CITY COUNCIL
CITY OF NEW YORK

---------- X

TRANSCRIPT OF THE MINUTES

Of the

COMMITTEE ON TECHNOLOGY

---------- X

October 18, 2018
Start: 1:37 p.m.
Recess: 3:01 p.m.

HELD AT: 250 Broadway – Committee Rm.
14th Fl.

B E F O R E: PETER A. KOO
Chairperson

COUNCIL MEMBERS: Robert F. Holden
Brad S. Lander
Eric A. Ulrich
Kalman Yeger

A P P E A R A N C E S (CONTINUED)

Kelly Jin, Chief Analytics Officer, City of New York and Director of the Mayor's Office on Data Analytics, MODA

Adrienne Schmoeker, Director of Civic Engagement and Strategy, Mayor's Office on Data Analytics, MODA

Albert Webber, Director of Open Data, Department of Information Technology and Telecommunications, DOITT

Donald Sunderland, Chief Data Officer and Deputy Commissioner for Data Management and Integration Department of Information Technology and Telecommunications, DOITT

Daniel Allen, Technology Policy Analyst: Appearing for Gale Brewer, Manhattan Borough President

Lindsay Poirier, Lab Manager, BetaNYC

Alex Camarda, Senior Policy Advisor, Reinvent Albany

[sound check] [pause] [gavel]

CHAIRPERSON KOO: [off mic] Good afternoon. I am Peter Koo [on mic] the Chair of the Committee on Technology. I want to welcome all of you to the Technology Committee hearing today in the City Council. We are here today to conduct and oversight hearing on the Open Data Law. The 2018 Open Data Plan and the several amendments to the Open Data Law that were passed by the committee within the past few years. New York City is widely recognized as the leader in the Open Data Movement among local governments. This success should be credited to both toward the underlying law itself as well as the delegation for its implement—implementation and hard work that has been demonstrated by the Department of Information Technology and Telecommunications, and the Mayor's Office on Data Analytics. Today, the committee will discuss the implementation of Open Data Laws, the challenges that have arisen and ongoing issues and the ways we can work together to solve them both administratively and legislatively. In addition, we will hear Intro No. 1137, sponsored by Council Member Adams codifying the Mayor's Office on Data Analytics also know as MODA. I look forward

to hearing from the panels today, and I would like to thank the Technology Committee staff and our data team for putting together this hearing. With that said, I would like to recognize the Technology Committee members and Council Members. We have Council Member Adams, and the other ones are on their, and our fist panel is Albert Weber from DOITT and our newly arrived MODA—MODA—MODA chief, Kelly Jin, right? Yeah. Thank you, yes.

LEGAL COUNSEL: And--

CHAIRPERSON KOO: One second, one second please.

LEGAL COUNSEL: One moment, please. I just need to swear you in.

CHAIRPERSON KOO: Yes.

LEGAL COUNSEL: I want to ask you to raise your right hand. Do you swear to tell the truth, the whole truth and nothing but the truth in your testimony before the Committee, and to respond honestly to Council Member questions?

KELLY JIN: Yes.

LEGAL COUNSEL: Thank you. You can start. [pause]

KELLY JIN: Good afternoon, Chairman Koo and members of the Committee Technology. My name is Kelly Jin, and I serve as the Chief Analytics Officer here in the City of New York as well as the Director of the Mayor's Office on Data Analytics. I'm joined today by Adrienne Schmoeker, MODA's Director of Civic Engagement and Strategy as well as Albert Webber, Director of Open Data at the Department of Information Technology and Telecommunications, DOITT as well as Donald Sunderland, DOITT's Chief Data Officer and Deputy Commissioner for Data Management and Integration. WE are here to discuss MODA and DOITT's work to facilitated citywide compliance with the city's Open Data Law, and achieve our vision for Open Data for All. Monday, October 15th was my first day as the City's new Chief Analytics Officer. I'm thrilled to step into this role, and work alongside our agency and outside partners, many of whom are represented here today to continue the great work happening within the city regard data analytics and open government. To provide a brief biography about myself, most recently I served a director focused on data driver investments at the local government level at a national philanthropy as a Policy Advisor to the

U.S. Chief Technology Officer and Chief Data Scientist at the Obama White House, and prior to that as a founder and co-lead of the City of Boston Data Analytics Team. While I have just started, I am preceded by the excellent work of the New York City Open Data Team, city agencies as well as the Council whose efforts have really made New York City's Open Data Program one of the best in the world. On behalf of the Administration, I would really like to extend gratitude to this committee for its ongoing and continued support of this important program. I would like to first begin by describing the structure of the city's Open Data Initiative. MODA is the business owner of the program, and our mission is really to make city data more accessible, and actionable through public data, interagency data sharing, and advanced operational analytics. This work would not be possible without open data. Last month we published the 2018 Open Data Plan and the Annual Progress Report on Open Data for All conveniently titled and also here the New York City Data at Work, copies of which are here in the room, and also available to committee members as well.

CHAIRPERSON KOO: [off mic] I'll give you one.

KELLY JIN: Oh, hello. [background comments, pause] And this report really describes how open data powers government efficiency, and effectiveness at MODA as well as across city agencies. The Data stories contained within, highlight the many ways in which public data can and has improved outcomes for New Yorker from prioritizing where inspectors root out tenant harassment with advanced analytics to coordinate a more efficient response from emergency service agencies to enabling Minority and Women Owned Businesses to have greater opportunities for city contracts. Our partners at DOITT are the technical manager of the program connecting the vital work of data publishing with city agencies, developing data set automations and maintaining the datasets digitally. Each week the Open Data Portal is visit— visited by over 30,000 users including students, researchers, entrepreneurs, non-profit employees who really use the data to conduct meaningful analysis, and inform unique projects. These projects include the Open Sewer Atlas, a digital resource that pulls

from data provided by the Department of Environmental Protection by DOITT and 311 as well to inform the work of water advocacy groups across the city. The development of community resources such as the Sewer Atlas would not be possible were it not for the hard work of the DOITT Open Data team. The team ensures that the continued quality of the open data inventory, and important information service for New Yorker and those who serve New Yorkers. Finally, the program would not be possible without the contributions of Open Data Coordinators also known as ODCs. They are who MODA and DOITT's agency—they are MODA and DOITT's agency level liaisons. Open Data Coordinators are appointed by the head of their agency, and are responsible for identifying eligible datasets across agency divisions enabling the delivery of open datasets to the portal and addressing public feedback on their agency's datasets. Thanks to Local Law 251 from 2017, which this committee passed last year, every agency is required to have an Open Data Coordinator. The success of open data really relies on the strength of the cohort of Open Data Coordinators. To that end, MODA and DOITT have made significant progress setting

up ODCs for success. This past year we have trained all ODCs on how to be more effective in the position, and a mandatory day long workshop hosted in partnership with Socrata, with Beta NYC, who I believe is represented here today, Laurenellen McCann of Build With, the Sunlight Foundation and the Department of Citywide Administrative Services. Also, convened ODC's in-person to describe—to discuss requirements and share opportunities and lastly, developed education al resources to assist ODCs in creating compliance and engagement strategies tailored to the unique data environments in their own agencies. We know that these efforts to build capacity are working because we recently surveyed Open Data Coordinators for feedback on this year's annual compliance reporting process who largely responded that they felt confident in their abilities to complete this year's compliance reporting requirements. MODA, DOITT and the ODCs have collaborated to add more than 2,100 datasets representing billions of rows of data to the Open Data Portal. The NYC Data at Work Report describes how several of these datasets are created and used. The largest datasets on the portal are taxi trip

