

CITY COUNCIL  
CITY OF NEW YORK

----- X

TRANSCRIPT OF THE MINUTES

Of the

COMMITTEE ON TECHNOLOGY

----- X

September 21, 2016

Start: 1:00 p.m.

Recess: 2:35 p.m.

HELD AT: 250 Broadway - Committee Rm  
14th Fl

B E F O R E: JAMES VACCA  
Chairperson

COUNCIL MEMBERS: Annabel Palma  
David G. Greenfield  
Barry S. Grodenchik  
Joseph C. Borelli

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

Dr. Amem Ra Mashariki  
Chief Analytics Officer and Director  
Mayor's Office of Data and Analytics

Albert Webber  
Department of Information, Technology and  
Telecommunication

John Kaehny, Executive Director  
Reinvent Albany  
Co-Chair, NYC Transparency Working Group

Noel Hidalgo, Executive Director  
BetaNYC

Joel Trinidad  
Open Data Provider

Tio Chino and Daphne  
Data Privacy Advocacy Group

Alexa Billier, Student  
CUNY Baruch College



[sound check, pause]

CHAIRPERSON VACCA: But it on vibrate so we don't—we won't be hearing extraneous noise. There's a lot of extraneous noise in these hearings to begin with I want you to know. [background comments] [laughs] I couldn't use this stuff. Okay. Thank you. It's 1:00 p.m. and today is September 21<sup>st</sup> and I want to welcome you all to the Technology Committee hearing here at the New York City Council. I'm James Vacca. I'm chair of the Committee on Technology, and I'm sure colleagues of mine will be joining me throughout the hearing. We have one colleague here, Council Member Barry Grodenchik from Queens, and I welcome him and other committee members that will be arriving. We're here today to conduct an oversight hearing on the Open Data Law, the 2016 Open Data Plan and the seven amendments to the Open Data Law that were passed by this committee within the past year. I sponsored two laws in that package, Local Law 110 of 2015, which requires the timely updating of certain public data sets on the Open Data Portal, Local Law 8 of 2016, which requires citizen examinations and verifications of the compliance of certain agencies with the requirement to public data

sets. Additionally, this committee passed five other bills, Local Law 106 of 2015 sponsored by Council Member Cabrera, which requires the preservation and archiving of data that would have otherwise been purged from the Open Data Portal. Local Law 107 sponsored by Council Member Gentile, which requires the creation of a data dictionary for every data set. Local Law 108 of 2015, sponsored by Council Member Ben Kallos, which standardized address data, a geo code address data. Local Code 7 of 2015 sponsored by Council Member Anabel Palma, which requires agencies to review FOIL requests that include the release of data for inclusion of that data on the Open Data Portal, and Local Law 109 of 2015, sponsored by Council Member Torres, which creates the response timelines for public requests on the Open Data Portal. Today, we will discussing the implementation of these seven local laws, the challenges that have arisen, the ongoing issues, and the ways we can work together to solve them both administratively and legislatively. The Open Data Law passed in 2012 requires DOITT to work with all city agencies to post public data online in a centrally accessible location, the Open Data Portal by 2018. As mandated

in the Open Data Law, DOITT must annually release an Open Data plan. The 2016 plan was released on July 15, and reports that the portal now contains over 1,500 data sets, over 200 of which were automated with an additional 212 data sets planned for release before 2018. Before we talk about the ways to improve the Open Data Law, I want to commend DOITT and the Mayor's Office of Data Analytics for their dedication to the implementation of this law. Like on so many other fronts, New York city is a nationwide leader on open data, and the hard work of DOITT and MODA along with help from the Council and dedication of advocates certainly deserves recognition. Yet, there is still work to be done, and the committee intends to advocate for ongoing improvements. This committee believes that data quality, and the ability to report errors found within the data have become a serious concern, and needs to be addressed. This committee also remains concerned with agency compliance both with the original law's mandates, and the mandates from the recent amendment made by the Council. The number of agencies reporting on their FOIL data releases as required by Local Law 7 were disappointingly low.

I'm hoping that today's oversight hearing will shed light on the successes of open data as well as on the many challenges we still face. With that said, I want to mention I've been joined by my colleagues. To my left Council Member Joseph Borelli. He's usually always on my right, but I'll take him on my left. This is actually on my right, and I will turn it over to Dr. Amem Ra Mashariki of the Mayor's Office of Data Analytics and Albert Webber of the Department of Information, Technology and Telecommunication, and we look forward to hearing your testimony. I do have to swear you in. Do you affirm to tell the truth, the whole truth and nothing but the truth in your testimony before this committee, and to respond honestly to council members' questions?

DR. AMEN RA MASHARIKI: Yes.

ALBERT WEBBER: Yes.

CHAIRPERSON VACCA: Okay, so thank you. Who would like to begin?

DR. AMEN RA MASHARIKI: I will, Chairman Vacca. Thank you.

CHAIRPERSON VACCA: Introduce yourself, please.

DR. AMEN RA MASHARIKI: My name is Dr. Amen Ra Mashariki. I'm New York City's Chief Analytics Officer as well as the Director of the Mayor's Office of Data and Analytics. I wanted to start by saying thank you very much for inviting us here to talk about the incredible strides that we're making—we've made over the last year with the work on open data. Again, my name is Dr. Amen Ra Mashariki. I'm the Chief Analytics Officer and Chief—and also Chief Open Platform Officer for the City of New York. I'm joined by Albert Webber, of the Department of Information, Technology and Telecommunication. We're here today to speak about the way MODA and DOITT have worked together over the last year to fulfill the requirements of the city's Open Data Law and its recent updates. With respects to open data and open data for all, when city leaders passed Local Law 11 of 2012, New York City set a high bar for effective and transparent government. Not only would we open our data, we would open it all. Last July, we committed to open data for all, a more inclusive vision in which all New Yorkers, not just the tech savvy can find value in all the data. We believe that every New Yorker can benefit from open data, and



open data can benefit from every New Yorker. Soon after we published *Open Data for All*, Council passed and Mayor de Blasio signed a package of bills bolstering the original open data law with requirements to make it easier for New Yorkers to access, understand and derive value from city data online. These new laws required us to be more responsive to public feedback, create better technical standards and geospatial data, and data retention, and strengthen our means of ensuring agency compliance. Together, they help anchor the Administration's commitment to transparency and equitable uses of technology around open data. One of the things that we've been able to do is build and open data ecosystem. MODA is the business owner of this initiative. As the hub of analytics in the city, MODA advocates for the use of open data in citywide data analytics. DOITT is the technical owner ensuring that technological capabilities are constantly improving to better meet user's needs. Agencies are the data owner. Our primary points of contact with agencies are their open data coordinating with whom we work closely to ensure that agencies comply with the law. These three entities

along with our friends in the City Council and the civic tech community help to form the city's Open Data Team. Together, we continue to improve the biggest and best municipal open data initiative in the world by focusing on one number, 8.4 million, the total number of New Yorkers. Over the last year, the new statutory mandates have catalyzed and alignment of our vision of open data for all with our technical tools and work processes. In order to implement these laws while growing, the usability, reach and impact of open data, we have drilled down on the following three areas: User research, feedback mechanisms and technical standards. With respects to user research, the first step in opening data should be focusing on demand. We support creative analytic thinking to enable our users to answer the question most important to them, but we also hope to spur demand in communities less familiar to open data. Open data for all means open data for app developers, agency analysts and mapping gurus, but also open data for Emma Yang, the 12-year-old New Yorker who published her analysis of NYPD data from the Open Data Portal, as the youngest ever recipient of Wolfburn's Research Mentorship program. To this end,

