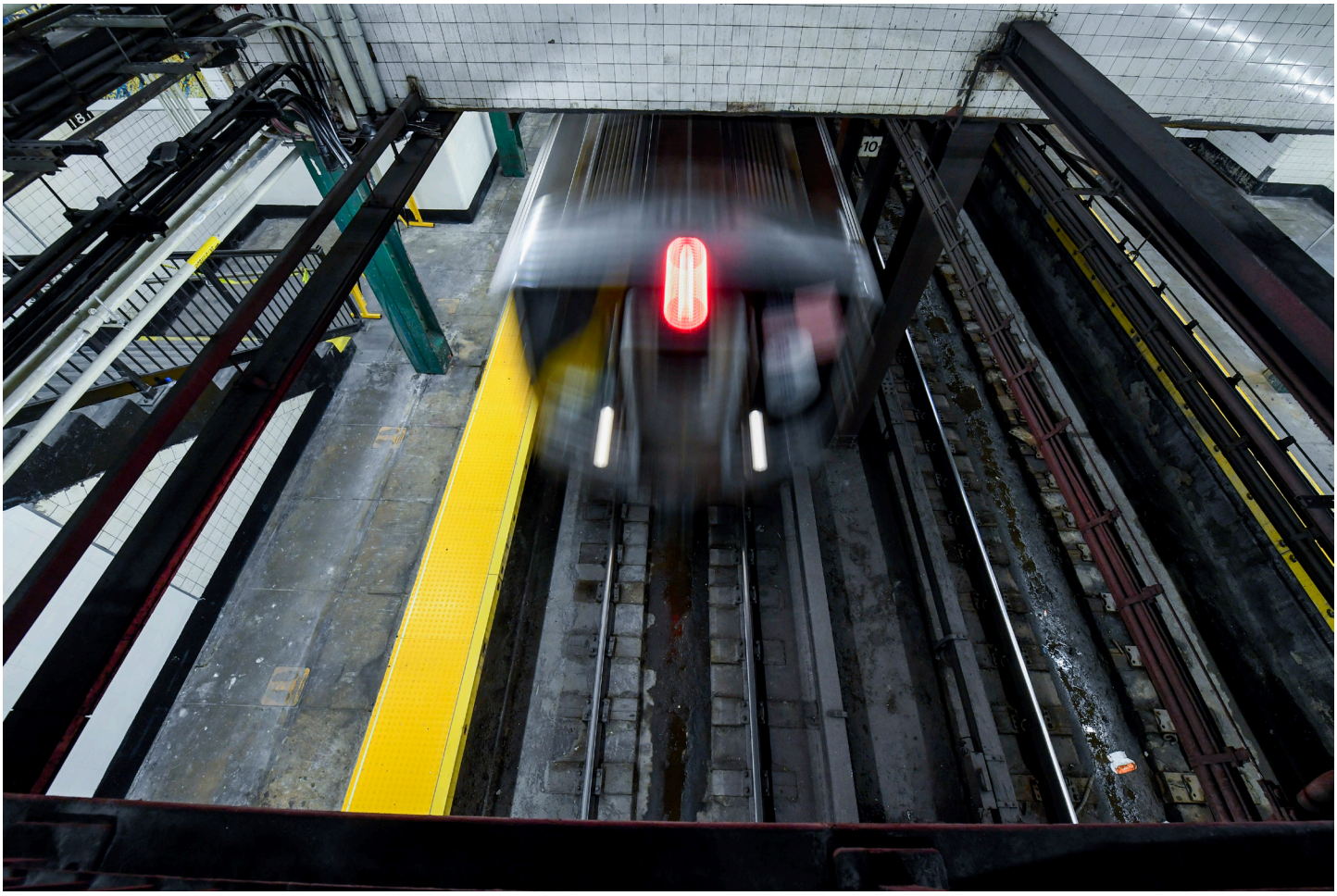
THE METROPOLITAN TRANSPORTATION AUTHORITY

OPEN DATA PLAN
2024 ANNUAL UPDATE



OVERVIEW

In October 2021, New York State Governor Hochul signed the MTA Open Data Act into law, prompting the MTA to renew our commitment to proactively sharing high quality, machine-readable data with the public through data.ny.gov.

We agree that open data and increased transparency is in everyone's best interest, including the MTA's. Open data can increase civic engagement and trust in government by allowing the public to engage with government officials with suggestions and solutions to improve its operations.

The MTA Open Data team is proud to release its second annual update to the original open data plan. 2023 was a year of immense growth for the program, and the team is looking forward to continuing to share more and better data for all.