record records, which account for hundreds of millions of rows of data annually, and are used by the Taxi and Limousine Commission to create better proactive policy at the agency. Among the most accessed datasets on the platform are the Department of Buildings' dataset on job applications and permits, which have helped the agency improve how it processes information, and the DCSA' Civil Service dataset, which is used by thousands of prospective civil servants to check the results of their Civil Service examinations. This year 38 agencies have published more than 600 datasets ranging from street furniture location to records on licensed dogs to information on housing developments that receive financial assistance. In addition, more than 400 new datasets were identified datasets were identified for future release. Still, we do not measure success solely by the number of datasets that we publish. Usable data is well documented data. Metadata is information that describes how data was collected and what each column in each dataset represents. As such, metadata is the key to making dataset understandable to every New Yorker, which is the spirit of the Open Data for All. More than 90% of

datasets on the Open Data Portal have data dictionaries and MODA recently completed our metadata for all initiative in partnership with the local Library of Science community to develop as a new standard and guide for creating best in class metadata. Starting in January, we will see improved metadata for the most used datasets on NYC Open Data, and all metadata for new published datasets will be required to meet this standard. Since the announcement of Open Data for All in July 2015, the Administration has been unrelenting in its efforts to put data in the hands of more New Yorkers. As noted in the 2014 NYC Open Data is an invitation for any one, any time anywhere to engage with New York City. This past year we've engaged a record number of users both online and in person. I would like to take this opportunity to outline a few use cases of open data and specifically highlight the Open Data for All Initiative. First, Open Data Week 2018 showcased the value of open data as a community building and problem solving resource. This past March, 1,800 New Yorkers attended events during the second annual Open Data Week. More than 51 government, academic, cultural, and business partners produced over 30

events including a data art exhibition, an open contracting treasure hunt and a tour of a data exhibit at the Museum of the City of New York. I wish I had gone to this. Students in Staten Island use data on school statistics to express the information about youth in foster care. Entrepreneurs learn how to use open data to build a business, technologists and designers collaborated on creative solutions for the L Train Shutdown, which led to a current proposal for an L Train Co-lab, a program and a place to focus on data driven problem solving of issues related to the L Train and sustainability impact as a result of the shutdown. Open Data Week showcased the breadth and the depth of open data users as well as those use cases. We expect to see even more unique ways public data is being used in Open Data Week 2019, which will be produced in close partnership with BetaNYC and expected to take place in March. We will be launching a call for event organizers in early November, and we welcome the Committee on Technology's participation in Open Data Week 2018—2019, and we look forward to exploring ways to engage New Yorkers in your districts around open data.

Additionally, we are pushing open data to be a tool to teach New Yorkers how their local government works. Three *Learn About New York City* events welcomed a combined 350 plus attendees to City Hall to how city officials from nine different agencies describe the operations behind the public data. The most recent event held in August placed the spotlight on transportation data from, the City and Limousine— the Taxi and Limousine Commission, the Department of Transportation Citywide Mobility Survey, the Department of Citywide Administrative Services Fleet Management Operations, and the Department of Sanitation PlowNYC program. We shared some of these stories and more in the Open Data for All progress report, which was accompanied by a video on the open data home page show the public data behind city operations that New Yorkers encounter every day. Dozens of examples of the way New Yorkers use open data can further be found on the open data website's new Project Gallery or which stories can be sourced— were sourced from a contest held earlier this year. Winners were featured in our second marketing campaign using LinkNYC kiosks. Finally, agencies are the key to scaling Open Data for All into a citywide

data awareness effort. For the first time this year we required agencies to go beyond publishing data and commit to engaging their communities with their datasets. Commitments include advertising and agencies' open datasets on its website and social media channels producing curricula for using its data, speaking about open data at public events or in schools, and writing blog posts. MODA is developing a tracker to help make sure that all 200 plus public commitments can be met by city agencies in the coming year. Local Law 11 of 2012, the original Open Data Law, sunsets this year. Thanks to legislation the Council passed last year, the program will continue into the future, and will be more transparent than ever. Thanks to Local Laws 251, we published a comprehensive inventory of all public datasets and the status of their compliance with relevant Local Laws. Additionally, a new compliance dashboard will be added to the Open Data website by the end of the year. We are also identifying additional ways to make governance of the Open Data Program more transparent and participatory. We invited high school students to participate in an Open Data Youth Leadership Council to generate ideas for bringing

public data into their communities and schools and MODA continues to recruit youth to its Leadership Council and we invite Council Members to please share this opportunity with your constituents. Before I close, I would like to address Introduction 1137, which would codify MODA into the New York City Charter. MODA was founded by an executive order in 2013, and has been a leader in civic analytics in the five years since. We are excited by this opportunity to formalize MODA's role in the Charter and are eager to discuss ways in which the bill's language can better reflect the current practices with the Council after this hearing. My colleague Albert Webber will highlight a few of those ways in which DOITT already works with MODA to achieve some of those specific responsibilities identified in the bill. I would like to thank the Committee on Technology for the opportunity to testify today. I would also like to thank the Mayor's Fund, New York City's Library systems, the Metro Library Council, Civic Hall, BetaNYC, Reinvent Albany as our city agency partners, open data users and all of our many community collaborators whose support truly makes this work possible. At this point, I will turn it over to

Albert Webber who will discuss our progress on data publishing and achieving compliance with the Open Data Law in more detail.

CHAIRPERSON KOO: Before we do that, I want to announce we have Council Member Yeger and Council Member Holden arrive. Yeah, thank you. [background comments, pause]

LEGAL COUNSEL: Do you affirm to tell the—to tell the truth, the whole truth and nothing but the truth in your testimony before the Committee today, and answer honestly to the Council Member questions?

ALBERT WEBBER: Yes, I do.

LEGAL COUNSEL: Thank you.

ALBERT WEBBER: Good afternoon, Chair Koo, and members of the City Council Committee on Technology. My name is Albert Webber, and I'm the Director of Open Data for the Department of Information Technology and Telecommunications also know as DOITT. With me are Donald Sunderland, DOITT's Chief Data Officer and Deputy Commissioner for Data Management and integration; Kelly Jin, Chief Analytics Officer for the City of New York and Director of the Mayor's Office on Data Analytics, and

Adrienne Schmoeker, MODA's Director of Civic Engagement and Strategy. As Chief Analytics Officer Jin stated, DOITT and MODA collaborate closely to ensure the city's compliance with Local Laws 11 of 2012, also known as the Open Data Law. Over the past several years, it has been a pleasure to work with the New York City Council Committee on Technology and advocates alike to make the Open Data Law even more effective and impactful. We believe New York City's Open Data Law is one of the strongest, if not the strongest law of its kind in the country. We should all be extremely proud of all the progress we have made together. Chair Koo, we look forward to continuing this work under your leadership, and to hosting you for an open data training in your district next week. For those who are unfamiliar, the Open Data Law requires that all public datasets maintained by New York City agencies are made available through a single web portal. The Open Data Portal powered by our vendor Socrata, is maintained by my team at DOITT. Just last year we revamped the open data website to be as user-friendly as possible. Today the Open Data Portal is home to over 2,100 datasets including datasets, tables and maps. City

agencies and officers and responsible for identifying and making available all public data that their agency products, and are also required to annually submit an inventory of public data assets that they have not yet made public along with the dates they intend to publish them. This inventory along with other compliance metrics was published on September 14th of this year as a part of the Open Data for All Report. This is a pivotal year for the Open Data Law as the original law required agencies to publish their public data by December 31, 2018, which is in a few short months. This horizon was a great goal to move toward, but we believe a good deal of work remains. As CAO Jin stated in her testimony, the open data team has been working hard to continue to meaningfully engage open data coordinators throughout the year to deeply engrain open data practices throughout agencies. New data is created each day meaning that agencies are continuing to identify new data assets. In fact, on top of the 625 new datasets published within the last reporting period, agencies and officers identified 419 new datasets to be planned for future release. This is also the first reporting period since the passage of Local Law 251

of 2017, which in part, prescribed the disclosure of the status of all datasets including the scheduled publication date, the actual data publication, and the location of the dataset, whether a dataset is automated and, if not, if it can be and other compliance revisions. To this end, we have published a data asset inventory and an open data plan tracker. Compiling this information in one place has given the Open Data team and the public a better view of agency compliance with data standards, timely publication and updates to datasets. Over the past year we have also dedicated ourselves to efforts that ensure that the 2,100 plus datasets are high quality and up to date. We archived or consolidated 97 datasets in order to improve how users search for and find the data that is useful for them without removing important historical data in the process. We have also taken extra measures to evaluate datasets for automation. Automation is important because it helps agencies quickly and frequently update dynamic datasets. The 311 service requests dataset is an example of an automated high value dataset. Because new data is being created every day, this dataset is automated daily without any specific action by 311 or