MODA has partnered with the academic community to examine where open data is reaching New Yorkers and where it's not. A Capstone at the New York University Center for Urban Science and Progress measured data poverty across the city. So we partnered with NYU CUSP on a Capstone project where we measured data poverty across the city. Data poverty refers to a condition in which communities or people lack access to, use of or representation within data that is nevertheless used to inform decisions that may affect them. This analysis is among the first of its kind, and will be released publicly later this fall. We also did Capstone with Columbia SIPA. We work with the Columbia University School of International and Public Affairs to assess existing and future opportunities for open data to advance the goals of the city's community base organization. Although municipal data is in high demand, CBOs pointed to several pain points in accessing and operationalizing open data. These results are informing conversations on ways to improve user experience on the portal. Local Law 107 of 2015 is helping ensure that data is more understandable for non-technical users by requiring

that every data set include a data dictionary defining the attributes and clarifying the usability of its contents. With respect to feedback mechanisms, the Open Data Portal is not a static product. It is a platform that should be used as a tool for public feedback. We are dedicated to soliciting user input at every step of the open data life cycle to improve overall data quality and increase data use. We want those who give feedback to understand what their contributions are being used for, what opportunities for change are available, and when they can expect to see the changes take effect. Substantive engagement is more than obligation. It must be default for open data. We did a comment analysis as well. This summer we reviewed every comment that has ever been submitted to the Open Data Portal, and developed a typology for the types of inquiries that come in. This has informed how we are categorizing future feedback in order to optimize the way we deliver customer service. Over the next year, we will be implementing a new technological solution to intake feedback with a centralized mechanism that allows for quicker responses and better tracking on our interactions with users. This will make it

easier for us to meet the requirements of Local Law 109 of 2015, which mandates timely responses to public requests. A better feedback mechanism gives us a better way to comply with the law. This spring as per Local Law 8 of 2016 on examination and verification, we worked with the Department of Investigations to develop a plan for examining the Housing and Preservation Department, the Department of Sanitation, and the Department of Correction for information assets that may—may contain public data. Before this year, we would work with agencies to craft publishing plans for their data. Now, the Examination and Verification Law, the Public Request Law and the law requiring coordination between open data coordinators and portal officers have enabled new means of locating data covered by the original Open Data Law. They also help enact a broader culture shift. As open data becomes the norm for city data, it makes agencies more aware of the data they have, and the data they produce and spurs better upkeep and disclosure of information. Many of the new pieces of legislation require updates to the Technical Standards Manual, the document of record that outlines all policies and protocols for open

data. As we evaluate each standard, we have three priorities, meeting the letter of the law, meeting the spirit of the law and maintaining usability to the maximum number of users. To do this effectively, public participation is vital. Take for example Local Law 108 of 2015, which mandates the formation of a working group to create standards for address and geospatial information on the Open Data Portal. Over the last several months, I convened a group of Open Data coordinators to draft recommendation to geospatial attributes, column headers and geocoding processes for data sets. The recommended geospatial standards reflect that data builds (sic) most frequently captured by city agencies, information that is in highest demand from public users, and attributes that will have the biggest impact on citywide operations once they are standardized. In the spirit of open data for all, we have invited the public to join the conversation. We received constructive feedback and helpful suggestions from members of the tech startups, national non-profits, research uni--universities and others. For example, once submission from the American Society for the Prevention of Cruelty to Animals, ASPCA, said, "These

new standards are a great step in the right direction and will allow non-profits like ours to avoid costly geocoding of addresses.” These comments are helping us refine our recommendations to the standard, which will ultimately be included in an update—in an updated version of the Technical Standards Manual later this year. I’d like to close by sharing a standout account of enormous potential of open data. This spring, Ben Wellington, a local data scientist who studied parking violation data from the Department of Finance, and found that some of the locations where cars were ticketed were, in fact, legal parking zones. When he brought his analysis to the attention of the—of the city, NYPD analysts confirmed his results also using open data, and took steps to ensure that ticketing officers would avoid the oversight in the future. Wellington, who runs the popular analytic blog, Icon NY, wrote the following when he saw PD’s response: “I was speechless. This is what the future of government could look like one day. This is what open data is all about. Imagine a city where all agencies embrace this sort of analysis instead deflect and hide from it. Democracies provide pathways to government to

learn from their citizens. Open data makes those pathways so much more powerful. In this case, NYPD acknowledged the mistake, is retraining its officer and is putting and monitoring to limit this type of erroneous ticketing from happening in the future. In doing so, they have shown that they are ready and willing to work with the people of the city, and what better gift can we get from open data than that. Open data lies at the nexus of digital services, public transparency and cutting edge civic analytics. But we are constantly aiming higher. We will continue to test our assumption, and try new ideas, collect information on our efforts and analyze our performance for points of improvement. We have committed to engaging the public throughout the process to keep us honest and on track as we continue to learn and improve. I would like to thank the Council Technology Committee for the opportunity to testify today, for its continued support of open data. At this point, I would like to turn it over to Albert Webber, who will walk you through our progress on each piece of legislation individually in order.

ALBERT WEBBER: 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 will be testifying today on behalf of the Department of Information Technology and Telecommunications. As you know, this year's annual update to the NYC Open Data Plan was released on July 15, 2015. In this plan, DOITT with the Mayor's Office of Data Analytics and MODA and city agencies collaborated to increase the quantity and quality of data in the NYC Open Data Portal. Since last year's update traffic on the Open Data Portal has increased receiving over five million hits with more than 2,000 user created use stream barriers. (sic) To date, nearly 1,600 data sets (sic) have been made available, 116 of which were released in the past year as per the Open Data Plan. These include restaurant and building inspections, the city record online, and the City Council own Constituent services. The infrastructure has been built for the automation of over 100 new datasets bringing the total number of automated requested data sets to over 200. Additionally, we have identified and published more than 40 unscheduled data sets. This demonstrates the commitment of agencies to make data available as quickly as possible. To continue

building on this past year's accomplishments, DOITT is actively seeking to expand its Open Data Team. We are currently interviewing candidates for three or four new positions that will be fully dedicated to the Open Data program. With this additional staff, we plan to improve among other things our communication the open data users, our technical infrastructure, and the quality of receiving data. The package of laws the City Council passed late last year and early this year is an integral part of open data retention. (sic) The following portion of testimony will outline our progress of each law.

