DEPARTMENT OF INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS TESTIMONY BEFORE THE CITY COUNCIL COMMITTEE ON TECHNOLOGY

Oversight - Open Data Compliance.

December 10, 2021

Good morning, Chair Holden and Members of the Committee on Technology. My name is Albert Webber and I am the Director of Open Data for the Department of Information Technology and Telecommunications, also known as DoITT. Joining me today is Evan Levine, DoITT's Chief Data Science Officer. I am here today to give the Committee an update on DoITT's role in Open Data, and I'm grateful for the opportunity to discuss this important program, which is a collaborative effort between DoITT and the Mayor's Office of Data Analytics (MODA).

The Open Data Portal has come a long way in the last decade-plus, starting out as the "NYC Data Mine" in 2009 as a limited resource for programmers and developers, to its now expansive, accessible, and user-friendly iteration. Since the passage of Local Law 11 of 2012, otherwise known as the Open Data Law, all public datasets by New York City agencies have been required to be made available through a single web portal. In our partnership with MODA, we are proud to implement a strong law not only requiring the publishing of datasets, but also containing robust amendments that keep City agencies and offices accountable to providing up-to-date, accurate, and usable data.

Today, the Open Data Portal is maintained by my team at DoITT. It is currently home to over 3,400 public data assets, including datasets, tables, and maps. Of course, the effort to upkeep existing assets and publish new ones does not occur in a vacuum – it is an all-hands-on-deck, year-round process with MODA and publishing agencies and offices. Each City agency and office is responsible for identifying and making available all public data that they produce, and are also required to annually submit an inventory of public data assets that they have not yet made public, along with the dates that they intend to publish them. This inventory, along with other compliance metrics, was published on September 15th of this year as part of the Open Data for All 2021 Progress Report.

As my colleague Martha has detailed, there are approximately 100 Open Data Coordinators (ODCs) who are designated subject matter experts and our liaisons to their respective agency's or office's publishing of Open Data, as required by Local Law 251 of 2017. On the Open Data Portal, there is a dataset that lists the ODC for each agency or office. DoITT and MODA work closely with them throughout the year, providing guidance and assistance on a wide range of topics, including: compliance requirements, responses to the public, publishing timelines, identifying public datasets, and automations, to name a few. In fact, we held a virtual ODC convening yesterday. It is DoITT and MODA's shared goal to empower ODCs with as much

information and support as possible so they may make the most informed decisions about data publishing for their own agency or office.

The investment we have made in establishing a community of ODCs has, as Martha noted, improved the compliance process, and has also been integral to the City's COVID-19 response. Despite the challenges presented by the pandemic and the City's response to it, the Open Data program has continued its normal course of business, in addition to supporting the publishing of datasets to the Open Data Portal related to COVID-19 and its variants. Additionally, data has been generated as a result of a number of new programs and policies the City has implemented to handle the shift in City life as a result of the pandemic, including the Department of Transportation's Open Streets locations and maps and the Department of Parks and Recreation's Parks Closure status datasets, to name a few. We have been proud to support these agencies in publishing this data to the Open Data Portal.

DoITT has also worked to implement technical changes to make compliance and publishing easier and more reliable, including improvements to our file drop automation capability, upgrades to our databases, and a number of processing enhancements.

Lastly, I'd like to address the Technical Standards Manual (TSM), the document that details that the Citywide policies, standards, and guidelines required to support the Open Data Law. This document has not been updated since 2018, but we have been working closely with MODA to update the TSM and provide the requisite public comment period. We hope to be able to share these updates within the coming weeks.

Thank you for the opportunity to testify today. I am happy to take Council Members' questions.

<center>###</center>

Toward a Third Wave of Open Data for New York City

Written Testimony by
Stefaan G. Verhulst
Co-Founder, The GovLab,
Tandon School of Engineering, New York University

Before the New York City Council Committee on Technology

December 10, 2021

***

Chairman Holden and distinguished members of the Committee, thank you for allowing to appear virtually before you today. My name is Stefaan G. Verhulst, Co-Founder of and Chief Research and Development Officer at the Governance Lab ("The GovLab") at New York University. The GovLab is an action research center whose mission is to strengthen the ability of institutions—including but not limited to governments—and people to work more openly, collaboratively, effectively and legitimately to make better decisions and solve public problems. We provide our advice and expertise to leaders and organizations around the world, including this committee, most recently in our testimony in January on the need to acquire a social license for smart city data.

I would like to start by commending the efforts of the NYC Open Data Program, which are an example for cities around the world. We engage with cities and organizations worldwide and, just this past week, were told that the City of Guayaquil in Ecuador, one of our Open Data Policy Lab City Incubator program participants, used New York City's open data as an inspiration for

much of its work for its own open data platform. New York City is a leader in the municipal open data field. The world watches and replicates what this program does.

