



Metropolitan Transportation Authority

State of New York

April 18, 2022

The Honorable Kathy Hochul
Governor of New York State
NYS State Capitol Building
Albany, NY 12224

The Honorable Andrea Stewart-Cousins
President Pro Tempore & Majority Leader
New York State Senate
330 State Capitol Building
Albany, NY 12247

The Honorable Carl Heastie
Speaker
New York State Assembly
932 Legislative Office Building
Albany, NY 12248

Re: MTA Open Data Plan

Enclosed please find the Metropolitan Transportation Authority (MTA) Open Data Plan (the "Plan") pursuant to Chapter 489 of the Laws of 2021.

In this plan, we detail our efforts to publish nearly 100 datasets that achieve our joint mission with New York State Office of Information Technology Services (NYS ITS) to increase responsiveness, improve knowledge of our operation, and respond to needs identified by the public.

The MTA moves millions of people every day, and we work hard to build trust and confidence in MTA services. It's critical that public transit is safe, reliable, and fast — and that the public has input on our ongoing efforts to improve. The open data program will serve a critical role for advocates, journalists, elected officials, and the riding public to participate as we continue to build accountability and transparency.

Thank you to Governor Hochul and the Executive Chamber, New York State Legislature, NYS ITS, and to the stakeholders who have contributed to this Plan for your interest in improving MTA data and the services we provide. We look forward to working with you as we evolve this program.

Onwards,

A handwritten signature in black ink that reads "Sarah Meyer".

Sarah Meyer
Chief Customer Officer, MTA

MTA Open Data Plan

April 18, 2022

Introduction and background

The MTA is excited to publish its Open Data Plan, pursuant to Chapter 489 of the Laws of 2021 and as part of an overall renewed commitment to transparency and accountability through open data under Governor Hochul's Transparency Initiative, memorialized in the [MTA's published Transparency Plan](#).

The MTA has a long history of publishing data, both open data and data visualizations that lower barriers to interpretation of our data. To date, in addition to the data published to the NYS portal [Open Data NY](#), MTA has published data on its own website, including ridership, performance metrics, static and real-time General Transit Feed Specification (GTFS), service information, system information, financials, and assets. Information is also published as monthly reports prepared for the MTA board and its related committees.

With this Plan, we are undertaking a concerted effort to consolidate datasets, codify and automate procedures to maintain all data, and improve the quality and granularity of our data.

This is a living document. Our aim is to be responsive to evolving operating realities and ridership needs. Through continued stakeholder engagement and feedback, we will advance our data practices as this program develops.

How we developed this plan

Inputs

The amount of data collected, processed, and analyzed by MTA is tremendous. In its rawest form, our data includes:

- records related to the subway, bus, rail, and paratransit service we operate, including scheduled service and service as operated
- workforce data
- transactional records, including fare and toll records
- financial data

In preparing this Open Data plan, we have reviewed much of this data that we utilize internally. We have paid particular attention to the data that we share publicly, including what is currently presented in our MTA Board and Committee materials, on our public-facing performance dashboards, the MTA Capital Program Dashboard, and otherwise published on our websites. We have reviewed past requests submitted under the Freedom of Information Law (FOIL) to identify frequently requested data.

We have additionally solicited direct input to the Open Data plan from our external stakeholders. We have met with a number of transit advocates, open data experts, and good government groups and assessed the priorities for data that they have offered. We have also invited the general public to contribute to our plan and have considered the input we have received. As was noted earlier, this type of engagement will continue as we implement the Open Data plan and while this provides an initial catalog for publication, and a schedule thereof, this will be an ongoing effort to provide important information to the MTA’s stakeholders and those with oversight over the Authority.

Assessment and prioritization

Given the sheer volume of data MTA produces, we have assessed the identified datasets to both determine which to include in the program and how to prioritize when our teams will publish each. The [NYS ITS Open Data handbook](#) outlines criteria to use for this, with value, quality, and readiness/availability being the primary measures. In terms of value, it lists these objectives, which we have used in our assessment:

- Increases accountability and responsiveness
- Improves public knowledge of agency and its operations
- Furthers the mission of the agency
- Creates economic opportunity
- Responds to a need or demand identified after public consultation

Scope

MTA’s Open Data catalog included in this plan consists of 99 datasets, which we have categorized as follows:

Category	No. of datasets
Operational metrics and data	34
System usage	22
Safety / Security	8
Financial	6
Workforce data	5
Customer feedback	5
Miscellaneous	4
Accessibility	4
Real-time data feeds	4
Capital program	4
Assets / geographical information	3

[View the full MTA Open Data catalog here.](#)

MTA publishes a number of real-time data feeds that provide detailed information about our public transportation services, primarily for customer information purposes. We maintain a Developer Resources microsite (mta.info/developers) from which these are available. We have included links to these fields within our Open Data catalog. Due to the volume of data that is transmitted through the feeds, we generally will not be publishing the entirety of that data in a historical repository. However, many of the performance metrics we are publishing are the result of aggregation of this operational data.

Following review of NYS guidance and datasets published on data.ny.gov by other State departments and agencies, we have not included within this plan one-off project- and study-specific data that is collected or produced, and then analyzed in our planning efforts.

We have not included within the open data catalog anticipated datasets that we are not yet producing. Again, this plan is a living document, and we will add future datasets at the appropriate times.

As we proceed with implementing this plan, we will build tools and expand our internal data infrastructure and pipelines, which we expect will reduce the per-dataset effort to initiate and routinely publish our data. Additionally, the normal replacement and improvement of data systems will contribute to this improved efficiency. We therefore expect we will add datasets to the catalog as this internal effort is reduced.

Implementation

We began expanding MTA's Open Data in December 2021 by publishing a substantial amount of performance metrics data. The catalog and timeline within this plan pick up from there and lay out how we will further expand our Open Data as well as improve the data we are already publishing.

Implementation of this plan follows a sequence prioritized broadly as follows:

- Continue publishing data supporting board and committee books
- Enhance the published performance metrics datasets to increase their value
- Publish financial data (budget and actuals)
- Enhance the capital program data we are currently publishing

Many of the datasets we have published to date can be improved, by adding supporting data fields that would provide greater context to the calculations or other fields within the dataset, improving the dataset structure to ease analysis, and/or increasing granularity. In some cases, we may choose to consolidate datasets where we believe that will improve analysis.

We will continue to solicit public input to our Open Data program, including both suggestions of additional datasets to publish and feedback about dataset structure. MTA will also participate in NYC Open Data community events.

We plan to improve the narrative description accompanying these datasets to better describe how the data is collected and used. Where applicable, we will clearly show the links between our Open Data and the relevant performance metrics in our Board and Committee materials and on the public performance dashboards (dashboards.mta.info).

Automation

99 datasets represent a large amount of data. In most cases, these datasets need to be refreshed monthly with new data, with 6 having shorter than monthly refresh cycles, including our popular Daily Ridership dataset that we update each weekday. To avoid being an excessive burden on internal resources, we are developing processes to automate the publishing of these refreshes. Additionally, we are building tools to remind staff of manual steps that need to be conducted and to alert our Open Data team if automation steps were not successfully completed or if datasets have not been refreshed per schedule.