Local Law 106 of 2015 with regard to the retention. This law requires DOITT to preserve record level data and to establish guidelines for the division of large data sets into smaller ones. The new guidelines will facilitate archiving of historical data as data sets get updated, allow users to follow trends over time, and improve the overall user experience. We are ahead of schedule for complying with this legislation. We have already begun working with agencies so data is preserved to maintain accurate and active lists for some of our more popular data sets. Appending instead of a place

in the existing records, and dividing some of the larger data sets into smaller, more manageable data sets. The Open Data Team has already drafted the standards for data retention and will open them up for public feedback later this fall.

Local Law 107 of 2015 provides the data dictionaries. This local law requires DOITT to include a plain language data dictionary with every data set on the Open Data Portal. The data dictionary has made data more understandable by our users, and fits into the vision of open data for all. To help ensure that that data dictionaries have a view to the public, we plan to collaborate with users to create data dictionary templates, and to open our open data—our open—our data dictionary standards for public comment later this fall. Since the law was enacted, we have—are already out of the box at least 70 data dictionaries, and we'll work towards including data dictionaries for all data set in the Open Data Portal by the end of 2017.

Local Law 108 will provide for geospatial and address those needs. This law requires the creation of standardized geospatial address layouts for all data sets containing the address information.

This standard will result in more robust data being added to the portal that allow users to easily perform spatial analysis across the multiple datasets, and build applications at levels with more information. As previously mentioned, Mayor de Blasio designated Dr. Amen Ra Mashariki Chief-Chief Analytics Officer to continue a working group that will provide recommendations of creating uniformed geospatial standards to both DOITT and the City Council. To develop these standards, we've held several meetings with open data coordinators, agency stakeholders, the Department of City Planning's Geo Support team and open data advocates. We then gathered this feedback, drafted standards, and opened it up to the public earlier this month. We look forward to finalizing these mandates and providing the technical value to agencies to adhere to this legislation.

Local Law 109 provides for the timely responses to public requests. This law requires DOITT to provide additional responses to requests for public data sets within two weeks of receipt, and the agency to which DOITT referred a request to make a final determination within two months of receipt.

Over the past year, the Open Data Team has worked closely with agencies to respond to data requests dating as far back as 2011. We found that many of these requests for data sets were—already existed, reinforcing the notion that users have a true knowledge of the city’s unique data landscape, and the agencies have been proactive about their use of data. Improving feedback has continued to be a prior of the Open Data Team moving forward. In the coming months, we will be improving and consolidating our communication channels to make sure there’s a clear line of communication between us and our user. We invite the public to continue requesting data sets, to ask us questions and to comment on our proposed data standards.

Local Law 110 of 2015 provides the time of the data set updates. This law requires data sets updated on agency websites to also be updated on the Open Data Portal and fits with our strategy to provide regularly and time requests on accurate data. For timely updates, our team will work with agencies to ensure there’s a schedule in place to maintain up-to-date data on the portal. Regarding manually updated updates, our new staff will play a major role

in ensuring that data is refreshed, but will also work to identify data and use of websites that will require the annual updates. There are roughly 1,400 data sets that are not automatically updated, and we will work to ensure that they are all addressed. Additionally, over the last year, our technical team has made huge strides in making timely data set updates. For the first time we began the process of automatically update geospatial data sets, a process that will save us hours of work every year. Additionally, partnering with the Department of Sanitation our team will be releasing the raw data source from the PlowNYC application, which has been in high demand for many years. The data will be made available in year real time than snow events with plow locations being updated several times per hour, and older records being archived. This is the first time we will be able to deliver automated data, and more frequently than daily, and we hope to apply these technology enhancements to other high demand data sets.

Local Law 7 of 2016 provides for FOIL. This law requires agencies to review responses to Freedom of Information Law requests that include

release of data to determine if such responses received public data sets that have not yet been included on the Open Data Portal. The results of this review were included in the last update to the Open Data Plan, and will be included in updates going forward. To facilitate agency compliance through its Open Data and legal staff provided guidance to agency oversees and counsel's office. It fully laid out what agencies needed to consider and was reading that call with us.(sic) This legislation is extremely helpful in expanding open data discussions within agencies, and highlighting the association between data that is requested through FOIL and open data both of which are indicative of promoting transparency. Reporting on these metrics included input from open data coordinators, FOIL officers, general counsel, and communications officers, and we anticipate they will continue to help agencies prioritize data going forward.

Local Law 8 of 2016 with regard to the examination and verification process. So, last, but certainly not least, Local 8 of 2016 require the Mayor to appoint an office or agency to conduct a series of examinations and verifications of

individual agencies' compliance with the Open Data Law. MODA is that appointed office, and began the process of collaborating with the Department of Investigation and its commissioner on examination and verification compliance. All agencies that must comply this year, the Department of Housing Preservation and Development, the Department of Corrections and the Department of Sanitation have been briefed on the examination and verification schedule, and MODA was to give surveys to each agency by the end of September. Additionally, suggestions from the public will be collected throughout the fall. Thank you for the opportunity to testify this afternoon. These have succeeded in making the Open Data Law stronger, and we have been pleased to update on our progress. Open data remains a priority of this Administration and we thank our partners in the City Council and the civic community for their continued advocacy. And this completes our prepared testimony, and we look forward to answering any questions.

CHAIRPERSON VACCA: I want to thank you. I want to note we've been joined by Council Member Annabel Palma to my right. Your testimony was good.



I-I want you to know I'm pretty happy with one thing, and that is the Plow New York City action taken.

Last year we had complaints. I think we've made a little bit of news here. It's not easy for this committee to make news I want you to know.

[laughter] But we did make some news today because winter is here, believe it or not, coming up on us, and here we had a situation where last year the Plow New York City information was telling us where these spreaders were going, but we did not retain the information of where they have plowed, and people are saying that there's people who have never plowed, and then when they went into open data, they could only find that information going forth not going back. So here we are told that, you're going to be address this year, and plow locations will be updated several times per hour, and all the records are being archived. This is the first time this is happening, and I think this will give citizens a real handle on whether or not their street has been plowed. I cannot begin to tell you how many phone calls I get from people telling me that their street was never plowed during the heavy snow storm. How many times those streets were plowed, but the reality was is

that it was plowed, but everybody dug themselves out. They dug their cars out on the street, shoveled it back, and if the temperatures don't reach 32 or above, the snow has nowhere to go. Nothing melts. So here we will have a record—Excuse me. So, here we will have a record of that taking place, and I think that's good. So I thank you. So that will be in place for this year. Yes. Okay. What's the most common reason people use open data? What are they looking—what do you find that they're looking for more frequently? I know when you call 311, the—the main answer—the main question probably is noise. So how frequent are the calls to open data about a particular—are there particular things people are looking?

