OFFICE OF TECHNOLOGY AND INNOVATION TESTIMONY BEFORE THE NEW YORK CITY COUNCIL COMMITTEE ON TECHNOLOGY
Oversight–Open Data Compliance
February 27, 2024

Good morning, Chair Gutiérrez and members of the Committee on Technology. My name is Martha Norrick and I am New York City's Chief Analytics Officer and Deputy Commissioner for Data and Analytics at the Office of Technology and Innovation (OTI). I'm joined today by Zachary Feder, the city's Open Data Program Manager. I'm excited to be here today to discuss our important ongoing work to fulfill the requirements of the city's Open Data Law and our progress connecting New Yorkers with the data that runs their city. I'd like to thank this Committee for its interest in, and support of, the Open Data program.

New York City Open Data is the city's most important public data resource, with billions of rows of data viewed by more than one million people each year. Prior to this administration, New York City Open Data was jointly managed by the legacy Mayor's Office of Data Analytics, or MODA, and the legacy Department of Information Technology and Telecommunications, or DoITT. Staff from each of these offices worked in tandem as the New York City Open Data Team. Technical work, from building the automated processes that move data from agencies to the Open Data platform, was led by DoITT staff, while agency and public-facing work was led by MODA staff.

When Mayor Adams' Executive Order 3 consolidated the city's technology agencies under OTI in January 2022, the Open Data Team was also united within a single office. This new arrangement strengthened the virtuous cycle of rigorous data analysis and high-quality data. The same necessary work to make datasets accessible to the city's data analysts and scientists – ensuring they're structured properly and documented thoroughly – is also vital for making that same data accessible to the New Yorkers who use Open Data.

NYC Open Data functions as a collaboration between our team and every city agency and office. Agencies share their data and their subject-matter expertise. We provide governance, help with data and technical knowledge, and lead broader public engagement.

The Open Data Team has three main functions:

- Supporting Open Data Coordinators (ODCs) and agencies
- Managing the technical platform and pipelines that underpin NYC Open Data; and
- Making public data more accessible for New Yorkers.

Each agency and office has its own ODC, approximately one hundred in total, who is responsible for identifying, structuring, documenting, managing, and communicating about their agency's public

datasets. Nearly all ODCs have other responsibilities in addition to their work on Open Data and regularly work with agency colleagues across information technology, data analytics, legal, and communications roles to fulfill their duties. It is also important to note that, given the breadth of information that most agencies share on Open Data, ODCs serve primarily as facilitators between their agencies and the Open Data Team. They rely on experts across different divisions and teams to contextualize the data their agency publishes. They then use that same context to create data documentation and respond to public inquiries about this data.

Supporting agencies and ODCs is critical given the number of agencies contributing to Open Data and the variety of forms that data can take. To start, we create and promote Open Data standards and best practices – derived from our own experience, conversations with city colleagues, and learning from other municipalities. For example, in 2022, we released a simplified "data dictionary" template for more comprehensive and easier-to-use data documentation, updated the Technical Standards Manual to refocus the document on ODCs and the dataset lifecycle, and published a data quality standards checklist used to vet datasets before publication.

Our technical work is focused on keeping datasets up-to-date. In recent years, our team has employed new tools to build automations quicker. These automations transfer data from its source city database to the open data portal at a specified frequency. We have also piloted an initiative to empower some agencies to create and manage their own updates and automations.

Accessibility is a hallmark of our Open Data initiatives. Our thinking around accessibility is directly informed by what we've learned from New Yorkers, best practices across the country and around the world, and NYC's Open Data Law. We remain focused on helping more New Yorkers understand and make meaningful use of this data through public programs such as the Open Data Week festival, which we're celebrating this year from March 16th through the 24th, and Open Data Ambassadors, which has brought introductory classes to academic institutions, library branches, and local electeds, including members of this committee. These initiatives are products of our longstanding partnerships with the civic technology nonprofit BetaNYC and art collective Data by Design as well as the enthusiasm of many New Yorkers who have volunteered to share their skills and knowledge.

I thank the Committee on Technology for the opportunity to testify today and will now take Council Members' questions.

###

To: NYC Council – Committee on Technology
From: Noel Hidalgo, Executive Director of BetaNYC

Re: Open Data Hearing

Tuesday, 27 February 2024

Dear Chair Gutiérrez, fellow Council Members, and Staff,

Now, more than ever, the City's data analytics team and the open data program need resources to ensure other agencies can do more with less. Since its inception, the data analytics team has streamlined internal and external data-sharing practices. Their work has improved people's lives.

