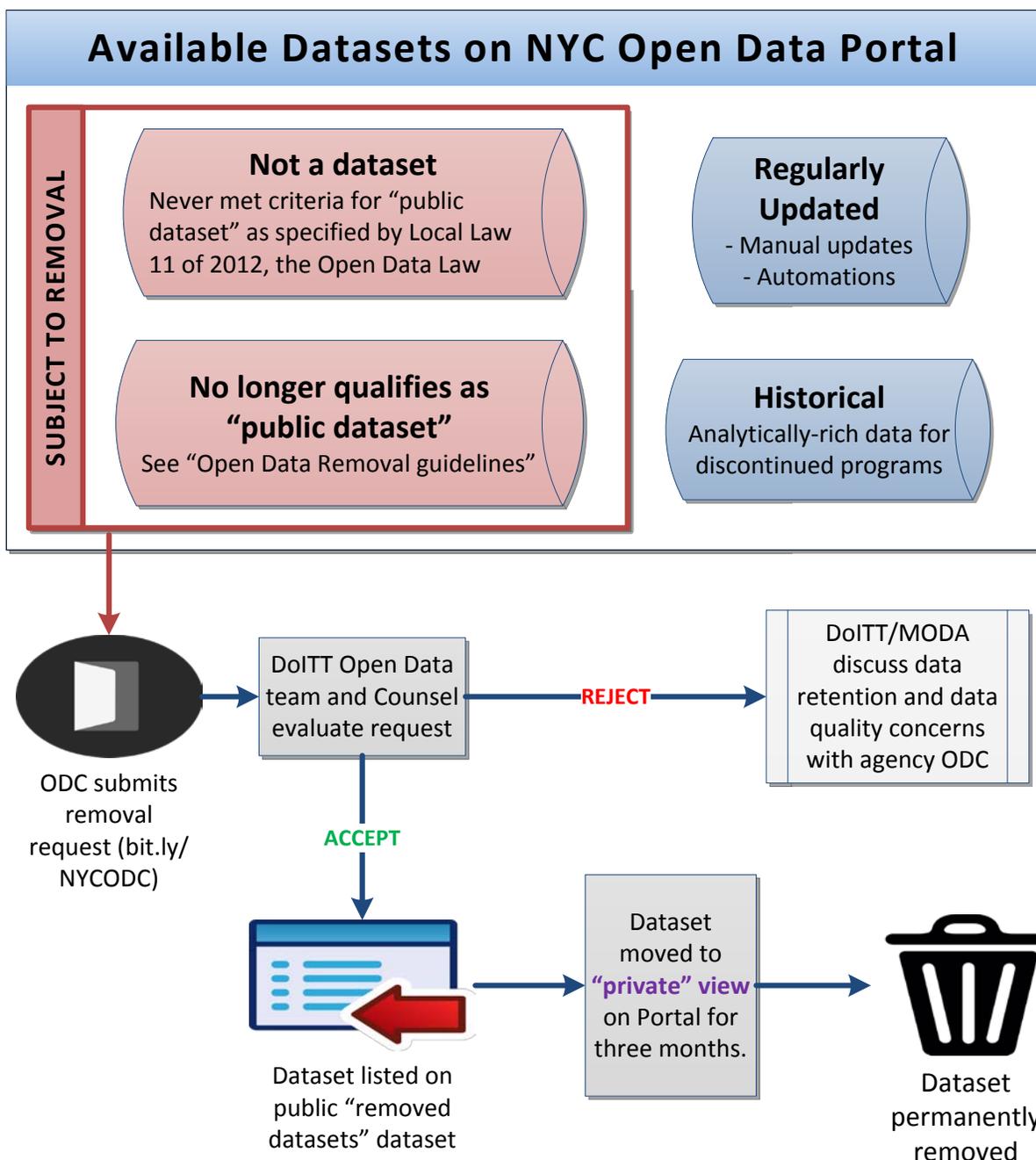


Removing data assets from the Open Data Portal - Process

The definition of “**public dataset**” in the Open Data Law has many elements and exceptions (see: NYC Administrative Code §23-501). Ultimately, the question of whether a dataset qualifies as a public dataset is a legal question for your agency’s General Counsel.

In some instances, a data asset (i.e., a table, attachment, or external link on the Socrata platform) *no longer* meets the definition of “public dataset.” In others, data published shortly after the open data law *never* met the definition of public dataset. Under these circumstances, the Open Data Team at DoITT and MODA may remove data from the open data Portal.

This purpose of this process is to remove low-value datasets so that users find more useful and up-to-date datasets when searching the Portal. To get started, ODCs may submit a removal request through the ODC request form at bit.ly/NYCODC.



Removing data assets from the Open Data Portal - Guidelines

According to Local Law 11 of 2012, the Open Data Law, your dataset may be removed if...

1. The dataset is not a “comprehensive collection of interrelated data.”

Is the underlying database your agency maintains no longer comprehensive? Is there a more comprehensive collection of data available? It may not be a public dataset any longer. Perhaps the data is part of a larger, more comprehensive dataset. **Hypothetical:** *In 2012, an agency piloted a program measuring sidewalks in The Bronx. But now in 2017, that program encompasses the entire city. The Bronx-only data does not qualify as a public dataset because it is no longer comprehensive. It may be removed from the Portal in favor of publishing the citywide sidewalk dataset for all boroughs.*

2. The dataset is not a “final version of statistical or factual information.”

Your agency is not required to publish draft versions of data: only final versions. If your agency mistakenly published data which you now realize was not the final version of the data, it may be removed from the Portal.

3. The dataset was not created “by or on behalf of and owned by an agency” or does not record “measurement, transaction, or determination related to [the agency] mission.”

If your agency no longer owns the underlying database from which the public dataset is generated, or your agency’s mission statement has changed, the public dataset may qualify for removal from the Portal. However, if the data is valuable for historic purposes, the dataset should remain on the Portal and its update frequency should be changed to “Historical Data.” **Example:** *The Fire Department’s Important Phone Numbers is a two-column list of phone numbers, which are administrative and not a “measurement, transaction, or determination” related to FDNY’s mission, making it eligible for removal.*

4. The source data does not belong to your agency.

The responsibility to publish data lies with the agency that produces and owns the data, not any agency that may be holding it. If your agency published data belonging to another agency, it may be removed from the Portal, in which case the Open Data Team will work with the agency that owns the data to publish it from the source. **Example:** *The Mayor’s Office of Data Analytics’ Multi-Agency Permits contains records on DOHMH and DOB permitting activities. MODA’s use case for publishing this data is no longer active and the information is duplicated in other published DOHMH and DOB datasets, making it eligible for removal.*

5. “Removing the redactable portions [of the dataset] is an undue financial or administrative burden.”

If recent changes to your data have made it incredibly difficult to redact the portions that shouldn’t be published (e.g. personally identifiable information), it may be removed from the Portal.

6. The dataset is stored on “a single computer or a portion of the City network exclusively assigned to a single employee”

The Open Data Law requires agencies to publish their canonical data, not a single employee’s working copy of that data. **Hypothetical:** *An agency maintains a database of its fleet. One agency employee pulls a report for a subset of vehicles used in the past 90 days out of that database into a spreadsheet on their desktop, and publishes it to the Open Data Portal. This dataset is eligible for removal.*