ALBERT WEBBER: So reviewing the data sets requested, combinations of communities, I'd say that the primary request is the violations that we've heard, violations from the Environmental Control Board and violations for restaurants.

CHAIRPERSON VACCA: Okay.

ALBERT WEBBER: [off mic] So—so overall it would be violations and additionally data from the MCA, which actually doesn't fall under the

jurisdiction of the city of New York. The violations and MTA data they're probably as much as has been requested. (sic)

CHAIRPERSON VACCA: So people look for MTA data but it's not there?

ALBERT WEBBER: [off mic] It is there. We—we do link to data that the MTA makes available, but we don't own or retain that data.

CHAIRPERSON VACCA: So what do people look most frequently for? I'm sorry.

ALBERT WEBBER: Other than that, violation. It's the most broad—

CHAIRPERSON VACCA: [interposing] You mean from Environmental Control Board violation data?

ALBERT WEBBER: [off mic] I mean from a variety of agencies. So whether it's violation through restaurant inspections the Department of Buildings and the Environmental Control Board. The violations tend to be the most frequently accurate data.

CHAIRPERSON VACCA: Now many of the portals will be zeroed in on—zeroed in on for updates? You said you're going to be providing updates to all the information or this going to be

portal by portal? How will the updates be provided?  
How frequently and where?

ALBERT WEBBER: With regards to all data  
sets?

CHAIRPERSON VACCA: Yes.

ALBERT WEBBER: [off mic] So, it-it  
depends on the updates that have been set by the  
agency. We work with agencies to I guess recommend  
appropriate frequencies that we think the public  
would want. At the end of the day, the agency decides  
within their business concerns what is the most  
reasonable update frequencies. From our end, we're  
trying to automate whatever it is that is mandated  
that is possible. We have 200 automations right now,  
but we're continuing to try to build that number, and  
then hopefully the new staff that we'll bring in will  
help us to also continue with that number.

CHAIRPERSON VACCA: Let me ask you so  
more questions. What agencies have not been as  
forthcoming as they should have been? I'm thinking  
that the police don't—I'm trying to be nice, you  
know, with politics-political-political terminology,  
but I'm thinking of the Police Department, and I know  
that we have gaps here, and I want to know what we're

doing. What agencies besides the police have been a little behind their timetable, and what are these, and what are the agencies number one? What have you done to--what is the extent of their being behind, and what have you done to get them to--to the table as for--as for the law?

DR. AMEN RA MASHARIKI: So I think agencies have made strides across the board. So with regard to the NYPD, crime data has been something that's been heavily requested over the last two years, and--and I believe it was last year, they actually released for the first time intimate level crime data. So we can go to neighborhood for the seven major felonies in New York City and actually see and map out where these crimes are occurring. NYPD has also added a number of data to their actually website, which we also went to with a variety of different data sets. So I would say that the NYPD has made real strides going forward. I mean overall, I--I wouldn't say any particular agency is not fully committed to doing what they're supposed to do. Agencies had progress across the board, and we've had automation through--through all sorts of agencies, and all sorts of new data sets as per

requested on the Open Data Portal. So I can't say there any one set that they haven't been doing what they need to be doing for the Open Data Law.

CHAIRPERSON VACCA: The Police Department, they--has complied?

DR. AMEN RA MASHARIKI: They've--I mean yes.

CHAIRPERSON VACCA: So they're--so I was wrong. They--they have submitted everything on time, and we have all the information on--on open data we're supposed to have in the PV. (sic)

ALBERT WEBBER: So they've--so if we break down each legislation one by one, I'd have to go back and get actually exactly what we may not be using that they've been using. More data set over the past year. They've responded to the public data set requests. I'd have to--I'd have to go back to any specifics, but they've made real strides over the last year.

CHAIRPERSON VACCA: Why--well, I'd--I'd like you to get back to me, and I'm going to--when you get back to me, I'll get back to you because there are agencies we've analyzed that we don't think we have total compliance with. I--I wanted to ask

further, though, about that. Can you meet the deadlines set by the legislation in all cases that the Council has legislated? Are you able to meet those deadlines?

DR. AMEN RA MASHARIKI: It would be difficult with the newly—the four lines that we’re going to bring on board. So with the four lines that we’re going to bring on board, we do feel confident that we can hit the numbers with all the goals in this legislation.

CHAIRPERSON VACCA: Let me—let me clarify that. Are you interviewing right now, or are you advertising?

ALBERT WEBBER: We’re interviewing.

CHAIRPERSON VACCA: You’re interviewing so the advertisement phase is finished.

ALBERT WEBBER: Yes.

CHAIRPERSON VACCA: Why is that some agencies now are hiring open data liaisons, and other agencies are trying to use existing personnel with previous responsibilities to be their open data liaison? There is not a consistency? I feel every agency at this point has to have a dedicated open data person, and many of the agencies do not.

They're looking to use people that wear several hats, and I think that that minimizes the importance of what we're trying to do when it relates to transparency and government accountability. So where is—where is our city in regard to that?

DR. AMEN RA MASHARIKI: So one of the things that we've done over the past year, which I think actually was the catalyst for people hiring, agencies hiring actual open data liaisons was that we moved from sort of an every so often meeting with ODC to a monthly meeting with ODC, and even in some instances like I have alluded to, the agencies that we know have a tough road to hoe, we sort of engage them one-on-one. And so we had a very sort of flexible scheduled engagement strategy with the agencies where they're seeing that ODC is actually—this is more of a full-time commitment, and there's a lot more that needs to be done here. And so, we've seen that move to sort of hiring of people. And I think as we move forward that's the strategy that we want to continue to grow within the agency is bringing on people who are full-time committed to overseeing. What you also saw was ODCs who were there, over the previous couple of years, move on in



their job commitments whether it's to another agency or somewhere else within that agency, and the people that replace them, they're meant to be full-time. So you're seeing that shift happen in many different ways, but we're pushing aggressively for the agencies to commit full-time ODCs because we're engaging them in a full-time manner.

CHAIRPERSON VACCA: Does the PD have an open data person?

ALBERT WEBBER: They do, yes.

DR. AMEN RA MASHARIKI: They do. The one and there's a difference. Okay, and what do you do when you note an error in one of the steps that you posted? Have you noted errors? Has the public brought your attention to errors? How do you find out the errors and what do you do?

