to the Open Data team to discuss this further.

There are currently over 230 datasets being automatically updated on NYC Open Data. For reference, here are a few brief examples of how those automations work.

- 311 – [311 Service Requests from 2010 to Present](#) are maintained in an internal data warehouse managed by DoITT. The Open Data team pulls the data from this location and publishes it to NYC Open Data.
- HPD – [Housing Maintenance Code Violations](#) are automated via a process where HPD places their data files on a FTP server, and the files are picked up by the Open Data team and published to NYC Open Data.
- DoITT – [Building Footprints](#) is maintained by the Geographic Information Systems (GIS) team at DoITT. These map files are placed on an FTP server which are picked up by the Open Data team and published to NYC Open Data.
- DOHMH – [DOHMH New York City Restaurant Inspection Results](#) data is stored in a DOHMH internal database. The Open Data team is provided with a read-only view of this database, and agreed upon records are pulled and published to NYC Open Data. **This is the preferred method for automating data.**

We recommend that you set up time between April – June to meet with your agency’s IT team and go over your Agency’s existing & planned open data assets to review what can / cannot be feasibly automated and in which timeframe. For clarification – if your agency determines that a data asset can be feasibly automated it does not mean that the data asset must be feasibly automated in the near future though the DoITT Open Data team will work with you to determine a feasible timeline for implementation of the automation in the coming months / years.