Thank you, Dr. Mashariki, and good afternoon Chairman Vacca and members of the City Council Committee on Technology. My name is Albert Webber and I am pleased to speak with you today about many of the improvements made to the NYC Open Data Portal since we last testified.

To date we have unlocked nearly 1,400 data sets and maps via the NYC Open Data Portal, sourced from more than 80 New York City agencies and entities. As part of the annual July 15 update to the NYC Open Data Plan, City agencies identified an additional 282 data sets to be released through December of 2018. We also work closely with the Mayor's Office of Data Analytics (MoDA) and agencies to continue identifying data to be released or listed in the Open Data Plan beyond the yearly July update, and we make monthly updates to the Plan reflecting those updates.

Data Integrity

To date we have automated the refreshing of approximately 120 data sets, to ensure their timely and accurate delivery. Many of the 120 data sets automated were in high public demand, including 311 Service Requests, Police Department Motor Vehicle Collisions, and Restaurant Inspections from the Department of Health & Mental Hygiene.

This year alone 18 new data sets have been automated, including Derelict Vehicles from the Department of Sanitation (DSNY); For-Hire Vehicle and Active Drivers from the Taxi & Limousine Commission; and the City Record Online, Green Book, and City Store from the Department of Citywide Administrative Services. Automation – aside from the initial work done to implement the process – means that no human intervention is necessary for a data set to be refreshed.

Additionally – and particularly topical given one piece of introduced legislation – we recently completed a proof-of-concept that should allow us to begin the automated delivery of geospatial data sets to the NYC Open Data Portal.

With regard to the 1,200-plus data sets that are not automated, we work closely with agencies and their Open Data Coordinators to manually refresh them. This process is often time-consuming and labor intensive, but our team diligently works to ensure timely data.

We are also undergoing an effort to consolidate data sets where possible. There are a number of data sets broken down by year, borough and other attributes, making it difficult for users to export a holistic view of the data they desire. By consolidating data sets we can allow users to more easily locate and download data, improving the overall user experience. This consolidation effort may decrease our data set totals, but in keeping with the themes of *Open Data for All* that Dr. Mashariki just discussed, this approach will in the long run facilitate fuller access to data sets and maps for our many users.

New additions

As we continue our efforts to make data sets more usable, we also continue the march toward opening new data. Since our last update, the following new data sets have been made available:

- <u>Universal Pre-K Data</u> – Locations of facilities, number of seats available, website, and contact information.
- <u>Workforce 1 Job Listings and Recruitment Events</u> – Job openings and recruitment events.
- <u>Civil List Data</u> – Salary information for all New York City employees.
- <u>TLC Trip Data</u> – Information on the millions of trips taken by New York City's yellow taxis on an annual basis.

Data sets scheduled to be made available by the end of the year include:

- <u>Bridge Ratings</u> – Bridges rated good or very good by calendar year.
- <u>Reportable Disease Surveillance Data</u> – Information on cases of reportable diseases (not including STDs, Tuberculosis or HIV) in New York City.
- <u>Events and Cleanup Costs</u> – Names, locations, and dates of events permitted by the Street Activity Permit Office and the DSNY cost of clean-up efforts billed to event organizers.
- <u>School Budget Overviews</u> – Showing individual school budget allocation by year.

Additionally, earlier this summer we partnered with our open data platform provider to launch a new feature called "Data Lens." Data Lens is a machine-learning enabled tool that allows for the creation of auto generated charts, graphs, and search features so that users, be they data scientists or everyday New Yorkers, can easily view and interpret data sets. In short this tool helps to make the viewing of large data sets less intimidating, and more accessible to a broader audience.

These ongoing efforts encapsulate the spirit of the *Open Data for All* initiative Dr. Masharki discussed earlier, and reflect the de Blasio Administration's commitment to keeping New York City at the vanguard of the open data movement. And as we continue this work, the continued support of and feedback from the city's vibrant civic technology community has proven invaluable – as has the support, led by Chairman Vacca and the Technology Committee, of the City Council.

A number of the bills introduced last month align very closely with the aims of *Open Data for All* – making open data more usable, more accessible, to New Yorkers of all stripes and in every corner of the city. Some of these bills would mandate efforts already underway, while others fit squarely with the de Blasio Administration's five borough strategy to enhance the open data initiative.

Taking a brief look at each in turn:

Intro. 914 would establish response timelines for public requests on the NYC Open Data Portal, and help set user expectations about when the data they have requested can be made available within the context of the City's larger Open Data Plan. We wholeheartedly support the intent of Intro. 914 and have no significant concerns about this proposal. As you know, as part of the annual update last July we committed to updating the plan on a monthly basis, and regular feedback on requests for new data sets fits into that increased engagement strategy.

Intros. 890, 898, and 915 embody proposals that speak directly to what *Open Data for All* is all about, and we look forward to working with the Council on each to explore ways that our shared goals may be realized.