We're a quarter of the way through the 21st Century, on the verge of a climate catastrophe, and need to ensure that the human investments we make at this moment will carry us through the middle part of the century. Consumer technology systems will morph rapidly, and the climate will dictate swift responses to people's needs. In sum, we need adaptable people who will be around for a bit.

Additionally, artificial intelligence and machine learning tools require good, clean metadata. We can NOT build a 21st-century government on the back of systems that contain poor data. Management practices that paper over insufficient data should not be tolerated — people who are juking the stats should be held accountable.

If we're talking about the City's open data program, we need to invest in data analysts and the technologists and service designers they work with. These people can tell us the difference between spread-sheeeets from Shinola.

Opportunities and Insights:

Staffing — The open data team is down five additional people from pre-pandemic staffing levels. While the Mayor has lifted elements of the hiring freeze, we ask that the Open Data Team be restored to its pre-pandemic capacity and add three more to assist with civic engagement, data literacy, and engineering.

Cultivating New Leadership — Agencies must modernize data systems and update the civil service system to ensure we have systems that work for this decade, not the last one. We need government technical talent ready to understand new technologies and ensure that Agencies are not being sold snake oil. We need a new generation of technical leadership that understands the value of service design, data portability and can build upon open technologies.

We must cultivate a new generation of technology leaders to do more with less. The analytics exchange, service design forum, and open data ambassadors are three programs demonstrating how technology/data/design leadership should cut across agencies. This must be combined with apprenticeship programs that prepare various college students to foster a career in public interest technology. Our Civic Innovation Apprenticeship programs can help model future career development paths.

Institutionalizing Open Data Coordinators and Open Data Ambassadors — We need people who know the ins and outs of datasets and can communicate that with others. We would like to see full-time Open Data Coordinators (internally) and Open Data Ambassadors (externally) within major agencies. ODCs and ODAs should be laser-focused on data quality and timeliness of publication; additionally, they are the right people to communicate the intricacies of datasets to broader audiences.

Accountability — As the Council is responsible for holding the Administration accountable, Commissioners and Agency heads must be asked about their internal infrastructure upgrades and agency data management practices at Budget hearings. Every Budget hearing should hold up whatever data is published in the dynamic Mayor's Management Report, have it cross-referenced by data published in the open data portal, and then compare that to what they tell you. FOIL requests and responses should be included in these hearings. The Open Data Team should not be held responsible for what is out of their control.

Data Publishing — Several agencies do not have their act together. The Department of Education continues to be an outlier with nonstandard data publishing practices. Additionally, I do not understand why the NYPD continues to publish monthly Compstat data in Excel spreadsheets at the police precinct level while not publishing data directly to the open data portal.

Digital and Data Literacy — we're on the verge of a transformative shift in how people interact with and incorporate information. If we are not teaching New Yorkers how to look at data or review source data, they will be prey to misinformation. Open Data Week, School of Data, Data Through Design, Open Data Ambassadors, and Mapping for Equity are just a few fundamental tools this government should adequately resource to increase data literacy while battling fake news and misinformation.

Thank you for giving us this opportunity.

Noel Hidalgo

Reinvent Albany Testimony to the NYC Council Committee on Technology For 2024 Oversight Hearing on Open Data Compliance

RE: Huge Backlog of Datasets Awaiting Automation Endangers NYC's Status as an Open Data Leader

February 27, 2024

Good morning Chair Gutiérrez and members of the Committee on Technology. My name is Rachael Fauss, and I am the Senior Policy Advisor for Reinvent Albany. Reinvent Albany advocates for transparent and accountable government in New York State. We were instrumental in drafting and passing New York City's 2012 Open Data Law and subsequent amendments. Thank you for holding this oversight hearing today.

Thanks to the leadership of the City Council, New York City passed the world's first open data law, and our success and failures are closely watched by governments everywhere.

Before getting into our analysis of the 2023 Open Data Report, I want to highlighted two principles that we believe will make NYC Open Data successful:

1. It is essential that the City Council continue to hold annual oversight hearings and look for ways to continuously improve the NYC Open Data program. The Council in the past has actively worked to increase the transparency, accountability, and effectiveness of the City's open data efforts by passing legislative mandates. The success of NYC Open Data depends on public interest and pressure. NYC leaders and career management still have not fully realized the operational and efficiency benefits that open data offers, and some actively hoard their public data and keep it from the public, other agencies, or even their own agency staff. Reporting mandates like the annual compliance plan, data dashboard, and agency audits are essential to compel agencies to comply with the Open Data Law, as is this hearing today.