DOITT. Just this year we automated 38 datasets bringing our total to approximately 250 automations. Furthermore, we found that 302 additional datasets are potential candidates for automation in the future. We will continue to work with agencies to identify feasible automations. We have also been steadily increasing compliance with other important provision of the Open Data Law. 2,000 datasets have data dictionaries, which help explain to users what column and rules represent. This supplementary document provides context that would otherwise not be apparent within the dataset. Additionally, 296 eligible datasets covered by geospatial requirements of Local Law 108 of 2015 have been geo coded. In other words, datasets that contain addresses must also contain specific standards fields such as latitude, longitude and Council District among others. It is our goal to get to 100% compliance on both of these requirements, and we have been actively working with agencies to reach that goal as new data sets are added. It is just as important to make sure that the data is understandable and usable as it is to simply disclose it. Before concluding, I'd like to briefly address Introduction 1137, which would

codify MODA in the New York City Charter. DOITT and MODA work closely together on the administration and implementation of the open data program, and we have found this to be an efficient and collaborative relationship that we will continue. Thus, we believe that Council Member Adams' bill is a laudable effort. However, there are some changes to the language we would want on with the sponsor and the committee. First, Local Law 11 of 2012 gives certain responsibilities to DOITT in relation to the implementation of open data, and we want to ensure that that language, which is in the bill enshrines the practices we have adopted to fulfill that mandate. We also want to stress the importance of DOITT's responsibility for the city's technology assets. We are the entity responsible for building and maintaining the infrastructure for citywide data sharing. In fact, under Commissioner Saini's leadership, we are in the process of enhancing citywide data sharing offerings. Although MODA is the business owner of the technology, DOITT remains the technology owner and service provider. This dynamic is akin to many other technology services we provide to agencies across the city. The success of

many technology services including the Open Data Portal has relied on these separate but collaborative roles. We look forward to working with the Council to make the legislation reflect the current practices that make our partnership so successful. Thank you for the opportunity to testify today. Open data remains a priority of this administration shining a bright light on our government and our city for all New Yorkers to see, and providing the tools to solve civic issues in creative ways. We thank our partners in the city Council and in the civic technology community for their continued advocacy. This concludes our prepared testimony and we look forward to answering your questions.

CHAIRPERSON KOO: Thank you for your testimony. Before we go to questions, I would like to invite Council Member Adams to give her statement.

COUNCIL MEMBER ADAMS: Thank you very much Chair Koo. Good afternoon, Chairperson Koo and Members of the Technology Committee and thank you for today’s hearing on my bill Introduction 1137. This bill would indeed codify into the New York City Charter the Mayor's Office on Data Analytics or MODA. The Mayor's Office on Data Analytics is New York

City's civic intelligence center allowing the city to aggregate and analyze data from across New York City agencies to more effectively address crime, public safety and quality of life issues. The office uses analytic tools to prioritize risk more strategically, deliver services more efficiently, enforce laws more effectively and increase transparency. The office's core functions include collaboration with the city agencies to implement data driven solutions to city service delivery issues; building a citywide data platform to facilitate data sharing; oversight of citywide data projects and implementation of the city's open law--Open Data Law. The objective information received from this office is a valuable tool for the New York City Council, and it helps us to be more robust and effective in our work. While the Mayor's Office on Data Analytics was created by Executive Order 306 under Mayor Bloomberg, we must ensure that this office survives successive mayoral administrations. I look forward to working with MODA and DOITT to fortify the bill's language and to make it as effective for all as possible. I thank you, Chairman Koo. I thank the Committee for your time

today, and I ask for your support of this important and very necessary legislation. Thank you.

CHAIRPERSON KOO: Thank you, Council Member—Council Member Adams. So, thank you all for your testimony. The committee together with our data team have reviewed the report and have several questions. First of all, the reports shows the hard work and dedication of the open—of the city's Open Data Team, and efforts toward the increasing accessibility and government—governmental transparency. We need to discuss what should be done to improve it. So, my first question is: Does the department authorize any public events or outreach to educate general public about the open data. In addition to the ones you mentioned in your testimony.

ADRIENNE SCHMOEKER: Hi, my name is Adrienne Schmoeker. I am the Director of Civil Engagement and Strategy for the Mayor's Office on Data Analytics, and in response to your question, the answer is yes. Open Data for All for us means making sure that the data is reliable, which is a lot of the work that Albert and his team over at DOITT are ensuring through the continued growth of the Automations Pipeline working with the public to

answer public inquiries that come in. We get about 25 to 30 inquiries per week to the platform. Making sure that open data is accessible is really important. There was mention of the Metadata for All Initiative. As a part of that, we had four community workshops of the summer in collaboration with local library groups in every borough. We also have the new video that was mentioned in the testimony, various tutorials and additional content has been added to the website making sure that the open data website is user friendly, ADA compliant are all things that have been a priority for us. We did user testing of the website again this summer to make sur that we're continuing to make sure that it's friendly for our users. And then lastly, making sure that the platform and the data are discoverable is key. If nobody knows that open data exists, people aren't going to use it. So, increasing marketing efforts. We had two marketing campaigns for the first time this year on LinkNYC kiosks. The first one was estimated to reach close to three million New Yorkers and we're still waiting for statistics on the last campaign, and then Open Data Week and the Learn About New York City events were mentioned in the testimony,

but I'm happy to provide more detail about those if you'd like.

CHAIRPERSON KOO: So, how do you advertise these events? How many people know about it?

ADRIENNE SCHMOEKER: That's a great question. So, on the open data website, on the contract us page there's an opportunity for people to sign up for a mailing list. So, we want to make sure people are opting into our communications and giving us their information of their own free will. So, we have a list serve of more than a 1,000 people at this point who have committed to hearing about different open data events, and so we send that information update, you know.

CHAIRPERSON KOO: Yeah, yeah. So, pursuant to the state law or to the—to Local Law 11 of 2012, the department should implement an online forum to solicit feedback from the public and encourage public discussion on open data sets availability onto that first to contact us option on the open data page. What are other avenues that the public can take in order to request data sets or leave the feedback? Do you understand my question?

ADRIENNE SCHMOEKER: I believe so.

CHAIRPERSON KOO: Yes.

ADRIENNE SCHMOEKER: So, the—the Contact Us page is a new implementation somewhat. It's been around for about 18 months. We decided to implement that after understanding that there was confusion from the public in understanding how to get in touch with the Open Data team. So, we took that feedback to heart. We developed this central place where the public could get in touch with the Open Data Team to request datasets as you mentioned, but also to report data errors, to submit data questions, to submit ideas for partnership. It's really our way of making sure that we're getting inquires in a central place from the public so that we can be accountable to following up with those inquiries after we pass them along to agencies, and as mentioned, we get about 25 to 30 inquiries on average per week.

CHAIRPERSON KOO: Okay. [background comments, pause] And during the 2015 hearing on open data, you noted that over the next year we will be implementing a new technology solution to take feedback through a centralized mechanism that allows for quicker responses, and better tracking on our

interactions with users. So, what are you—where are you in this process?

ADRIENNE SCHMOEKER: So, yes so--

CHAIRPERSON KOO: [interposing] This is the same, yes?

ADRIENNE SCHMOEKER: With the Help Desk Tool implementation we did procure a new technology. The platform—the software is called Screen Door. It's a product of a company the Department of Better Technology, which was recently acquired by City Base. That's the platform that we used for the—the Help Desk technology.