ALBERT WEBBER: So we've received—we-we, you know, come to—come across errors in a variety of fashions. One through comments in the Open Data Portal. Through social media, sometimes people Twitter accounts reviewed and they Tweeted these errors. So I think it also depends on the type of error. So I think errors that an agency may be aware of, or that happens somehow through the automated

workflow, if we see that on our end, we can—we certainly control it, and we'll work to correct those errors. And I think there is also the—the fact that errors that—that may come in as a part of operations from an agency. So for instance a ticket for—is accidentally with the data via 3016 as opposed to 2016. On our end we're not going to mandate an agency changes their operations. So, if the agency does find that error and points it out to us, and wants it corrected, we'll work with them to put that—that fix in place. But we won't mandate agencies to change their operations because it's possibly a clerical error that happened to go back in.

CHAIRPERSON VACCA: So if a member of the public finds an error, who do they call? Where would they go if they find an error? Should they call your office or-or text you or email or what's the process?

ALBERT WEBBER: A comment. As of right now, it's a comment on that data set.

CHAIRPERSON VACCA: A comment to

ALBERT WEBBER: Daily no. I'd say weekly.

CHAIRPERSON VACCA: Weekly. So if someone reviews, then you can—you can make that chance.

ALBERT WEBBER: [off mic] Yes, but I—what I'd like to point out is that with the new staff we have having on board, that would be a part of the things in their purview, review the comments, and reviewing data set nominations, reviewing the errors and just overall looking to clean up data on the Open Data Portal. So as of right now, with our current staffing, we're not reading the comments on a daily basis. With the new staff, we will be increasing the frequency in which there will be more comments and requests.

DR. AMEN RA MASHARIKI: Can I?

CHAIRPERSON VACCA: Yes.

DR. AMEN RA MASHARIKI: Can I also add that we make ourselves available in many different forums for engagement. You will always see MODA and DOITT's staff at Packathon (sic), at Data Solutions at data events in and around the city where we engage the people who are actually using the data, and almost invariably we engage in conversations around why I'm using this data set here for challenges and

seeing that data set. So that's—that's one form. I have alluded to this social media, and so we're consistently sort of watching our social media, and people do tend to engage us there. Then there's also via the—via the portal, and another mechanism that we have is from MODA's perspective we've had many businesses and uses of data specifically around sort of, you know, B data and so on and so forth reach out to us and say, hey, we'd love to come in, and sit down and talk to you about what we're doing, and how we're using open data and some of the challenges we're seeing. So we make our doors open to any organization that reaches out and wants to have a meeting, and—and—and sit down with us. So we have—we have any number of levels of the folks to begin to engage finally with the community.

CHAIRPERSON VACCA: I wanted to bring up about FOIL. We did touch on FOIL before. We have agencies that are not compliant with the FOIL—FOIL regulations. Sanitation is one of them. Buildings is another one. Only about 30 agencies provided their data on FOIL for the Open Data Plan, and why—why they're not complying with the FOIL requirements

on Open Data, and what are we doing to get them to comply?

ALBERT WEBBER: [off mic] So responses to FOIL, to those FOIL kind of metrics are ongoing. Since the plan was published on July 15<sup>th</sup>, there have been a number of updates to that. We will get (sic) bills updates for our members on the Open Data Dashboard. So that there's transparency and it is reported. I think what's happened over the last year is that there's been—there's been a lot required of our agencies and ODCs and to your point, ODCs sort of need help. I mean we're asking them to really get a dictionary for responses and an unprecedented number of comments and the physical feedback. And I think that agencies are working their hardest to provide those numbers. With—with regards to—in some cases this changes the way that FOIL is recorders had to I guess track or review the FOIL requests that were coming. So we're now asking you to look at from the perspective of Local Law 11 whether or not the data they released is a public data set or just a plain data set, and I think there are some changes that they're going to have to adapt to and go through. But I do think the numbers that are confusing

decrease as, you know, agencies get used to this piece of legislation.

CHAIRPERSON VACCA: I—I hope that what you're saying is the truth because I—I have to tell you that we noted that the Department of Transportation has indicated that 521 FOIL responses contain public data not on the portal. How can so many instances occur? Five hundred and twenty-one FOIL responses are not on the portal. That seems like a very big gap. Is this agency being cooperative with you, or is this a question of getting the—the knowledge out there as to how to do this, or when or whatever?

ALBERT WEBBER: It usually has been cooperative. I mean they are an agency that probably receives one of the highest number of public data requests is on emissions. Even prior to open data being in existence, I guess officially in the City of New York, they had their own open data portal. So they were already releasing data on a regular basis. So we have a number of data sets from DOT. They have reported on FOIL, and we've brought this then a responsive data set requests. So by the letter of the law, they've—they've been working very hard to be

compliant, and maybe it's in the required data with all this other data. (sic)

CHAIRPERSON VACCA: Well, this number stands out. There's something not going right here.

ALBERT WEBBER: [off mic] And I think—and I think one thing that's going to help is before a legislation is passed, we've opened up the conversation before the release of data within the agency. So now we have open data coordinators involved, and general counsel, FOIL officers, and communications officers. So we've increased the conversation with all of these pieces of legislation. So I do think that we will see that we need further compliance by dealing with the other agencies on the portal.

CHAIRPERSON VACCA: Now the more 109 of 2015 is the Response Timeline Law.

ALBERT WEBBER: Uh-huh.

CHAIRPERSON VACCA: And that specifies that requests from the public must be addressed in a specific three-month period. An effort was made by MODA to comply yet almost all the requests are still being marked open even if they've been answered, and

some of them still have not been answered. It always seems like the good stuff. (sic)

ALBERT WEBBER: [off mic] So we're continuing our work with agencies to response to those requests. I think there--there are similar requesting back in 2011, and I think that's a--I guess some feedback for all our agencies to get to, but the agencies that continue to send in their responses, I think at this point we're around 75% when it comes to issuing final determinations, and agencies are still continuing to issue those--those responses. What we're doing on our end to ensure that agencies are completely aware of what needs to be done by the law, as Dr. Ra Mashariki had mentioned hold ODC calls on conferences to ensure that they're aware of all the piece of legislation. We've actually--we have gone to agencies and sat with them in some cases to sort of help review nominations or any other data sets that up there to help--help them comply as much as possible.

CHAIRPERSON VACCA: Let--let me say this, I appreciate that all the agencies are working together collaboratively, and I do know that things have gotten better as compared to when we had the



original legislation in my-originally at the site hearings. So I'm here for as Chair of this Committee, but that's--this is all about--this is the reason I sponsored Local Law 8 because I wanted a compliance mechanism. There has to be accountability at a certain point. Now, I understand that DOI recently approved the examination and verification process that will be used by MODA to examine agencies for compliance with open data. Now, can you describe the examination process--process that will be used based on that review by--the review of the assessment done by DOI?