2. Reinvent Albany believes automating datasets is by far the single most important thing the Open Data staff should be doing. Unfortunately, the Office of Technology and Innovation's (OTI) Open Data team has 435 datasets waiting to

be automated. It took the last twelve years for OTI to automate the 437 datasets currently on the Open Data Portal – and during this time OTI/DOITT had an abundance of developers. We are very concerned by this backlog of automations and wonder why OTI is so behind in this crucial area.

Now I'd like to focus on some highlights from the 2023 Open Data Report, which cover open data efforts over the last few years.

- We support creation of the citywide dashboard with open data compliance metrics (2021)
- The continuation of Open Data Week is essential, with more visibility for open data projects via the "Powered by NYC Open Data" initiative, as well as ongoing partnership with BetaNYC (see 2024 schedule)
- We also appreciate the Office of Technology and Innovation (OTI) publishing and working with vendors on a number of policy documents and updates:
  - Data quality standards (2022)
  - Guide for accessibility/screen readers (2022)
  - Revised, updated technical standards (2023)

These efforts are part of continuous improvement by the Office of Technology and Innovation (OTI) that can help the open data program succeed. However, compliance is dependent on city agencies devoting appropriate attention and staff to open data, and OTI having sufficient staffing to ensure that dataset publishing is on track and that automation of datasets is occurring. Automation is particularly important if the Open Data program is going to be self-sustaining.

The 2023 Open Data Report shows that the open data team at OTI is composed of six people – this is a relatively small number for such a huge area of responsibility. New York City government is about five times bigger than the size of the Metropolitan Transportation Authority in terms of staff and budget, yet the MTA has three full-time open data staff.

Regarding OTI's 2023 compliance plan, the NYC Open Data Release Tracker, and Published Data Asset Inventory, we appreciate that there is some transparency and public reporting of changes to agencies' open data plans, as required by Local Law 251 of 2017. This reporting helps the City Council and public better understand implementation issues. This testimony will cover three implementation issues: automation, dataset removal, and using FOIL to determine what public datasets should be published.

Automation is critical to the sustainability of Open Data because it ensures that new data is made available to the public automatically, regardless of staffing and scheduling issues. We urge OTI to accelerate automation and publish a schedule to automate all eligible datasets.

Reinvent Albany Analysis: Status of Dataset Automation - Eligible Datasets As of Feb 22, 2024 Published Data Asset Inventory			
Year Dataset Made Public	Total Datasets Eligible for Automation	Automated (as of 2/22/24)	Not Yet Automated (as of 2/22/24)
Unknown	93	78	15
2011	71	2	69
2012	1	0	1
2013	120	43	77
2014	33	22	11
2015	67	32	35
2016	98	58	40
2017	47	32	15
2018	150	92	58
2019	73	28	45
2020	51	26	25
2021	24	4	20
2022	22	9	13
2023	22	11	11
Grand Total	872	437	435

Unfortunately, automation of datasets is lagging. The Data Asset Inventory shows that 872 datasets currently on the NYC Open Data Portal have been flagged for automation. However, 425 of them – nearly 50% – have not yet been automated. A number of these datasets date back to as early as 2011, as shown in the table above. Unfortunately, the data asset inventory does not show the amount of datasets automated in a given year.

Dataset removal should receive greater Council scrutiny. Overall, 22% of datasets have been removed from open data plans since 2018. Reinvent Albany examined agencies with at least 20 published datasets that removed a larger than average amount of datasets from their plans or the portal itself. In some cases datasets are removed because of consolidations, or because programs are not continued.

However, another commonly listed reason was that the data is in an annual PDF report – this is precisely the type of data that should be published as open data, since it is likely statutorily mandated and regularly updated. (Note that the City Comptroller removed a number of datasets from PDFs from their plan. These were scheduled to be published in 2017, so reflect the decision of a prior administration.)

Another reason provided is that the dataset is for internal deliberation (Department of Transportation removed many datasets for this reason). One purpose of the Open Data Law is to give important insight into internal agency operations that are not currently visible, but affect the public. Below and on the next page is a table of the larger publishing agencies and their removal rates.

