



# THE METROPOLITAN TRANSPORTATION AUTHORITY OPEN DATA PLAN 2023 ANNUAL UPDATE





# BACKGROUND

The MTA is thrilled to publish this update to the Open Data Plan as a continued demonstration of its commitment to transparency and accountability through open data. Open data is, by definition, readily available and accessible to all users in a machine-readable format, with no restrictions on its use. As a public benefit corporation, the MTA is excited to operate with greater transparency by working toward robust open data offerings. We are proud to have incredibly engaged employees, customers, and stakeholders who want to do more with our data and are committed to supporting their needs as best we can.

In 2022, the responsibility of the MTA Open Data Program transferred from Customer Experience to the newly formed Data & Analytics team at MTA HQ. This move will allow for more seamless data management, improved pipelines and processes, and ultimately more efficient publishing of new datasets. The team is led by MTA's Data Coordinator, Jon Kaufman, Chief of Strategic Initiatives.

Under the Data & Analytics team, sustainability and movement building are two primary goals of open data program management. We want to ensure that any datasets shared publicly are built from clean, well-documented data pipelines that are as automated as possible. We also want to share ownership over as much of the new dataset creation and refresh process as possible, so that individual data owners remain plugged in to the full open data lifecycle. And finally, we want strengthen the shared vision of why open data matters across our internal MTA employees.

We are excited to share our progress and always look forward to hearing from you. Contact our team at [opendata@mtahq.org](mailto:opendata@mtahq.org) and we will be in touch!

# YEAR IN REVIEW: OPEN DATA BY THE NUMBERS

2022



**1st** New York State Agency or Public Authority to publish our Open Data Catalog as a dataset on [data.ny.gov](https://data.ny.gov)

**42** new datasets published in 2022

**46** data stewards identified to help maintain our open data

**58** visualizations built on [metrics.mta.info](https://metrics.mta.info)

# IMPLEMENTATION

In 2022, our strategy for opening up our data was focused on bringing already public information (e.g., from our Board books and public dashboards) to [data.ny.gov](https://data.ny.gov) in easy-to-use tabular formats. The MTA has for many years shared data in ways intended for presentation or display, but we are moving quickly toward sharing in machine-readable formats.

To make our data more accessible as quickly as possible, some of these new datasets in 2022 were published in formats and granularity based on legacy systems and software. While this provided a huge benefit to external data users who previously combed through PDFs or data visualizations to transfer data points into a tabular format, it also meant more to clean up in the future when improving data quality and granularity.

In the summer of 2022, we began work on [metrics.mta.info](https://metrics.mta.info), a data visualization tool that retired MTA's practice of sharing data primarily through PDFs. The goal for our metrics site is to have it built fully from open datasets. While this is not yet completed, users are able to download data directly from the site when data is not yet available on [data.ny.gov](https://data.ny.gov).

After the launch of [metrics.mta.info](https://metrics.mta.info), we turned our focus to building internal support for open data and better analytics. In December 2022 we launched the Data Leads forum, which brings together leaders from across various MTA agencies and departments to address common issues on topics like data quality, data management, and data resources. The forum now meets bi-monthly and will support the MTA's goals of sharing more and better data.

## **AUTOMATION**

We have made enormous progress in streamlining the data upload process to remove manual steps. While only a few datasets currently run through a completely automated upload process, 2022 was a year of exploration where we discovered what is possible in removing manual steps from the publication pipeline. Automation is important since reducing the workload associated with each additional dataset will allow MTA to share more in the future.

## **PUBLIC ENGAGEMENT**

Since 2022 was a year of transition in program management, getting oriented to the workload and issues ahead was our team's number one priority. We did speak at a public event in November 2022 with R-Ladies, an organization dedicated to promoting diversity in data science. Just in the first few months of 2023, we have made significant progress and will be joining many more events this year to share the details with the public.

## **GOING FORWARD**

In Q2 2023, we will release an updated version of the Open Data Catalog to account for changing priorities, common stakeholder requests, and the need to improve data quality on existing datasets. This first quarter of 2023 has been a huge success for the team, and we look forward to continuing this momentum through the rest of 2023 and beyond.