DR. AMEN RA MASHARIKI: Sure. So, it's a--sort of a three-pronged process that merely some of what DOI has--has done and also some of what MODA and DOITT have done in the past. So it starts off with a survey. So we push out to the respective agencies a survey that has--and I can share some sample questions on the survey. But we push out a survey and the agencies are now required to fill--fill that survey out. And it really sort of walks them through a--a litany of questions around datasets, and we go through-- We--we constantly cull the comments, and request that on--online. So we'll put that, you know,

have you responded to this? Where is this data set and so and so forth. So the agency has a period of time to respond to that survey. Once we have the survey and we've called that survey, then we push it up for public comment. So then you make the public to give the community an opportunity to ask questions around well, what about this data set? I actually thought about this dataset and so on and so forth. And so when we do that, we use that as the process to go back and forth. So, now we take the public feedback and we use that process to go back and forth with the agencies. But ultimately where we land, we ask the agency to certify, and so we give them a document to certify these are the data sets that we now through this examination and verification. So this is sort of the verification part, but these are the data sets that through your own sharing of information. What we've seen through requests from the portal and then from public feedback is data sets—public data sets that need to be made public, and then it's their job to then certify and verify and sign off on that, and then that becomes public. So that's the process. You know, I—I think—can I also add to the workflow that we have in place with

respects to accountability. The way you—we looked at this was not only we had to keep moving on releasing data sets as per the original Open Data Law, but then we had seven new laws that we needed agencies to move fairly aggressively and quickly to comply with. And so we put that on their place, and we created a strategy in partnership with the ODC to do that. And so where you see—and I’ve—we’ve had full compliance and communication and—and cooperation with agencies on geospatial standards and data dictionaries and others. And so where you see agencies really moving aggressively to be helpful, we sort of take all of that into consideration. So in some instances with some laws, we—we—we want to ensure that there is movement. Some agencies are moving a lot more expediently than others, but we want to ensure there’s movement. The next step is I reach out to the ODCs, and quite frankly maybe some of the leadership within the agency to say hey, you’re moving—you know, we—we—we would appreciate it if you moved more aggressively on this particular law or this particular law. What we’ve—what we’ve refrained from doing is sort of picking out one law where we see an agency. They may be moving aggressively and—

and helpfully with others, and they're moving slow on-on these. We sort of remind them—we sort of remind them and continuously engage them. I speak with the team almost daily around are we getting movement? Is there passion there, or are we getting no response? We're not getting anything, right, and—and always the answer is we're moving. They may not even when we pass this. (sic) Once I reach out, then the next step is if we don't get the movement that we want in order to be compliant then we push this—a—a list to Commissioner Rose, and then the Commissioner has Commission—Commissioner communication. So right now that's the work. We refrained from sort of pulling the trigger on that, so to speak, because there is consistent movement. Speed can be faster in some instances with some agencies, but there is movement, and we're—we're appreciative of that and working with it.

CHAIRPERSON VACCA: Okay, we've been joined by Council Member Greenfield. Any questions for the panel from the council members present? Okay. No questions. I want to thank you so much for coming, and enlightening us, and it's something that

our committee will be monitoring further, and I know we'll be in touch with you. So thank you.

DR. AMEN RA MASHARIKI: Thank you very much.

CHAIRPERSON VACCA: Another panel would like to be heard. John Kaehny, Reinvent Albany; Noel Hidalgo, Made in New York City; Joel—Joel whose handwriting is terrible. [laughter] That's a video? Oh, Noel, is this yours? Wow, you should be—you should be a doctor. Joel Trinidad. Yes, come up. Can you make one more seat? We'll have four. If you would like to testify, you need to fill out a slip so if you'd like—you might as well testify, they're here. Okay. Alright, thank you so much and I have to swear them in, right?

LEGAL COUNSEL: No.

CHAIRPERSON VACCA: Oh, no?

LEGAL COUNSEL: You don't need to.

CHAIRPERSON VACCA: I don't swear you in. I just swear at you. Okay, he took it. [background ground comment] Noel, do you want to start off? Or John, do you want to start? Okay.

JOHN KAEHNY: Hi, my name is John Kaehny. I'm Executive Director of Reinvent Albany, and also

Co-Chair of the New York City Transparency Working Group and I'm testifying today both on my behalf and that of my co-chair Gene Russianoff of the Transparency Working Group and NYPIRG and of GWG member groups, Common Cause New York and Citizens Union. So first off, just thank you, Chairman Vacca, and the Technology Committee. Really, the most important part of my testimony is right now, and that's just emphasizing to you how incredibly important it is that City Council stay actively and energetically involved in pushing, cajoling and cheerleading for open data in this city. The origins of open data in New York are as a partnership between the Mayor, agencies, Council and the public, and your role is absolutely crucial, and without you being here, open data will not succeed in this city. So thank you. Thank you very much. A couple of quick comments. Really New York City just so everyone knows is looked at a leader in open data. We are watched carefully by cities all over this country and the entire world, and what we're doing here is very, very important for people everywhere. And in a lot of ways what we're doing now is entering a new phase of open data, but the phase that we're entering into

is that of fixing problems and listening to the public and making sure that the data is high quality and grows. The first part of open data is the part that we started with in 2012, which was publishing that data, and we've made tremendous strides now. So as you heard from DOITT and MODA roughly three-fourths of all the data sets that were called in that original 2012 Open Data Law are now published and online. DOITT is making great process in automating data sets in particular where I have to point out New York City is the gold standard for automating data sets, and that's real important because it means that automatically updated data once fixed at the agency is then fixed for the public, too. So it saves lots of time and money, and I want to point that out. I also want to point out that open data is starting to become part of the everyday conversation in New York City. When you go to a City Council hearing or a community board meeting you hear about 311 data. You hear about NYPD and DOT traffic and crime data, and it's really becoming a tool that more and more people are becoming familiar with. So open data and the idea behind open data is working in New York, and that's an important thing for us to emphasize.

Ultimately, going forward we think the really big payoff from open data is one that doesn't get talked about a whole lot, but it's going to be saving city agencies tens of millions of dollars by making it easier for agency employees to access their own agency's data, and that of other agencies. Because like us and the public agency people have to go looking and often googling to find data from other agencies. So we think one of the big payoffs going forward is to get as much data made public as possible both for the public and government. We're very optimistic, and I'm happy to be here today. The City Hall has made a huge push behind Mindy Tarlow, the Director of Mayoral Operations and Ann Roest, the Director of DOITT. Probably over the last four or five months, they've been vastly accelerating the tempo in Administration's effort and investment into open data, and they've called on many open data stakeholders, expert, non-expert advocates and other to get their ideas and to address long-standing problems with open data. So we think that over the last six months, the Administration, prodded in part by the seven very important amendments that Council passed. It has really, really started upping their