Reinvent Albany Analysis: Status of Dataset Removal As of Feb 20, 2024 Open Data Tracker						
Agency	**Released**	**Removed from the plan or portal**	**Scheduled for release**	**Under Review**	**Grand Total**	**Percent Removed**
Office of the Comptroller (COMPTROLLER)	12	41			53	77%
Brooklyn Borough President (BBP)	16	13			29	45%
Department of Parks and Recreation (DPR)	44	29	7		80	36%
Department of Correction (DOC)	11	15	15	1	42	36%
Department of Transportation (DOT)	89	51	4		144	35%
Department of Finance (DOF)	26	17	6		49	35%
Department of Citywide Administrative Services (DCAS)	24	13	2		39	33%

Reinvent Albany Analysis: Status of Dataset Removal As of Feb 20, 2024 Open Data Tracker						
Agency	Released	Removed from the plan or portal	Scheduled for release	Under Review	Grand Total	Percent Removed
Department of City Planning (DCP)	39	23	8		70	33%
NYC School Construction Authority (SCA)	23	10			33	30%
Department of Environmental Protection (DEP)	33	14			47	30%
Department of Consumer and Worker Protection (DCWP)	24	12	6		42	29%
Taxi and Limousine Commission (TLC)	7	10	22		39	26%

The Freedom of Information Law (FOIL) process is still not being fully used as a tool for determining new datasets to publish, despite the 2023 Open Data Plan: FOIL Metrics tracker. While some agencies appear to be using FOIL as a guide for publishing data, other agencies appear to be getting a large number of FOILs for data that is public, but not published or planned to be released, like the Department of Environmental Protection (1,452 FOILs). However, it is unclear exactly how many unique public datasets are being FOILed, but not published. We urge OTI to clarify this metrics dataset to determine exactly how many public datasets are in this category. See the listing below of agencies with 10 or more FOILs for public data that isn't currently released.

Reinvent Albany Analysis: Usage of FOIL as Guide for Open Data Release As of Sept 15, 2023 Open Data FOIL Metrics				
Agency Name	# of FOILs w/data (public dataset or not)	# of FOILs w/public datasets, not published or scheduled for release	# of FOILs w/public dataset already published or scheduled for release	# of FOILs responses resulting in data being posted voluntarily
Department of Environmental Protection (DEP)	2158	1452	706	**0**
Department of Transportation (DOT)	2940	475	2830	**0**
Department of Consumer and Worker Protection (DCWP)	212	247	71	32
Department of Buildings (DOB)	2137	117	2020	**0**
Department of Parks and Recreation (DPR)	473	39	434	0
Department of Finance (DOF)	61	23	24	0
Mayor's Office of Contract Services (MOCS)	17	17	0	0
Office of Administrative Trials and Hearings (OATH)	17	17	0	0
Police Department (NYPD)	160	13	52	0
Department of Correction (DOC)	11	10	1	0
Department of Health and Mental Hygiene (DOHMH)	41	10	31	0

Thank you for allowing me to testify today. I welcome any questions you may have.

THE COUNCIL
THE CITY OF NEW YORK

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☐ in favor ☐ in opposition

Date: 2-27-2024

(PLEASE PRINT)

Name: MARTHA NORRICK

Address: Chief ANALYTICS OFFICER, City

I represent: OF New York

Address: _____

THE COUNCIL
THE CITY OF NEW YORK

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☐ in favor ☐ in opposition

Date: 2-27-2024

(PLEASE PRINT)

Name: ZACHARY FADER

Address: OPEN DATA PROGRAM Manager,

I represent: OTI

Address: _____

THE COUNCIL
THE CITY OF NEW YORK

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☐ in favor ☐ in opposition

Date: _____

(PLEASE PRINT)

Name: Rachael Fauss

Address: 377 Broadway

I represent: Reinvent Albany

Address: _____

Please complete this card and return to the Sergeant-at-Arms

THE COUNCIL
THE CITY OF NEW YORK

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☑ in favor   ☐ in opposition

Date: 2/27

(PLEASE PRINT)

Name: NOEL HIDALGO

Address: 85 DRIGGS AVE, BKLYN

I represent: BETA NYC

Address: _____

Please complete this card and return to the Sergeant-at-Arms

THE COUNCIL
THE CITY OF NEW YORK

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☑ in favor   ☐ in opposition

Date: 2/27

(PLEASE PRINT)

Name: Christopher Leon Johnson

Address: ▮▮▮ Buffalo Ave ▮▮▮

I represent: self

Address: _____

Please complete this card and return to the Sergeant-at-Arms