CHAIRPERSON KOO: Yes. So, in the report it appears that a number of datasets that shall be on the Open Data Portal has yet to be published on the portal. What are the most common reasons for the delay?

ALBERT WEBBER: So, identifying data and eventually publishing it to the portal it—it generally varies per agencies. A lot of times what we'll find is that that as datasets were identified in the plan over the course of a number of years, sometimes you can find that there are changes in open data coordinators, changes in the priorities of the

agencies, cleaning the data, structuring it to make sure it's in a user-friendly format. So, you know, we're looking for it to be open data for all. So, we want to make sure that when we release this data it's clean, that I t's understandable. So those are some of the reasons why data could b delayed.

CHAIRPERSON KOO: So, if agencies are not complying, what has been done to ensure compliance? If they're not complying, what—what can you do?

ADRIENNE SCHMOEKER: So, we work with more than 60 open data coordinators from agencies, and then more than 40 open data coordinators from various initiatives, smaller commissions across the city. We've really made it a priority over the last two years to give the—this cohort of individuals across the city resources that they need to be successful in their roles. Now, if something is a little delayed or not on time we'll get in touch with the open data coordinator, look to understand what's going on, and if necessary, we'll escalate communications to agency leadership.

CHAIRPERSON KOO: Okay, so what can be done to ensure compliance? You noted in the—you—your already said that. Okay. [background comments,

pause] So, only 38 datasets were automated on the report. So, that 302 data sets can be visibly automated? So, what types of datasets were automated?

ALBERT WEBBER: There were a number of datasets that were automated, things from the Department of Environmental Protection, things from the Business Integrity Commission, but I do want to speak a little bit about, you know, the numbers of the 38 automations. So, those are the new datasets that we've automated over the last the reporting period. What the technology team at DOITT is also working on other automated data avenues, things like enhancing datasets, adding fields that the public is always asking for, geocoding data, and then also working on the technology infrastructure that's in place that's allowing us to automate data in a closer to real time fashion. So, while they're worth 38 automations, there is a lot of technology work being done behind the scenes that's—that's supporting more automation going forward.

CHAIRPERSON KOO: So, what type of data sets cannot be automated and why?

ALBERT WEBBER: Data that's not in machine readable format. Of course, when it comes to open data, anything that contains personally identifiable information or sensitive information, but it usually comes down to the structure of the data of where the data is stored.

CHAIRPERSON KOO: Okay. Geospatial references. Data is required by Local Laws 108 of 2015. Could you identify the challenges that you or agencies face in complying with this requirement?

ALBERT WEBBER: Yes. So, when it comes to geocoding data, the successive geocoding it really depends on the quality of the information that we're putting into the geocoding system. For the data that we've geocoded so far, the process—the process can be labor intensive. So, what my team was working on DOITT was we identified the datasets that qualified under this law. What we had to do eventually was for the ones that weren't automated, download those datasets, run them through the geocoding tool. It sometimes gives you multiple outputs that we need to recombine and push back up. So, the process can be a little bit labor intensive. There's a certain amount of technology or technical skill that's needed. So,

DOITT has taken on a lot of that work on behalf of agencies, but the overall process of—of cleaning the data, having it prepared to be geocoded is-is—it tends to be what can draw out the process.

CHAIRPERSON KOO: Okay, thank you. So, we received concerns about the size of the dataset files for downloading. Would it—would it be possible to include the information about the file size before its downloaded?

ALBERT WEBBER: Yes.

CHAIRPERSON KOO: Because if the file is too big, sometimes people don't realize it and then they click on it and then it takes forever to download. So, if you tell them beforehand like how big this file is, they'll think about it, and they'll think whether they have enough time or if they have enough space to store the data.

ADRIENNE SCHMOEKER: So, I think that that's a very important piece of metadata, but I believe we have some information about on that a primer page, but to your point, if people are not aware of that, they're not finding that information that's what matters more, right? So, continuing the down the road and trying to make sure that the

accompanying information about the dataset is understood. That's something we'd like to learn more about from people that you're hearing we're having this issue, and we can work to find ways to make that more noticeable firm to firm.

CHAIRPERSON KOO: The issue of privacy has been raised during our previous hearing. What does the department do in order to protect privacy of the New York City residents in relation to the Open Data Portal?

ALBERT WEBBER: With regards to open data, we lean heavily on our open data coordinators to coordinate within their agencies to ensure that personally identifiable information, that sensitive information is not included in the dataset. We lean on the agencies because no one understands their data as much as they do. We do brief checking on our end, but we heavily rely on the agencies to ensures that that information doesn't get to the portal.

CHAIRPERSON KOO: Okay. [background comments] We have Council Member Holden. Would you like to ask a question?

COUNCIL MEMBER HOLDEN: Yes, thank you, Chair. Over the years. I don't know if you can

answer this, but over the years what agency has really been extremely slow to respond to releasing data—data? I mean I don't—I—I probably am not going to get an answer on that, but I just wanted to ask it to hear is there one particular agency that you'd like to see move a little faster? Honestly.

ADRIENNE SCHMOEKER: Okay. Well, I would say that the Open Data Program has been in operation for about eight years, and the law was passed six years ago, and to Albert's earlier point, there is turnover who at the agency becomes an Open Data Coordinator, and even within the leadership of that agency. So, we've seen some agencies even in the past year who might have been a little bit slower to pick up speed over the last few months as we've done proactive outreach to make sure that an Open Data Coordinator who has the bandwidth to take on those responsibilities is appointed into that role, and then given the instruction and support to be successful in their role.

COUNCIL MEMBER HOLDEN: Okay, and my second question: There are numerous non-existent website links all over NYC websites pages or webpages. Could someone from DOITT perform dead link

analysis to track down the correct website URLs and notify the responsible parties that their websites have dead links, an updated coordinator? I know that's a—do we need more personnel? Is it because there's a lot of dead links? For instance, even on—on the open data About NYC Open Data web page, they have two dead links out of three, and about, you now, it says--two out of the three links on that page (1) in NYC digital and the other one is in Code Corpse. It's—it's—this is what happens. You get, you know, a blank page that's not available, and that's on your site, you know. So, it's really you get this. I don't know you can see it in the camera there. So, there—there needs to be on your own site, there needs to be, you know, a designated—somebody following up a little bit especially about NYC data. I mean, what--? Do you have an answer for that?

ALBERT WEBBER: I mean we can—we can definitely do analysis on that open data website because we want to make sure that there—that the data is accessible, searchable and available. So, with regards to the Open Data Portal, we—we'll definitely work to ensure that there are no dead links on it.

COUNCIL MEMBER HOLDEN: Some—somebody just has to run through and just like use it--

ALBERT WEBBER: [interposing] Yes, right.

COUNCIL MEMBER HOLDEN: --you know, because there's a lot of—it's—it's like frustration. So, nothing happens. I think I'm getting out of here. You know, so we need—we need really somebody, you know, an intern just to actually go through this thing check it.

ALBERT WEBBER: I agree, and then on our Contact Us page, we also has a section that's—that's meant for data questions and data errors. I'm not aware of any dead links on the site right now.

COUNCIL MEMBER HOLDEN: [interposing] Well, I'll give you the links to it.

ALBERT WEBBER: We will that be great from you guys.

COUNCIL MEMBER HOLDEN: I'll hand it to you. Yeah, and, you know, there's—is anyone using—here's another for—on the open. Is anyone checking the agency open data sets that should be released quarterly, monthly and weekly, et cetera. You know, our—that's being done. An example: DOT has a dataset or weekly street resurfacing by boroughs,

which has not been updated since 2014. So, what's the purpose, you know? And so that's—we have to get at the certain agencies and—and check them and make sure that, you know, it's working. You know, so you—you obviously will work on that and check some agencies?