game on open data, and I'm very happy to be able to report that today. Because open data is hard. It's about transformation, and transformation is not always easy, and I want to thank you for the questions you asked today. I think they were right on target, and the target is how does the public talk back to the city and tell them when a data set is wrong or has a problem or something like that. And that's what some of the open data mimics were about, and that's what I think a lot of open data oversight and efforts are going to be about going forward. That is interactive open data, and that was a real huge shift in how government works because never before have we tried to create some process some way so that the public can tell agencies how to fix their information. We don't—we don't even do that really. So this is a pretty new thing, and we expect a lot of growing pains, but it's a super exciting time and—and we appreciate you being here in the committee and pushing that area. Along with thanking Mindy Tarlow and Anne Roest, I want to thank the guys that just testified, Amem Ra Mashariki from MODA and Albert Webber from DOITT. They really are incredibly impressive and hardworking civil servants and

they're—they're guys to work with, and we feel very privileged that the city has put so much talent into open data, and that's a testimony for us that—that City Council. I mean, pardon me, that City Hall really does care about this, and is trying to meet all of the laws' mandates, and really can meet the spirit of the law. I just want to highlight a couple of things real quick for you that happened this year that we think are pretty cool. One is A+ to DOITT for automating 100 data sets. Super important, and we hoop they—hope they can keep going this year and do the same thing. We give the Administration an A grade for publishing a lot of important data sets this year, the city budget, the city record online, NYPD's seven major felony crimes and a huge tack in (sic) is the information data set. We want to give a shout-out to the GLC (sic). You asked who the laggards were. We'd say they're one of the leaders, and early adopters of open data, and they're showing how great it is. We're also very happy that DOITT and MODA are now leading and responding all of the comments and requests on the open data portal. That was a—a huge issue for our groups the last two years, and we've very pleased to see that they're used in

the Open Data Portal. We're also happy that the Administration is attempting to comply with the Open Data Law as much as they can, and doing things like publishing their annual update on time, which is great because agencies don't always do that kind of thing, and they've done so much extensive analysis of open data user-users, including reviewing all the comments dating back to 2011, and looking at some of the web analytics. This is very high quality good stuff. They didn't even brag about it, but we'll brag about it for them. Some of the ongoing issues that your questions touched, but I'm going to hit them again is that there still is no clear process for fixing public complaints about data quality problems, and there's a lot of them. This is going to be something that my colleague and—and many of us are going to address again and again. But Mr. Mashariki gave the example of Ben Wellington from IQuant New York getting the police to fix how they deal with certain parking tickets. Very few people can do that. Very few people can call up the press office or call up MODA directly and get them to fix that directly. So we need a public process for dealing with the fixed data sets. A lot of the bills

that the Council passed are direct hits. They're totally appropriate, and we want to thank you again for that. But some of the issues that they deal with non-standard agency data, and the need for data dictionaries are really hard. And you asked if the city was going to be able to comply with all the mandates. They're not. We don't think they're going to be able to comply with the data dictionary requirement. Not because they not going to try, but because this is a big change, and this is tough, and nobody really knows how to do it, and this is— In the area of data dictionaries, but probably New York City is being watched by more other governments than any other. So how do you get data dictionaries that work for 1,600 data sets and over 100 plus for the agency? A big, big challenge, and we're going to work hard with the city Administration, but we want to flag that because we don't think that they're going to hit the target. We hope they do, but we think this is crucial because without a plain language data dictionary—the data dictionary is the thing that explains what the problems with numbers are. That's what it is in plain English. Nobody knows what the data is, and right now there's a lot

of data sets that nobody knows what they are except for the three or four people in the agency that produced this doc. So really they're gibberish or they're a foreign language that has not been deciphered yet. So a data dictionary is the Rosetta Stone that unlocks the data for the public and for other agency users. So we're going to be really paying attention to that as super important. We have some recommendations that we list out in some detail, but I wanted to call out a couple things to you about high-high demand data sets that are not coming forth, and you asked about NYPD data. CompStat 2.0, which is the Police Department's online crime website is not open data, and it's strange because it's a website that shows the data, and matched the data, but the data that underlies it, the information is not available on the Open Data Portal or NYPD website, and we think it should be, and we'd like to see Council push NYPD to make that available. So it's kind of an odd form of transparency where you— you put in online, but you don't provide the underlying data. Another data set that we know is very high demand is the DOT Paving and Milling Schedule, which is in a PDF format, and we think that

you should push them to make it into a machine readable format that can be mapped and put on line. So those are just some specific things about what agencies can do. I have a couple just quick recommendations for this committee, and we have written recommendation for the Administration that I will not speak to right now. One, we'd like to see you amend Local Law 7 of 20116, which is the FOIL bill, a very small change that would require agencies to name the data sets that have been FOILED. So we don't just know the number of data sets that were FOILED, but we actually get to have a list of them. So we can see. In the case you mentioned of DOT, what were those data sets that people were asking for and that they FOILED for? And then when we see the name, it allows the public and the Council to ask okay, why don't you put them on the Open Data Portal? Because that's the idea behind that law. So it—it does seem like it needs to be very slightly amended to name the data set requested, and whether or not those data sets are scheduled for publication on the Open Data Portal. We'd also like the Council to consider mandating through legislation the creating of a public data issues trackers. Like the federal

government's data.gov issues, which is what the Federal Open Data Portal uses to track publicly and transparently all of the open data questions they get for their gigantic Open Data Portal. And we think that might be a model here, and that Council should look at that and consider whether or that needs to be mandated. And lastly for you, we want to—I'm circling back to what I said at the outset, I really, really thank you and encourage you to keep going with oversight hearings, and also to meet informally with the City's open data community as much as you have time for and your staff has time for. We think it really, really matters and it helps public advocacy tremendously to have you asking questions, and to being engaged and interested. Thank you.

CHAIRPERSON VACCA: Thank you and I certainly am willing to meet with the group again. We met with them previously. I thank you for your work. I do want to note that you said some nice things about the city and the Administration and TLC, and I think sometimes we don't know nice things. So I'm glad you said that because I know the Administration has worked hard on this, and I know that TLC has worked and I appreciate their help. But

we tend to always articulate the negative side. I think you. I—I do want to say that Mr. Mashariki is still here, and that is a unique feature of the meeting by the whole, because usually--and I'm here in the Council more than ten years now--the Administration people that testify leave when they're finished and don't hear from the public. So I want to thank him for staying and listening, and I think that speaks to his level of commitment and engagement. I did want to give you an experience where in my own case I used open data. I had a young gentleman unfortunately who was killed in a traffic accident on a major stretch in my district on East Fremont Avenue, and I was then doing research and I used open data, and I found out that within four years there were 100 accidents on this stretch, and 100 accidents in four years. DOT had proposed a Vision Zero road common plan, and the community board voted no, and then a year and half had passed, and they had taken no further action. So I called DOT in. I made some tweaks to what I thought the plan should involve, some tweaks, but that generally I supported the Vision Zero plan, and it's now been implemented. I had been criticized pillar to post