While the city's work is commendable, we believe there is much work that city leaders can still do to improve lives, be transparent, and support action toward those issues that New Yorkers care most about. We would encourage New York City to continue its status as an international leader in innovation by initiating and embracing a [Third Wave of Open Data](#).

The Third Wave of Open Data, a framework our Open Data Policy Lab developed in consultation with numerous open data leaders worldwide, comprises a set of actions and priorities that goes beyond the initial first wave focused on freedom of information, and the second wave that sought to establish a platform for data release. It involves an approach to open data not simply for the sake of opening data but to enable impactful and responsible re-use of data to tackle societal challenges and to provide for equity and opportunity, especially through inter-sectoral collaborations and partnerships. The Third Wave seeks to build upon the existing foundation of open data and integrate the following priorities that can increase the value of current programs:

- Publish with purpose: it seeks to understand the demand side of open data, how people will consume and use the data released, and prioritizes action toward those areas that are urgent, impactful, and legitimate.
- Initiate Data Collaboratives: it seeks to open both private and public sector data and enable partnership between a variety of actors and experts; and
- Prioritize data rights and data responsibility.

For this committee's consideration, we provide four recommendations for how the City of New York can improve its open data program to be a leader in the Third Wave and ensure that its open data program evolves at the pace necessary to meet the needs of its residents. These

recommendations are not comprehensive but, rather, represent major priorities that the City could focus on to expand the impact of its current work. These recommendations include:

A. **Invest in Data Stewardship**: [Data stewards](#) are organizational leaders or teams empowered to create public value by providing responsible access to data (and data expertise); identifying opportunities for cross-sector collaboration and responding proactively to external requests for functional access to data, insights, or expertise. They are groups or individuals active in both the public and private sector, promoting trust within and outside their organization. In its [final report](#), the European Commission's High-Level Expert Group on Business-to-Government (B2G) Data Sharing noted the need for public and private sector organizations to establish data steward roles to enable responsible, accountable data sharing for the public interest.

Consistent with this idea, the NYC Open Data Program legally established [Open Data Coordinators](#), existing agency staff members designated by each city agency to take on the additional responsibilities of (1) identifying and enabling the delivery of datasets to the open data portal, (2) develop their agency's portion of the legally-mandated annual open data compliance report (3) liaising with the open data team at the Mayor's Office of Data Analytics and Department of Information Technology and Telecommunications, (4) addressing public feedback on datasets, and (5) performing public outreach. New York's decision to establish these liaisons is a laudable first step toward establishing a culture of data stewardship in New York City government. However we recommend that investment in agency-based open data capacity be increased by making these positions full-time open data stewards at the agency level. This will allow individuals to focus their full time

and effort toward improving open data, identifying opportunities for collaboration, and engaging with the public.

B. **Democratize Methods to Surface Demand**: The NYC Open Data Program has a legally mandated public data request avenue through which the public can request that certain datasets be made public. While this is a valuable resource for those familiar with data and familiar with data the City of New York collects, it is a limited pathway to collect and identify the demand and needs that could be met via open data by other less data-literate communities.

We recommend that the NYC Open Data Program explore additional programs and processes to surface needs, questions and demands from New Yorkers, to then influence future data releases. We have found particular value in smarter crowdsourcing techniques. Our [100 Questions Initiative](#), for instance, uses practitioners who possess both relevant domain knowledge (such as on migration, gender, or disinformation) and data science expertise) to collaboratively draft important research questions and then uses public voting to prioritize which questions to focus action toward. Our [Data Assembly](#)—which we launched in partnership with New York and Brooklyn Public Library systems—assembled small broadly representative assemblies of stakeholders representing data holders, data users, and New York residents to learn from one another, deliberate on complex topics, and share their opinions in order to identify local concerns and inform decision-making as it related to data re-use for pandemic response and other crises. These three groups, mini-publics, produced insights that proved valuable in

soliciting diverse, actionable public input on data re-use for crisis response in the United States that could inform a responsible data re-use framework.

C. **Facilitate access to more non-governmental data**: The NYC Open Data Program hosts a handful of non-governmental datasets. Creating a formalized pathway for hosting non-governmental datasets is part of the 10 year strategic plan for NYC Open Data. We recommend accelerating the responsible re-use of non-governmental datasets through the NYC open data platform and data collaboratives.

As COVID-19 and the [NYC Recovery Data Partnership](#) led by the Mayor's Office of Data Analytics demonstrates, there are missed uses and missed opportunities for impact when non-governmental datasets are not made more readily available for analysis. The City can easily make data more accessible for re-use by centralizing it in a common space. We recommend creating a policy about non-governmental datasets, developing a governance structure to manage them, and creating processes to handle those requests for that data.

