METROPOLITAN TRANSPORTATION AUTHORITY
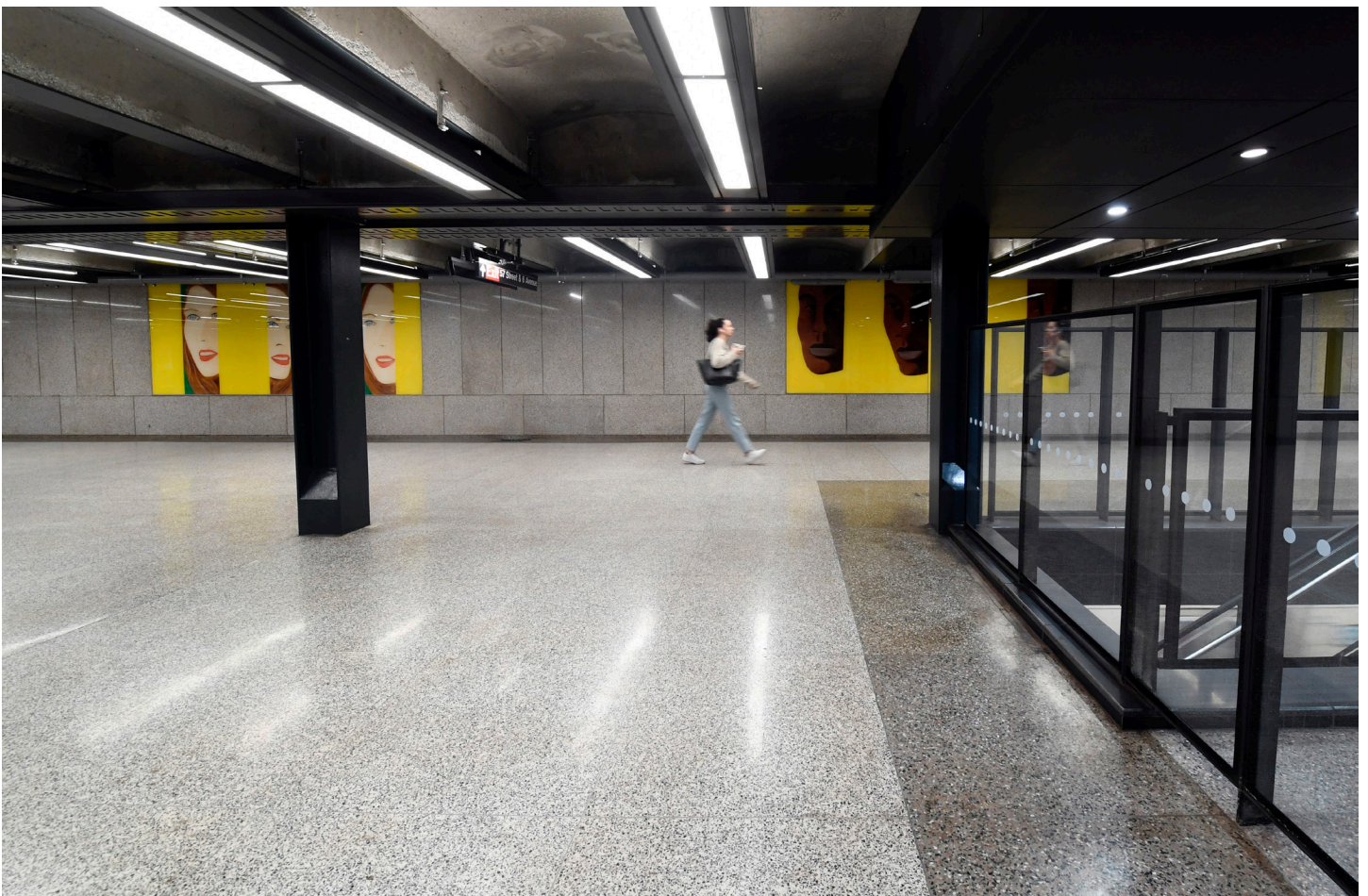## OPEN DATA PLAN
2025 ANNUAL UPDATE

The Metropolitan Transportation Authority (MTA) and Open Data team is proud to present the third annual update to the Open Data Plan, reinforcing our commitment to transparency, accountability, civic engagement, and data-informed decision-making. Open data—information that is publicly accessible, in a machine-readable format, and can be freely used and shared—plays a critical role in fostering trust and collaboration with researchers, developers, policymakers, advocates, and riders.

2024 marked three years since Governor Hochul and the New York State (NYS) Legislature enacted the MTA Open Data Law. In this time, our team has rapidly grown the agency's portfolio of open data offerings, including over 160 open datasets and several supporting resources that empowers decision-makers and riders alike to improve mass transit in New York.

Looking to 2025, our Open Data Program remains committed to providing the public with transparent, high-quality data that enables stakeholders to contribute to and advocate for a more efficient, reliable, and safe transit system for all New Yorkers.