YEAR IN REVIEW: OPEN DATA BY THE NUMBERS

2 full-time open data scientists hired and onboarded	**204,681** dataset downloads	**7** data dashboards consolidated
720 attendees at open data events	**14** external-facing stakeholder engagement events	**50** datasets published or upgraded

2023 NEW DATASET HIGHLIGHTS

We published a lot of new and newly cleaned up datasets in 2023! While there are too many to effectively detail in this update, we do want to feature a handful of our favorite releases.

HOURLY SUBWAY AND BUS RIDERSHIP

Two of our biggest (literally and figuratively!) new dataset releases this year were about subway hourly ridership and bus hourly ridership. For the subway, hourly data is broken down by station-complex, payment method, and fare class. For the bus, hourly data is broken down by route, payment method, and fare class. This is a new level of data granularity for the MTA, and it has been well received.

20-YEAR NEEDS ASSESSMENT

The MTA public five open datasets in conjunction with the release of the MTA's 20-Year Needs Assessment, a comprehensive blueprint that outlines the agency's long-term capital needs. This is the first time the MTA has publicly provided data on the condition of most of its assets and fleets. Additionally, part of the 20-Year Needs Assessment includes the first-ever quantitative and comprehensive assessment of the costs and benefits of potential expansion and enhancements. This "Comparative Evaluation" dataset includes criteria such ridership, time savings, and cost of potential projects.

STATEN ISLAND RAILWAY

We published three datasets in the fall focused on Staten Island Railway for On-Time Performance, Scheduled Trips, and Mean Distance Between Failures. These are the first standalone datasets for the sole rail line on Staten Island and go back to 2006, farther than nearly all MTA datasets on Open NY. The Scheduled Trips dataset is the first by the MTA to show scheduled service levels over time, and the On-Time Performance data includes delays caused by holding trains for connections with the Staten Island Ferry.

OPEN DATA IN ACTION

As the MTA continues to progress in publishing new and better datasets that can be used to draw actionable insights, we also dedicated time toward building resources and tools that will better help data users understand and work with our data.

In early 2023, we published our Open Data User Personas, a human-centered design tool that facilitates the consideration of real data users with a variety of goals, needs, and technical abilities. We designed our user personas through multiple stakeholder workshops with our existing open data community and refer to them often as we build out new tools and resources to help folks explore MTA data.

We also published a GitHub repository for Open Data Week 2023, with the goal of demonstrating how to use Python to analyze and visualize MTA Open Data sets available on data.ny.gov. This session gave a peak into how the code for metrics.mta.info is built and provided the data user with sample Python code they could use to process, analyze, and visualize open datasets. The live session, which is available on YouTube, was a great success!

Improving our public-facing dataset documentation was also a focus this year. Starting with our subway performance metrics datasets, we enhanced the clarity, completeness, and accessibility of the dataset documentation, focusing extensively on the dataset methodology sections of the overview documents that accompany each dataset.