ALBERT WEBBER: Yeah, we will definitely coordinate with Open Data Coordinators to make sure everything is updated in a timely fashion. That's also why we emphasize automation of data so much. I mean it makes it easier for everyone, and it makes the data—it gets the data where it needs to be quicker than we could manually. So, we will work with the Open Data Coordinators, and then also continue to emphasize automation. It—it gets the data where it needs to be quicker than we could manually. So—so we will work with those Open Data Coordinators, and then also continue to emphasize automation.

COUNCIL MEMBER HOLDEN: And—and you know, like I—I would guess you want to publicize your success work. The Twitter account at NYC Analytics, this says it has been it's been stale for three years, and on your website, the last news and media

posting was in 2015. So, the Analytics is, you know. Has nothing been done new and exciting since 2015? You know, nothing to publicize?

KELLY JIN: I can say there have been plenty of very exciting things that have happened since—since 2015. This is definitely on the top of my list.

COUNCIL MEMBER HOLDEN: But it's not—it's not being done.

KELLY JIN: It's one of my—top of my list items to look into.

COUNCIL MEMBER HOLDEN: It's a long time--

KELLY JIN: [interposing] I'm going to have to.

COUNCIL MEMBER HOLDEN: --with nothing, nothing happening, and just my last question I guess, can existing data be cleansed? You know something as simple as coming up with a master file for 311 operators to use so that they are consistent about what they type in. For instance say, you know, like they type in an address let's say 8000 Cooper Avenue on day, and then 80-00 Cooper Ave. and then another day the third one types in Avenue instead of A-V-E.

So can we create a consistent master data—master file so that it will save time and I think it will be more accurate? Is that possible?

ADRIENNE SCHMOEKER: Well, I believe what you're speaking to is the idea of data standards—

COUNCIL MEMBER HOLDEN: [interposing] Right.

ADRIENNE SCHMOEKER: --and we have that for open data when it comes to geospatial standards--

COUNCIL MEMBER HOLDEN: [interposing] Right.

ADRIENNE SCHMOEKER: --but every agency has their own services and processes and that they deliver to New Yorkers, and creating those standards at the agency level for data collection upstream goes beyond the scope the Open Data Program, but completely agree that data standards is something that really helps the usability front for the end user.

COUNCIL MEMBER HOLDEN: Alright. Alright, thank you for now.

CHAIRPERSON KOO: Thank you, Council Member Holden. [off mic] We are joined by [on mic]

Council Member Lander, and Council Member, do you have questions?

COUNCIL MEMBER LANDER: Thank you, Mr. Chair. So, I apologize. I'm a committee that was meeting across the street that one on freight, servicing freight and open data like two absolutely fundamental and critical things that most New York are not as focused on. So, I'm sorry that they conflicted today because I am very interested in bot of them. So, I haven't had time to read the entirety of the testimony, but let me ask one or two questions. You know, and it strikes me I think having kind of the open data—you know, this hearing kind of focusing on both the—the broadest requirements of the Open Data Law, and the specific kind of strategic efforts of the Mayor's Office on Data—Data Analytics, it's sort of instructive in thinking about there is sort of a spectrum we have here. We want as much as possible to be as transparent and an as open as possible, and obviously that is more than any finite set of people are going to be able to go and use for very focused and strategic purposes in making government work better, and part of what we want is an ecosystem that sort of

balances that. So, my first question comes I guess I see in the—the Open Data Law testimony a reference to the extra work you do guys are doing on the 311 data calling it a high value dataset. So I guess I wonder as you're thinking about the datasets you have, and obviously the law says all of them have to be put out there in this way, but clearly some of them are really rich and important and the ones people want a lot, and then some of them are much beloved very important data, but less likely maybe to get a lot of attention from the public. So do you have a hierarchy of which dataset are—you consider high value that get an extra deep dive to make sure the integrity is good, the updates are good and—and if so, how are you determining. Well, let me I guess ask if you have it before I ask questions about it.

ADRIENNE SCHMOEKER: So, there short answer is yes. So, with an inventory of more than 2,000 data assets, we do have to make sure that we're prioritizing in some way when we're thinking about we're going to make sure is a high quality to the public, but that being said, we work to make sure evert asset put up on the platform, you know, has to have a data dictionary. If there's an address needs

to have the spatial standards attached, but for example, one initiative that we did this summer is called Metadata for all, and we prioritize having the Library of Science community review and assess the metadata for top 100 most viewed datasets on the platform. We were able to pull that information and using some data we have from our technology provider, Socrata, and we also use Google Analytics to be able to understand usage. So, we are also a data driven program in addition to being a program about data.

COUNCIL MEMBER LANDER: And tell me a little bit more about, you about that usage, what that means? You know, what it is that you're looking at when you've chosen the hundred highest used or highest volume?

ADRIENNE SCHMOEKER: Sure. So, for that initiative specifically, we were looking at browser views. So, how many people are just coming onto this dataset and viewing it, and the difference there between browser views and broader view are API calls is a signal about machines are looking at the data, and when undergoing a usability initiative, we wanted to make sure we were understanding what are the most popular datasets that people are trying to use? So,

that's why we chose that metric as a way of figuring out well how do we make these datasets that are the ones that people are going to most even more accessible by improving the metadata?

COUNCIL MEMBER LANDER: And have you done some surveys of your users to try to understand the using—the data using community if finding works, finds not works, might have suggestions for improvements?

ADRIENNE SCHMOEKER: So, engagement with our users is really important to us, and we've gone about getting that feedback in a number of different ways. We did an audit of our Help Desk tool just this past summer to be able to understand where we can improve. We have quarterly events so that we're actually meeting people face-to-face answering what questions they have, what concerns they have. The Help Desk itself receives about 25 to 30 inquiries per week. So, we have people flagging data errors, data questions for us, and we're getting back to them in a timely manner, and we also—about—about annually have been conducting research about who our users are more broadly. So, about two years ago, we conducted research with our local firms here in New York called

Reboot that helped us develop user personas to help understand that non-profits use our data, students use our data because by design, the Open Data Program is built such that we don't necessarily know who all of the people are who are coming and downloading data sets.

COUNCIL MEMBER LANDER: Have you though about kind of including and ask with every download. You know, take a three-question survey, register with, you know, any of the kinds of tools that interact when you—you could say no I just want my data. Leave me alone, but--

ADRIENNE SCHMOEKER: [interposing] Not.

COUNCIL MEMBER LANDER: --you might be willing to say sure I'll take your two-question survey or yes I'd like to create user profile for easier use next time.

ADRIENNE SCHMOEKER: It's a good question. There are some data platforms out there that require you to sign in or require you to answer some questions before you can access the data.

COUNCIL MEMBER LANDER: I don't want to require, but, you know, three's a requirement and there's offer.

ADRIENNE SCHMOEKER: Yes. As of now, we've decided not to create that barrier of entry before getting to the data, but it's something we can look into.

COUNCIL MEMBER LANDER: [interposing] And I agree with that. I wouldn't want it to be a barrier to the data but, you know, that's—that form when you're on hold of the thing and they say after you get your customer service would you take a 2-question survey? I imagine there would be a way that you could structure this that would not be a barrier to the data, and you could easily just click no, you know, or I don't know. It—it might be worth looking at. It might give you some automated tools for understanding the user base better with—in a way that wouldn't add lot of staff time and for you so--

ADRIENNE SCHMOEKER: [interposing] It's a good idea.