D. **Explore Opportunities at an Even More Localized Level**: Finally, the Third Wave of Open Data identifies the need for increased investment in opening data at the subnational level. Given that New York City houses a population the size of many countries, the function of the NYC Open Data Program, in some respects, resembles more of a national-level open data program than a subnational one.

We recommend continued exploration of open data engagements at an even more localized level in New York City—continuing existing engagements such as Open Data Week, the Open Ambassador Program, and supporting data-driven efforts by community boards. There may be additional opportunities for research and pilots in this domain as well, new models that may be worth exploring through an open data application at a borough or community-board level in New York City. We would encourage city leaders to look at some recent efforts taken around the country to promote even more localized data re-use, such as Tulsa's Urban Data Pioneers, Boston's pilot of DTPR and San Jose's Data Equity Framework.

***

To ensure New York City embraces the Third Wave, my colleagues and I encourage city leaders to pursue these four actions: investing in data stewardship; democratizing methods to surface demand; hosting more non-governmental data; and exploring local localized opportunities. We believe these actions will allow New York to continue to excel and preserve its status as an international leader in open data.

Thank you for your time. I welcome any questions you might have.

Reinvent Albany Testimony to the NYC Council Committee on Technology
For 2021 Oversight Hearing on Open Data Compliance

December 10, 2021

Good morning Chairman Holden and members of the Committee on Technology. My name is Tom Speaker, and I am a Policy Analyst 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.

Today, in the last month of the eight-year de Blasio administration and on the cusp of a new mayor and City Council entering office, Reinvent Albany's testimony will highlight eight points.

First, it is essential that the City Council continue to hold this annual oversight hearing and look for ways to continuously improve the NYC open data program. The Council has actively worked to increase the transparency, accountability and effectiveness of the City's open data efforts by passing legislative mandates.

Second, the success of NYC government open data depends on public interest and pressure. NYC leaders and career management 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 agencies. Reporting mandates like the annual compliance plan, data dashboard and agency audits are essential to compel agencies to comply with the Open Data Law.

Third, having a Mayor's Office of Data Analytics is generally a good idea and especially important to the success of Open Data. Officials in the Mayor's office have clout and access to Deputy Mayors that the Department of Information Technology and Telecommunications (DOITT) does not. We hope the Adams administration has a robust MODA.

Fourth, despite almost a decade of experience and a real investment of time and talent, NYC government has barely begun to reap the benefits of open data. By far the most expensive part of creating an agency dataset is gathering data. Once the dataset is created, publishing that data is cheap. And once the data is published, an infinite number of users can benefit from it. It is simply stunning to hear agencies complain about the time and effort needed to publish data in the Open Data Portal, since publishing data for broad use probably costs half a percent or less of what gathering and maintaining the data does. To use a paper analogy, it's like taxpayers paying an agency to create a phonebook of city services and then printing one copy and sticking it on a bookshelf in some manager's office.

Fifth, the mayoral transition between the Bloomberg and de Blasio administration almost destroyed NYC Open Data. MODA had to be rebuilt from scratch, DOITT was adrift, and many agencies defaulted to ignoring the requirements of the Open Data Law. We hope this Committee and City Council can help NYC government avoid this when Mayor Eric Adams takes over.

Sixth, MODA and DOITT are struggling to get agencies to fulfill basic aspects of the Open Data Law. According to the Open Data Dashboard, only 59% of datasets are updated on time and only a third of datasets scheduled for publication in the last 12 months have been available on time. Yes, COVID has been hard for everyone, but this performance is unacceptable and shows that almost a decade in, complying with the Open Data Law has not been institutionalized by agencies.

Seventh, we believe the key to sustainable open data is automating datasets so that agencies do not have to update data by hand. Our understanding is that DOITT has struggled to automate datasets. You should keep a close eye on dataset automation as it is the key to making open data cost-effective and sustainable.

Finally, you can think of open data as the canary in the coalmine of NYC government's information architecture. Agencies that are good at managing their data do not have much trouble publishing data to the Open Data Portal. That said, it is totally unclear to us whether high-level NYC managers and IT managers understand that a highly functioning open data system would save the city huge amounts of time and money and foster much more sophisticated and widespread data analytics. To this end, we suggest you and the Adams administration take a close look at DOITT's new nearly $200 million contract that includes a redesign of DataBridge. This could be a huge opportunity to automate the flow of agency data to the public Open Data Portal. At a minimum, it is a great chance for MODA to rationalize NYC data management.