Intro. 890 relates to the retention of archival data on the NYC Open Data Portal, and rightly references the Technical Standards Manual as a place for the appropriate language to reside. To the extent that old data sets are removed from the Portal without explanation this proposal is in keeping with *Open Data for All*.

But mandating that old data sets, replaced by those from the same period that are more accurate and complete, have to remain on the Portal indefinitely, is problematic. This can be confusing to visitors and runs counter to making City data more understandable and usable for all. There will also be some costs for additional storage capacity as data sets – like some of those we mentioned earlier, containing millions of rows of data – become increasing large and complex, but we look forward to discussing with the Council a path forward on data set retention.

Intro. 898 requires every data set on the NYC Open Data Portal be accompanied by a plain language data dictionary and is aimed at exactly the right place: Making what can be very bureaucratic data more usable and understandable to the layperson. Data dictionaries and supporting data set documentation were at one point common on the Open Data Portal, and can be a valuable standard to more formally implement.

Of course, the role of providing headings, explaining columns, and defining the technical terms and acronyms included in data sets needs to be completed by the agencies submitting their data to DoITT and MoDA for posting and/or automation (not by DoITT or MoDA directly), and we therefore agree with its inclusion as a requirement that can be made more specific as part of the upcoming Technical Standards Manual refresh. We look forward to discussing with the Council the most effective means of implementing data dictionaries.

Intro. 915 requires that public data sets updated on other City agency websites be posted to the NYC Open Data Portal within three days. This would help ensure continuity and consistency of user experience when searching for official New York City data, which is directly in line with both the original intent of the City's Open Data Law, as well as with *Open Data for All*. While DoITT is constantly working with agencies to automate data sets, there still remains work to be done. There will be some cost and resource considerations to be addressed in pursuing this effort but we share with the Council the goal of keeping City data accurate, updated, and consistent across all channels – and we look forward to working with you on making that vision a reality.

The next proposal, **Intro. 908**, establishes a new process for making agency FOIL responses publicly accessible. The spirit of this proposal squarely aligns with the de Blasio Administration's commitment to the promise of transparent, free, and accessible open data. As part of that promise, we are committed to ensuring that agency responses to FOIL requests are widely accessible to the public.

To that end, the Department of Records and Information Services is currently developing an OpenFOIL Portal that will launch by the end of this year. This centralized website will empower the public to easily submit and track FOIL requests for all City agencies, as well as access the vast majority of records responsive to previously-filed FOIL requests, even where there is no accompanying "data set."

Such proactive disclosure through a centralized repository will dramatically expand public access to government records. Therefore, rather than duplicating work to be completed in the coming months, we are happy to explore crafting a technical solution whereby the records that will be posted to OpenFOIL are automatically transferred to the NYC Open Data Portal.

The final two proposals, Intros. 900 and 916, have elements with which we agree but also present significant staffing, cost, timing, or legal concerns.

Intro. 900 requires every data set on the NYC Open Data Portal that contains address information to utilize a standard field layout and presentation of that address information, including corresponding geospatial reference data.

Having a single City standard for address and geospatial information data is ideal, but we do not believe it is a goal best achieved through open data legislation. While Open Data would benefit from such a standard, the standard itself should be developed via a larger collaborative process whereby agencies, advocates, and users can work toward a generally agreed-upon, adaptable standard for address and geospatial data. Once established, the systems utilized by agency databases feeding the NYC Open Data Portal would adhere to that standard when producing data for publication. Older legacy systems, once replaced, would then also be required to incorporate those standards. This is a worthwhile effort, but one that will require additional resources and time – substantially more so than the 90 days proposed – to implement effectively.

In the interim, we continue to make all the relevant data – tabular, shape files, etc. – available on the Portal itself for interested individuals to use in geocoding the data sets they are most interested in. This indeed is one of the earliest, truest promises of open data: Provide it to the public and let them do with it what they will, without further interference.

Intro. 916 would require the Commissioner of Investigation to conduct audits of certain agencies' compliance with the Open Data Law. While we – and all public officials – can always do a better job of serving New Yorkers, the City's Open Data Law is among the most ambitious, comprehensive policies of its kind in the world, and has set a model for many governments to follow. We are happy to work on its continued implementation and committed to doing even better still. There are a number of steps between making the accountability improvements Dr. Mashariki mentioned (for example, more clearly stating on the Portal which data sets are either removed or delayed, with detailed explanations) and requiring the Commissioner of Investigation to conduct audits in an area where the City has demonstrated such noteworthy, transformational success in just over three years. Most significantly, we are concerned that the proposed audits would require the Commissioner of Investigation to inspect *all* of an agency's data, including even sensitive, confidential, or privileged information, in order to determine the existence of public data sets. We have legal concerns about the Commissioner's authority to access such information, as well as concerns about the utility of such audits generally in advancing Open Data goals.

Thank you for the opportunity to testify this afternoon. Open Data has been and remains a priority of the de Blasio Administration and we thank our partners in the City Council, and across the civic tech and advocacy communities, for their continued efforts in making New York City's successful implementation even better.

This concludes our prepared testimony and we look forward to answering your questions.

Thank you.