COUNCIL MEMBER LANDER: Alright, and then my—my other question, and I got a chance to scan through the open—open data testimony more than I did the MODA testimony, but I—so it may be that you answered this in your—in your testimony, but in the time that I've spent on the MODA website, I found

some very good examples of—of MODA's Strategic Initiative. So, like if we had a problem we wanted to solve, we assembled the team, we looked at data. Here's what we did, and that's a great way of working when City Hall wants a problem analyzed. We've-we've talked with you guys I know about Vision Zero and how we get smarter in our enforcement. That's great. There also, the—the goal of working with our civic tech community so that—and then, of course we just—we provide all the data and people can do whatever they want, and if they come up wit great solutions wonderful. There seems like there's a sort of middle ground, which more problems to solve than the Mayor's Office on Data Analytics can solve, but with more focus than here is all of the data we got, and whether that takes the form of, you know, whatever people are doing, structured hacks or particular competitions where here's a somewhat larger set of things we think could be improved. We don't have the resources at MODA to like dig in on all of them. So, what--? And again, if I'm asking you to repeat stuff you said in the testimony, I apologize, but what are the ways in which you're trying to maximize what the broader community can do, you know, on things that

are high enough for priority that we could name them, and say, we think there's something here if they're not quite high enough for priority given staff at MODA itself to make it a deep dive of yours?

ADRIENNE SCHMOEKER: So, I would say that we've build strong relationships with the civic tech community in New York through Civic Hall through BetaNYC, through others, and have a continued dialogue around, you know, where there's interest from their side in addressing local problems. I do think there's opportunity to scope more problems and to put that out to this group of talent and good will, which we're very fortunate to have in the city. I wouldn't—I would say that scoping robust projects that can then lead to a robust implementation pathway is not a small task to take on, but there is opportunity there.

COUNCIL MEMBER LANDER: And I think this is exactly why I'm asking. I—I—my hunch is given what you guys know from the projects you've done, you're in a stronger position than the agencies are to understand that a successful project would look like, but that if you had some initiative some space where you kind of helped folks do that. You know,

you worked with agencies, CIOs. You helped them understand how to frame a questions, and what a real project would look like that you'd be in a lot stronger position to kind of put that out into the world. Like sometimes the agency might decide, you know what, that's so worth it, we're going to invest some money in it next year. You might decide it's worth it for you to pick up. It's a high enough priority, but I could also see a step on the ladder, which was, you know what, like that's well enough framed now we could go out to the community. Who knows, with a prize, with—you know, who knows what? You guys would have a better sense of what approach might be best, and that we could significantly increase the—the—you know, not infinitely, you know, but I mean how many projects can you guys do in a year with the staff. I don't mean that critically. Like you can't do that many--

ADRIENNE SCHMOEKER: [interposing] Yeah.

COUNCIL MEMBER LANDER: --can you?

ADRIENNE SCHMOEKER: No, I mean I—I—I think there are a lot of—there's a lot of—there are a lot of challenges to solve in New York City, there are 300,000 who work for the City of New York, 8.5

million New Yorkers. If we can tap into their creativity to help solve problems, the city is going to be the better for it. I think that steps we have taken towards getting agencies to feel comfortable about how to problem solve with the public especially using open data as that vehicle and that tool is that this past compliance cycle we created a requirement where city agencies had to identify three civic engagement commitments that they were going to follow through on in the next year, and so it's a baby step. It's not here's a problem. Could you solve it, but already having an agency be proactive. When they publish a dataset Tweet about it, blog about it. Don't just put it on the portal and hope someone is going to discover it is a first step we're taking towards getting agencies to understand that beyond open data compliance, you're putting information out there, and this information that the public could use to help you problem solve. So, we're moving in that direction, and there definitely is opportunity there.

COUNCIL MEMBER LANDER: Thank you very much. Thank you, Mr. Chair.

CHAIRPERSON KOO: Thank you. So, we are joined my Council Member Ulrich. Yeah. Do we have

any more questions? Seeing no further--no further questions, I want to thank all the panelists for coming here to testify.

ALBERT WEBBER: Thank you.

JON KAUFMAN: Thank you.

CHAIRPERSON KOO: So, you may step down, and we have the second panel from the public, Yeah. [background comments, pause] The second panel consists of Dan—Daniel Allen in behalf of Manhattan Borough President and Lindsay Porter (sic) from BetaNYC and Alex Camarda—Camarda from Reinvent Albany. Now you can identify yourself and start now.

DANIEL ALLEN: [off mic] I am Daniel Allen.

CHAIRPERSON KOO: Turn on, your mic on.

DANIEL ALLEN: Hi, I'm Daniel Allen. I'm the Technology Policy Analyst for the Manhattan Borough President's Office. Thanks for having me. I'm going to presented Gale's testimony on her behalf. My name is Gale Brewer, and I'm the Manhattan Borough President. I want to thank Chairman Koo and members of the Committee on Technology for holding this hearing. As you may know, I was the primary sponsor of the Local Law 11

of 2012, the Open Data Law –[sneezes] excuse me—as a member of the City Council. Implementing the Open Data Law continues to be a major undertaking, and I would like to thank that Department of Information Technology and Telecommunications, Commissioner Saini for his—for his efforts, the excellent work of DOITT and the Mayor's Office on Data Analytics, and the Mayor's of Technology and Innovation. It makes me confident that New York City will continue to lead the nation on municipal data initiatives. Open data become very important for the city and our constituents. I fully support Intro 1137, which will codify MODA into the City Charter. It is a necessary step as we look to improve the institutional framework behind the Open Data Portal. I am proud of our city's influential position in Open Data, and thriving civic hacker (sic) community that is leveraging the many opportunities created by transparency. I want discuss the work of BetaNYC and how open data is revolutionizing the ways in which citizens and city government interact. Over the past few years I have partnered with Noah Hidalgo and Beta NYC to run the Civic Innovation Lab, a program dedicated to improving community board's use of data

and technology while training the next generation of civic leaders in the CUNY Service Fellow's Initiative. The program has been a huge success providing community boards with useful tools such as BoardStat, Board Track and After Hours Variance Dashboard. Another such tool is SLAM or the State Liquor Authority Mapper, which aggregates data on active liquor licenses, sidewalk café licenses, 311 complaints about bars, restaurants, clubs and restaurants health inspection and NYC onto a single map. This tool saves community boards considerable time and resources when forming a resolution on a liquor license application or a sidewalk café application. BetaNYC has been instrumental in implementing these technologies and instructing Community Board members and staff on how to properly use them, and addressing tech capacity issues for all 12 community—and 9 community boards. The improvement in tech and data resources greatly helps community boards fulfill their responsibilities and contribute to government decision making. The more the city communicates through its data, the more everyone stands to gain. Accordingly, when datasets are updated with greater frequency, the data is more

actionable. The Open Data Law is an engine for economic development, and we're only beginning to see its full potential. In order to achieve that potential, we must make strides to integrate open data coordination with day-to-day operations in every agency. The current wait times for requested datasets are far too long due to the volume of requests as well as the manpower we currently have to meet them. Some of our partners have been told it will take at least two years to release requested datasets. In other instances, agencies have data available on their websites, which is not available on Open Data Portal or updated on a regular basis. Utilizing Open Data teams, composing part by staff members involved in other agency operations will create inefficiency including staff members in the open data process will allow agencies to apply the expertise of city workers in determining useful data to release, and will also provide context to data descriptions, which only comes from years of experience. Providing Open Data Coordinators with appropriate resources as they guide their agencies towards compliance, ensures the city remains transparent, and increases the Open Data Portal's

value. In March, the Open Data Team partnering with BetaNYC, the Sunlight Foundation, and the Department of Citywide Administrative Services to host the first full day of training for NYC Open Data Coordinators. While I applaud this step forward, I believe this training should not be limited to a one-time event. Rather, it should be available quarterly as we on-board new Open Data Coordinators and expand to other agency staff members. The Open Data Law has brought the civic hacker movement to the forefront of good government oversight. Open data has become more than just transparency, it makes government more accountable, serves as a teaching tool for undergraduates and a pipeline to good jobs, empowers citizens and small businesses, and improves city services. There is much progress to make on this inside front, and I will do all I can to ensure the vision of 2012 continues to expand these initiatives and that New York City remains the national leader in municipal data innovation.

CHAIRPERSON KOO: Thank you. Yeah. Thanks Beta—BetaNYC.