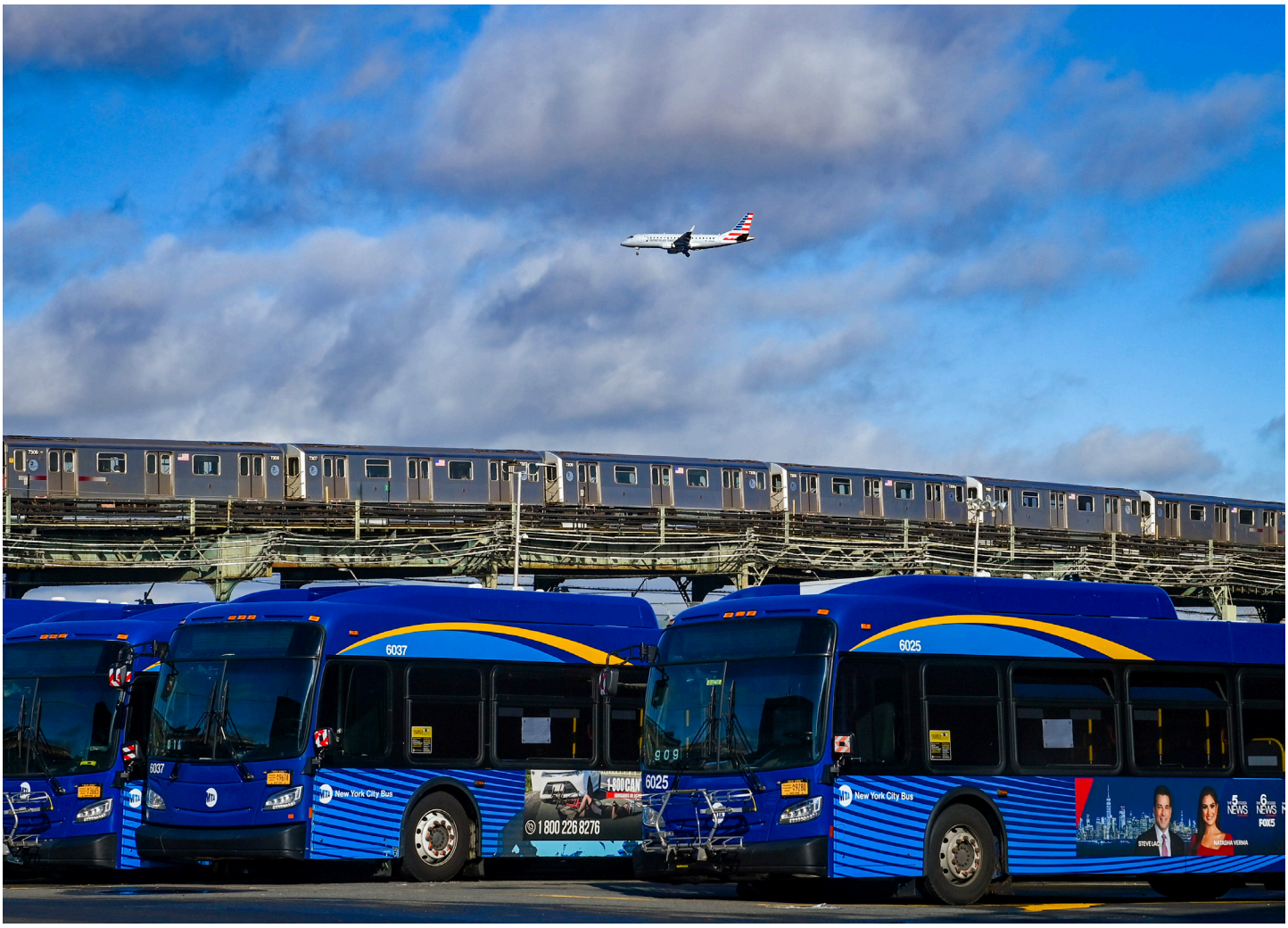
At the end of the year, we launched the MTA Data & Analytics blog, offering a platform for sharing insights, updates and best practices related to MTA data. The blog features contributions from the broader Data & Analytics team and will cover a wide range of topics that provide further insight into data at the MTA. Through the blog, we aim to foster a community of readers and supports of MTA data and analytics, encouraging collaboration, knowledge sharing, and continuous learning.

And finally, throughout the year, we continued to build out our open data visualization site, metrics.mta.info, enhancing its functionality, performance, and user experience based on feedback from our stakeholders. We added new interactive features, expanded the range of available visualizations, and further implemented data connections with the New York State open data portal. Our ongoing efforts have empowered users both inside and outside the agency to make informed, data-driven decisions and drive positive outcomes for New York City’s transportation system.

INFRASTRUCTURE

The MTA is committed to developing a sustainable open data program. In line with this vision, we dedicated time in 2023 to developing robust internal data infrastructure, which in the long-term will significantly enhance the efficiency of our open data program. In short, we cannot provide public stakeholders with timely and comprehensive data without having access to this data ourselves internally, so beginning to build out centralized internal data infrastructure was a necessary step to ensure the success of MTA's open data program.

To read more about our data infrastructure, check out our blog post How we build analytics at scale at the MTA.



2024 FOCUS AREAS

This year should be an even bigger and better year for MTA Open Data! We have a lot of ambitious goals on our plate, including internal-facing work such as finalizing our internal infrastructure needed to support the program and continuing to automate our dataset publication. We also hope to open source the code for metrics.mta.info and create clear guidelines for members of the public to contribute new visuals and code improvements to the project.

New datasets slated for publication can be found in our publication plan, but we did want to highlight a few big releases to expect this year.

OPERATING BUDGET DATA

At the end of Q1 2024, the MTA plans to publish the agency's operating budget data in machine-readable formats. Datasets will include the MTA's statement of operations, subsidy information, number of employees by agency and position function, and other critical elements of the operating budget. Currently, this data is published as PDFs for the MTA Finance Committee and regular Financial Plan updates, but is not provided in machine-readable format. This information will still be regularly reported in those forums, but with the added benefit of open data access.

CONGESTION PRICING DATA

Beginning with the launch of the Congestion Pricing program, which will charge drivers a toll for entering Manhattan below 60 Street, the MTA will begin publishing data related to program implementation. Datasets will be released in machine-readable format on a rolling basis and feature information including toll crossings, revenue generated, and vehicle speeds.