**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.