LINDSAY POIRIER: Chair Koo and members of the New York City Council Technology Committee, my

name is Lindsay Poirier, and the Lab Manager at BetaNYC, a civic technology organization with 4,000 plus members that has spent several years dedicated to conducting research, developing curriculum, and producing tools to support the equity—the equity of and accessibility of the City's Open Data resources. We work in partnership with the Manhattan Borough President Gale A. Brewer's Office, the CUNY Service Corps and FCNY to improve digital and data literacy and train a new generation of civic technologists. Over the past several years we have collaborated closely with MODA and DOITT's Open Data Team and we are grateful for the opportunity to work closely with Adrienne Schmoeker, MODA's Director of Civic Engagement and Strategy to ensure that our shared goals of open data for all are achieved. We are in preliminary talks with MODA to support each other's efforts around New York City School of Data, and Open Data Week 2019 celebrating the sixth anniversary of the Open Data Law. We look forward to working with the city's new Chief Analytics Officer to continue pursuing this aim. For the past three years, BetaNYC has been conducting research into the information infrastructure supporting New York City community

boards. More specifically, we have sought to understand how community boards are currently leveraging data resources as evidence to support their resolutions, and for what use cases they would like to have better access to open data resources. Through this research we've been able to design tools and curriculum that configure the city's open data resources into dashboards, maps and visualizations that are much more accessible to the public than raw data sources. In October, we published two reports that outline our research into the information infrastructure supporting community boards, summarize our findings and offer recommendations to community boards, civic technologists, city agency representatives and elected officials. We have included an executive summary of these reports with our written testimony, and both can be downloaded in full from our website. Our testimony today is largely informed by this research. In regards to the proposed legislation, we are completely in support of Introduction 1137. We wish for MODA to be an independent agency, but are excited to see its powers written into the City Charter. Explicitly, we're excited to see MODA be a steward of an open source

analytics library that can increase visibility into how agencies develop and use algorithms. If properly implemented, this could help advance other initiatives we support such as open algorithms. In terms of feedback on the 2018 Open Data For All Report, the report marks the hard work and dedication of the city's Open Data Team and demonstrates that they are working towards making open data more useful and accessible to the public. Most—most notably, the team has published 629 new datasets bringing the total number of datasets on the portal to 2,154. We believe they should have the proper resources to manage these datasets as the number continue to grow. The team has engaged 1,800 plus New Yorkers at events during open data week 2018 and hosted three sold out events in 2018. This demonstrates their efforts to engage the public in topics related to open data and advance data literacy for all. And the team is working to identify research and highlight real world use case for open data and to design projects around these use cases. This demonstrates their commitment to user-centered design. BetaNYC believes that the implementation of the Open Data Law could be strengthened in the following areas: (1) While 805

of the datasets eligible for the Geospatial Standard have been geocoded, some critical datasets are not in compliance with Local Law 108 of 2015. BetaNYC understands that the city's Open Data Team is working under incredible constraints. The team is currently managing over 2,000 data sets each requiring regular quality assurance and documentation, and most requiring geo referencing all while both MODA and DOITT have been operating without key leadership figures for several months to years. For many on this team, managing the city's data assets is just one component of their job description. Budgetary resources should be allocated to ensure that the Open Data Team can prioritize performing quality assurance and getting the existing data assets in compliance with more recent addendums to the Open Data Law. While 89% of datasets have data dictionaries, many are only sparsely documented making it not only difficult for the public to interpret what different categories mean, but also opening up the possibility that the public will interpret the data incorrectly and draw inappropriate conclusions. BetaNYC is in support of the Open Data Team's Metadata for All Initiative, which has advocated for incorporating big

narrative description of the contents of each dataset published on the Open Data Portal to its documentation. We believe this effort will require considerable time and resources including meeting with the data producers for each data set at each agency to document key terms and concepts and translating the subject matter expertise into terms the public can understand. The initiative should be funded adequately. (3) Community Boards have described wanting access to certain information that is currently not on the Open Data Portal either because no agency is collecting the data, for example vacant storefronts, it is not in accessible format. For example rent stabilized units or it is not yet available on the portal. BetaNYC has submitted requests to the Open Data team for a few of these data sets. In one case we learned that the data would not be published for a year and a half, and in another case we learned that the dataset had not yet been scheduled for release. We hope to start productive conversations on how we can ensure that data that currently exists in the community has deemed a priority can be published in a timely manner. (4) While agencies have committed 230 plus

forms of civic engagement around open data, we hope to see resources allocated to arrive—to allow for more meaningful forms of engagement. Currently 5 of 70 agencies have committed to hosting focus groups with uses of the data. Four of 70 agencies have committed to producing tools and sharing them to project's library, and one of 70 agencies has committed to producing curriculum on their data resources. User engagement is essential to ensure that the data is structured to meet diverse needs and that jargon is properly explained in data documentation. However, we also recognize that the Open Data Coordinators are strapped for time and resources. To make broader civic engagement possible, we believe that every agency should have an Open Data Team, which dives technical subject matter expertise in representing diverse offices within the agency that can collaborate to support data quality assurance, documentation, public engagement, and tool building. Funding should be allocated to support this. There should be more—(5) there should be more opportunities for collaboration between Open Data Coordinators at different agencies. Often the most important data insights do not emerge from analyzing

and visualizing one data set produced by one agency, but instead by integrating data for multiple datasets. However, because the city's data resources are often produced in silos, it can be extremely difficult to configure multiple datasets into a single view. Each city agency had their own unique way of identifying businesses, restaurants, buildings and lots, and when their datasets characterized these features, they typically only use their own standards of identification—of identification to reference them. For example, BetaNYC has tried to design maps of potentially vacant storefronts throughout the city by integrating several datasets from DCP, DOHMH, CCA and the State Division—Division of Licensing Services reporting the location of commercial units and active business licenses. However, because businesses are referenced with a single set of identifiers in each data set reporting license, this has been close to impossible. Coordinating efforts across agencies could highlight opportunities to link information across data sets. Local Law 250—(6) Local Law 251 of 2017 required not only that DOITT review the Technical Standards Manual every two years, but also that they establish a method to which the public can

comment on it. There are many areas where technical standards can be improved. Agencies often geocode addresses differently, use different terms or naming conventions to refer to the same concept or use different stylistic conventions for filling in standard data values. For example, in the 311 service request dataset, the community board column is formatted 01 Manhattan whereas in the DOB's Building Permit datasets the Community Board Column is formatted 101. While agencies understand these nuances, it can be very confusing for users who may draw their own conclusions for why words are classified differently or values are input differently in different datasets. In promoting interagency coordination around data quality and release efforts, DOITT could more readily identify mismatched schemas and stylistic conventions in the datasets and use this feedback to strengthen the Technical Standards Manual in ways that make it possible to link data across datasets while also supporting the public in developing a civic vocabulary. We would like to work with DOITT to host events and solicit broad public feedback on the Technical Standards Manual, and finally (7) we hope

that future releases of the Open Data for All Report can include a headcount of MODA positions filled, positions available, and the annual budget. Thank you for your time. [background comments]

ALEX CAMARDA: Good afternoon, Chair-

CHAIRPERSON KOO: [interposing] Alex Camarda, Reinvent Albany. Yeah, please.