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

The Good:
- Over the last 20 months and since the dawn of the pandemic, we have seen explosive interest in open data events. Since April 2020, we have engaged over 1,680 people at one of our open data themed events.
- This year, over an 8 day span, we supported over 150 Open Data Week events. It was a bit of a blur but we hosted daily open data trainings, morning coffee hangouts with MODA and agency staff, and helped each other process the data behind the pandemic.
- Next year, we are excited to co host Open Data Week 2022 and celebrate the 10th anniversary of the City's Open Data law. If you have an idea for an event go to www.open-data.nyc
- Next year, we are excited to rekindle our open data ambassadors program with MODA and METRO. This program recruits government employees, public library staff, and every day New Yorkers to learn how to represent Open Data to all New Yorkers. Through an intensive training, we help you figure out how to speak open data to everyday situations. The last time we ran this program, we partnered with the Queens Public Library and were able to provide open data trainings in every single community district.
- Ten years in, we can see that the current that propels open data is getting stronger.

What have we learned in 10 years?
- The scaffolding to support the open data program is well thought out but it needs dedicated technical resources.
- Agencies continue to share open data based upon how they record data. While this is what we wanted when we started, this is not what will make open data useful for the next decade.
- We are still waiting for agencies to properly resource data analytics teams and dog food their own data.

Before I run out of time with outlining the challenges, it is clear that MODA needs more technical resources to pull data from agencies, then to help agencies use the data. They should not have to share these resources with DOITT. We know that using data identifies inconsistencies.

For the next decade of open data, we need the following - better metadata; we need agencies to use their own data from the open data portal, this is how they will clean their datasets. MODA needs the resources to support its own technical needs and support other agencies. We need an updated Technical Standards Manual. We need to improve our auditing and examinations workflow. Lastly, we need transparency from DOITT — we need to know what services they are purchasing, how they are supporting MODA, how they envision digital transformation for City Agencies, and how DOITT is including users' voices / needs into that transformation.

The current challenges:
- **Technical Standards Manual (TSM)** — As part of extending the open data law's mandate, the Mayor and Council should be receiving an updated copy every two years and DOITT should have a publicly accessible website for the public to comment on.
    - The only TSM we can find is from 2018 and it is a PDF.
    - There is an older version of the TSM as a website with the ability to make public comments.
    - Additionally, the TSM was once on GitHub so the public could track changes. Not only is the TSM gone from GitHub, so are all of DoITT's open source repositories. Not only is this the opposite of how Governments are operating, this makes me question DOITT's leadership.
- **Inconsistent Automation** — DOITT has not dedicated enough resources to automate and keep data sets up to date. While we see they have put resources into supporting the open data program, they have not put resources into making open data operational. From our point of view, there is a deficit in technology resources.

- There is no comprehensive cloud strategy. We're looking at one Covid related emergency services contract for nearly $200 million that includes support for the construction of a data bridge but have yet to hear about any success. $6 million alone is dedicated to redesigning Data Bridge. There needs to be clear communication from DOITT on how the City's IT infrastructure will be modernized.
        - Moving forward, we're starting to wonder if there should be some additional metadata on datasets saying how they get to the open data portal. Example, where is the data coming from? Is it coming from databridge or directly from the agency?
- **Not enough oversight** — MODA and DOITT don't have the staffing to oversee every agency's implementation of open data. Not only does the Council need to hold the Mayor accountable. We'd like to see a renewed law around auditing and examinations. We know that this was time consuming for the MODA team and think it is critical that MODA gets additional resources to perform data audits and automations. The Mayor's office and Agencies need to be held directly accountable.
- **Timeliness of data release** — We are super thankful for MODA's open data dashboard. It lays bare the issues around data updates; it also outlines how things have improved. From our understanding, this dashboard has helped Open Data Coordinators (ODCs) validate frequency, clean up metadata, and improve ODCs ability to keep datasets up to date. Yet, we can see there are major data quality issues.
        - Right now, 40.72% of datasets are not updated on a timely basis.
        - There is no clear understanding of when datasets get updated versus when they say they should be updated. This speaks to an issue around better defining a dataset's metadata. Also, this speaks to a lack of operationalizing open data. If there were better practices of open data being used daily, data quality issues would be more apparent. (This gets back to better metadata > oversight > resources > operationalizing the data.)
- **Data Quality issues** — The City's open data dashboard makes it clear that many agencies need to weed their garden.
        - Agencies like the DOE need to clean up their existing datasets, they have over a thousand datasets far exceeding any other agency by 900 datasets.
        - Nearly a ⅓ of agencies either don't update their data in a timely manner or haven't cared to fill that data in.
        - 61 of 97 agencies have "null" as a response to datasets "Percent of planned releases released on time within the last 12 months"; at the same time, every single one of these agencies say they have zero overdue datasets.
        - Entities like NYC 311 need to clean their legacy data, there are many little data quality issues with addresses that need to be fixed.
        - We are very supportive of MODA developing an analytic tool to help increase metadata quality, but this will take time to have every agency clean up that data.