2024 YEAR IN REVIEW: OPEN DATA BY THE NUMBERS

28
NEW DATASETS

144
SUPPORTED OPEN DATASETS

115,979,419
RECORDS IN OUR LARGEST DATASET, BUS HOURLY RIDERSHIP

116,097
DOWNLOADS OF OUR MOST POPULAR DATASET, DAILY RIDERSHIP

20
SPEAKING ENGAGEMENTS

1050
ATTENDEES AT OPEN DATA EVENTS

2024 NEW DATASET HIGHLIGHTS

2024 was a year of bigger datasets and more data automation. We published over 10 years of delay-level data for Metro-North Railroad and Long Island Rail Road, enormous datasets on subway origin-destination ridership estimates, granular data on bus speeds at the timepoint segment level, and detailed data on subway end-to-end running times.

In the first quarter of 2024, we published MTA's operating budget in open data format, making us the first transit agency in the United States to achieve this. This marked a major milestone for the MTA, as previous efforts to improve transparency around operating budget data were labor intensive and reliant on repetitive manual processes.

The operating budget is available in three datasets: Statement of Operations, Subsidies, and Headcount. This data release was accompanied by public engagement efforts, including a series of public-facing events we dubbed our "budget press tour". During these events, we held live demonstrations showing users how to use these datasets and explained how we built them. We also published a detailed post on the Data & Analytics blog, offering a deep dive into how we approached the technical challenge of creating these datasets.

Shortly after the release of the operating budget open datasets, we published an archive of GTFS real-time data with the Service Alerts dataset. This dataset provides a historical record of service disruptions, planned work, and other incidents that impact service. By making this available on open data, researchers, advocates, and policymakers have a new tool available to analyze incidents and service disruption trends over time.

We also dedicated months preparing data to support the launch of the Congestion Relief Zone. In 2024, we published a variety of supporting datasets to help analyze congestion relief's impacts, such as bus routes and taxi zones in the Central Business District (CBD); the CBD Geofence, which is a geographic dataset of the tolling boundary that is being used by companies like Google Maps and Waze; and taxi and for-hire vehicle trips in the CBD.

OPEN DATA IN ACTION

In the past year, we dedicated efforts internally to refine our existing resources and decision-making processes, ensuring we can provide high-quality and transparent data offerings. We also continued to engage with our data users who support the Open Data Program and are always eager to explore more MTA data.

PUBLIC ENGAGEMENT

In fall 2024, we launched the MTA's first Open Data Challenge and called on all interested stakeholders from across the country to build a product, such as a web app, data visualization, written report, map, or even a piece of art, using the MTA's open data collection. Participants had one month to create and submit their projects, and we were overwhelmed by the enthusiastic and creative responses. We received over 100 submissions from students, researchers, policymakers, developers, and everyday customers of incredible quality and scope, with participants making use of the MTA's breadth of open data offerings.

Throughout the year, we also continued hosting and presenting internal and external events to showcase new data releases and tout the success of the MTA's Open Data Program. Some notable events from the year include traveling to Washington, DC to present at the Transportation Research Board, NYC Open Data Week, updating the MTA Board on our progress and achievements, and launching our budget press tour. In one of the most exciting events of the year, we participated in the CUNY Macaulay Honors College Datathon, which challenged students to solve real policy challenges facing NYC using MTA open datasets.

INTERNAL IMPROVEMENTS

In 2024, we created the Open Data Steering Committee, marking a major shift in how we facilitate internal decision-making on open data releases. The Committee suggests and approves high-level, priority subject areas for the open data team to focus resources. The Committee convenes an annual planning meeting and approves new datasets and topic areas, which will be included in the release catalog in the following year. Datasets requested by internal or external stakeholders, are related to agency priorities, or fill a gap in the current open data catalog are prioritized for discussion during the annual planning process.

Earlier in the year, our developers also reached a big milestone: all data visualizations on MTA's performance metrics dashboard, metrics.mta.info, are built using datasets available on the Open Data Portal. The website is also built with open-source tools, which means metrics.mta.info could be built by anyone, as all the information needed to replicate it is fully accessible!

Finally, we improved our dataset metadata (data about the data!). Some datasets have been on the Open Data Portal for over a decade, and the accompanying metadata could be stale or even inaccurate. In 2024, we cleaned up the metadata to ensure consistency across all datasets. We also standardized how we categorize datasets, making them easier to find through the catalog search feature.

2025 FOCUS AREAS

We've accomplished a lot in three years, and we're excited to take the Open Data Program to new heights. In 2025, our team will focus efforts to publish more geospatial data to the Open Data Portal. Today, basic spatial data about the locations of our train stations, bus stops, and routes is inconsistent across operating agencies; as we focus resources here, information about our core infrastructure and service can be useful for interested mapmakers.

This year, we're also aiming to grow the depth of our data offerings. So far, our program's primary quantitative measure of success is the number of datasets we have published. While this metric is important, measuring our program by this number alone is flawed as it does not capture the depth of data available in our datasets. We are excited to dedicate time this year to enhancing the contents within datasets when possible.