ALEX CAMARDA: Good afternoon Chair Koo, and members of the Technology Committee. My name is Alex Camarda. I'm the Senior Policy Advisor for Reinvent Albany. I'm not going to read my entire testimony. I'll just summarize key points from it. First, we want to applaud the appointment of Kelly Jin as the new Director of MODA. This was a hired that was long in the making, and so we're pleased that she's aboard. Second, we want to support Intro 1137, Council Member Adams' bill. We actually testified before the 2019 Charter Revision Commission for the codification of MODA in the charter, and we're pleased to learn that the Council can do that on its own. So, we do echo the suggestions by DOITT and the Administration that the bill should more fully include their role in—in the process of making data available to the public. And then third, we did

evaluate the Open Data 2018 Annual Report, and have a number of things that we would like to highlight that we think we think are good developments, and then also highlight some areas that we think improvements can be made. So staring with the—the positives, we're very impressed that the MODA team and DOITT were able to ensure a great deal of compliance with geocoding, and also with data dictionaries. There's a very high percentage of datasets that have both in the portal and that's very encouraging to see. We're thrilled by the—the community that MODA has built through the many open data events that it has. If you've ever been to an Open Data Week event—if you haven't been to one, I highly encourage you to go to one. You really see that there's many stakeholders that—that care about the Open Data Portal and the data sets that are made available. I think some of the data that it's released in the reports supports that. I think one statistic that hasn't been highlighted there's actually one million unique users that view six million pages of data from the Open Data Portal in—just in the Fiscal Year 2018. So that really speaks to the—the use and interest in the datasets. City agencies have a—have a mixed record

on compliance, but one thing that they have done well is they've actually put 1,627 datasets into the portal that were actually not in their annual plans. So, we think that shows that they're thinking about putting datasets in the portal, and while they don't always stick to their plans, they are actually organically putting datasets on the plan or in the portal that they didn't initially—initially intend to. We think the new civic engagements commitments—commitments by city agencies is an encouraging step. Obviously using the data is important. We can warehouse data in the portal, but if it's not being used, it really has no purpose. So, the—the use of the data is much more important than just releasing the data, and I think the report speaks to that in highlighting the use cases upfront. With regard to improvements, it was already highlighted by the chair and by DOITT some of the challenges around automation. We think automation is really important because then we—the public will have the data in real time, and as was said before by the administration, then they don't have to spend as much time reminding agencies to put updated data in the portal. So, that's something that we deeply care about. We're

encouraged to see that more datasets are planned to be automated, and hope that occurs. I spoke about the agencies putting datasets in the portal that were not in the annual plans. Unfortunately, they're also not putting a lot of data in the—in the portal that is in their annual plans. By our count, only 42% of the datasets agencies promise to publish were actually released in five years. That could reflect shifting priorities, but we'd like them to—the agencies to stick more closely to their plans and, you know, it's great to release data in a non-year plan, but should also release the data that is in your plan unless there's a really good reason not to. We think some of that—I'm not sure--this wasn't highlighted, but MODA actually creates datasets about the datasets, and we went through those. There were some inconsistencies that we saw namely around compliance by the agencies that we think need to be reconciled. And then lastly, the question was raised, you know, which agencies are releasing many data sets to the portal or doing a good job with open data versus others that may need improvement. So, we went through the data sets provided by MODA, and that's actually the last several pages of our

testimony. Beginning on page 4, you can see the agencies are ranked by the most datasets released to the portal. Obviously that's only one indicator. If—if an agency releases many datasets to the portal, it doesn't necessarily mean they're high value datasets, but the Department of Education is by far and away the leader having published 563 datasets. DOITT, as you might expect, is second. It's 157. There are several big agencies that have released, you know, a 100 datasets or more or tens of datasets. There are many others that have released under 20 even under 10 datasets. One that's sent out to us the CCRB has actually released 120 datasets. It's a pretty small agency. Yet, they NYPD has only released 14. So, that's one that we think should be looked at more closely. Thank you and I welcome any questions you may have.

CHAIRPERSON KOO: [off mic] Thank you. Yes. I have a question yes to BetaNYC.

LINDSAY POIRIER: Uh-hm.

CHAIRPERSON KOO: In your testimony it is noted there are many datasets, and if I would like to do this, they are totally unavail—unavailable on the

Open Data Portal, and have yet to be scheduled for release. What are these datasets?

LINDSAY POIRIER: Well, some of the datasets that they would like to have access to just don't exist throughout the city at all. So, for example, there are—there are no datasets characterizing the number of vacant storefronts throughout city. So, that's not on the portal because it just doesn't exist yet. In addition to that there's a number of datasets that aren't on the data portal because they are in a PDF format. So an example of this would be rent stabilized units, which is managed by the Rent Guidelines Board. There are another of other datasets that have come up through our research, datasets from DOT and DOB that—that community boards specifically reached out to us to ask to have available in the portal that have been a little bit slower to—to be scheduled for release.

CHAIRPERSON KOO: [on mic] Alright, thank you. Council Member Lander, do you have a question?

COUNCIL MEMBER LANDER: Thank you, yeah. Thanks to all of you for your testimony and your leadership on the issue for sure. On your final

point, Alex about agency differentials and particularly something like the NYPD where obviously the have a massive amount of data, of which have only released 14 datasets raises questions. But there's no denominator—I guess it's challenging because there is not a denominator, right. So one suspects that's a bad grade because it's out of a whole bunch that should be there, but I guess currently the agencies get to decide what—I guess is my question like, you know, how—how do we evaluate better? Like how do we know what we suspect that CCRB is doing a good job leaning into open data, and the NYPD is doing a bad job hoarding their data. Like how—can we tell that from this report? If not, we need changes to the law that would, you know, from when we do evaluation of Minority and Women Owned Businesses, right, we're evaluating out of like total number of dollars contracted. So, how—how do we—how do we get at that?

ALEX CAMARDA: [interposing] So, it's an important question. You know, when the law was originally written, the agencies had to do an inventory by the end of this year. That was updated in amendments by the Council that made it more of an annual process, and I think that recognized that the

inventory of city agencies is always changing as to how many datasets they are going to have. So, it's actually hard to get the denominator you spoke of.

COUNCIL MEMBER LANDER: But also, they get to do it themselves. They got that right so--

ALEX CAMARDA: [interposing] Yeah, this is something that they're determined.

COUNCIL MEMBER LANDER: --I just—so, I know that's not a dataset—

ALEX CAMARDA: [interposing] Yeah, you know, I think our

COUNCIL MEMBER LANDER: --that some of audit calls for, who's auditing that call that—that, you know, are they umpire as well as the producer?

ALEX CAMARDA: Right. I mean I think our suggestion would, and I think the Council can play a pivotal role in this. During the—the budget hearings obviously there's many questions for the agencies that come forth about their operations, and their financial needs. I think integrating into the Council hearings asking about datasets, and what have you released and what's a high value dataset, and what do stakeholders want to see in terms of the information that agencies make available. I think if that's

engrained in the process that that would be a good way to at least identify which datasets the agencies have and which are of most interest to the public and to the—and to the Council. I think if that's done on an annual basis, we'll see more of a prioritization by the agencies. I mean I can tell you just having done it, if you go on the website of almost any agency, there's usually data sets you can identify that they've made available to the public seemingly important, but yet have not put all of those in the portal necessarily, and I think it's just probably competing priorities. They're focused on their operational functions, but I think to the extent that Council and others highlight that, it will happen more frequently.

COUNCIL MEMBER LANDER: So, I'll take your, you know, your suggestion, and try to incorporate that into some of my budget questions in the way, you know, members do around diversity among senior staff, and a range of other questions. But maybe I'll offer a suggestion back as well because it seems to me that one very valuable role for the civic tech and advocacy communities might be to work with advocacy partners who care about the issues and might

not know that, you know, it's-so I suspect that the advocacy groups that think about policing wouldn't be surprised by the statistics you gave, but probably are not paying that much attention to it, and there might be some opportunities for partnership for advocacy linking you and, you know, subject matter advocates that might help drive some change here in ways that we could be partners as well, but—but would be driven well from advocates on the outside.

ALEX CAMARDA: Thank you.

CHAIRPERSON KOO: Thank you, yeah. See no more—no more questions, you can state that, and thank you for your participation.

ALEX CAMARDA: Thanks. [background comments]

CHAIRPERSON KOO: Any more public participations. Seeing none, this meeting is adjourned. [gavel]

C E R T I F I C A T E

World Wide Dictation certifies that the foregoing transcript is a true and accurate record of the proceedings. We further certify that there is no relation to any of the parties to this action by blood or marriage, and that there is interest in the outcome of this matter.

Date November 10, 2018