in—and on literally Facebook this, and that. I've been called every name, things that I have been called I've never been called before, but I don't think I am, and— but it's going well. And I think—I think open data provided that impetus for my office to do that type of research. You know, I know communities that always say we want more police, or we want more traffic agents. They often just go to meetings to say we want more and more and more, but they don't have the documentation. They don't do the research. Here it's at their fingertips. Here we can do the research. I do question whether or not many of the community boards and the community groups have the knowledge about how to access open data, and I think that that has to be an inherent part of our city's effort, and that is to train the community boards first. They are city agencies. They advocate, but they're advocating with one hand behind their backs unless they have the training. I understand that many of them will not to be trained. Many of them do not think they have the time, and many of them do not think they will ever be able to be trained. Well, I hate to leave them behind, but if that's the case, we have to move the boards ahead

because they're in the main line of making sure people get city services. And when you go to agencies asking for personnel or asking for service delivery changes, this open data and what we've done here at the Council is really their friend. So, we do need that type of training, and I've been saying that for some time, and I hope that we can get down to the bottom of it. There really should be an appointee from the agencies, or appoint people-- appoint people at DOITT who are going out to the community boards and engaging them, and reassuring that this--that you don't have to be a computer genius to find out how to use open data, and that they should be reassured. And they can--once they find out it can help them, I think they will use it. Okay.

Next.

NOEL HIDALGO: I can't believe you said all those things in preparation to my testimony. I wonder whether or not you're psychic. So, Chairperson and council members it's a great honor to be here. My name is Noel Hidalgo. I'm the Executive Director at BetaNYC. In front of you, you have seven pages of kind of our detailed review of the activities that we've done over the last year. As

I've been told before, I shouldn't read every single word. So I'm going to just summarize it in—in kind of a plain narrative voice. First is a huge thank you to not this Council, but also to the Administration and the work that the Open Data Team, not only MODA but also DOITT, for their tireless efforts as well as other agencies that show up at these different types of events who actually take their expertise, take the time to demystify what they know, and how these technology systems have been put in place. Our organization, BETA NYC is really focused on figuring out how to demystify this government technology environment that we're in, and to use technology, data and design to hold government accountable and to really—to give everybody an opportunity to express their voice. So speaking of where we are with the city's Open Data Law, in the last 12 months our community has grown 700 new members, which is kind of a spike in where we've been since 2009. We're not over 3,700 civic hackers who are really ready to kind of understand this environment and—and put their talents to use to help our neighbors. In the last 12 months, we've hosted four significant events, which is I think kind of

helped pull back the curtain on where we went to go.

It was the NYC School of Data. We partnered with the

New York City Parks Department on Trees Count Data

Jam. We hosted two conversations, one on affordable

housing data, and another one with NYC 311 where they

released the city's second largest data set, I think

about 35 million rows of call entry data. DOITT was

also there with records on this. So through these

events, we really see that the open data community is

at a particular maturity that it needs more

investment. The seven pieces of legislation that

were passed are taking us in the right direction, and

they have really strengthened the city's open data

practice. Dr. Mashariki used ecosystem. I will

interchangeably use ecosystem and practice, and we—we

were honored as, you know, advocates and community

members to partner with the city and different city

agencies to help do these prototypes, and we think

that that is the appropriate step. I mean we are

partners, and so we were included. So, you know, we

can say that, but we really think that that's

exemplary of other communities that should be done

in—with other agencies. We commend the Mayor's

Office of Data Analytics for engaging all communities

as much as possible to gather that feedback and to incorporate the City's data—data users' voice and to the future direction of where we're going. And we absolutely agree with everything that our friends at the Transparency Working Group, Reinvent Albany and NYPIRG through their testimony. One of the things that you brought up is figuring out how to make data usable for community boards. We absolutely embrace the nation—the Administration's notion of open data for all, and for the last year we've partnered with Manhattan Borough President Gale A. Brewer, the city's Open Data Fairy Godmother on figuring out how to bring community boards into the 21<sup>st</sup> Century. And we spent a year kind of looking at their—oh, (a) we got a whole bunch of CUNY students. We put them through a boot camp, and so we're starting to prototype this open data education kind of curriculum, and in that process we did a gap analysis on community boards. And within Manhattan alone we discovered that community boards have a desire to use open data, but don't always have the bandwidth in regards to how fast their Internet is, the education, the tools, or the capability to process all that data. And frustratingly, you know, there is zero

best practices out there on how to teach the city's open data to itself. And luckily we partnered with Pratt Institute out of Brooklyn through their Savvy Program, and we've outlined a framework that we were able to (a) apply to 311 data, and then apply to the Trees Count data program, and so we have a—a good understanding of what needs to happen in-in the way forward. But we need to actually go out there and teach more community groups to—to use this data, and we need to figure out how to use the data.

Therefore, including this feedback and improving our curriculum, but sadly, we haven't been able to find any financial support to go out and teach the city actually its open data. In regards to kind of the things that we're starting to see within the seven pieces of legislation, the—the Data Dictionaries Law is a great start, and I love that the Administration has stated over and over and over again that the seven pieces of legislation are the floor not the ceiling. And we feel that every data dictionary should contain many tutorials explaining how to best explore the data, and embrace the data portal's functions. You know, in that, we also hope that the Administration starts exploring other tools that

better suit kind of the diversity of users that we have. One of them is a bulk data users. Our current open data tool doesn't allow for bulk—easy bulk data downloading, and that's a frustration that we've heard time and time again. But at the same time, we're confronted with a situation where the tool that we have is the best to use for the general public. You can go to the city's Open Data Portal. You can— if you have a fast enough internet connection, if you have a large enough monitor, if you have a fast enough computer, and you have struck a little bit of luck, you can actually sit there and play around with the city's data. And frankly, this is kind of the best tool that's out there, but frustratingly it doesn't meet all of the needs of everyone. And so, over the next year, we're going to continue to explore kind of these insights in these seven pieces of legislation that have been passed, and work with the Manhattan Borough President, and the Fund for the City of New York to figure out how we can develop more insights within the next step of open data. Some of our—some of the insights that we've already gathered within this year is that we hope to see that there is a workflow, a concrete workflow that

actively engages the public around what are the issues within a data set, and how do you understand a data set. We're starting to call this a user centered data release workflow. We helped pioneer this with the Parks Department and MOD and MOTI at the New York City Parks Data Jam. And for those of you who were geeky, it's a user-centered design process that really explores kind of like how do you start to uncover the key parts of a data set, and how to best use the data set. The other big insight that we had from our events this year was the New York City School of Data. We featured—over 372 tickets were sold. We checked in over 260 people. We had 18 sessions, 40 presenters. Sixteen of those presenters came from New York City government, and we had three elected officials. Council Members, we hope to have you this year in celebration, or next in celebration of the City's Open Data Law. One of the key components of understanding this ecosystem, and this practice is how do we include parents? I'm not there yet, but I know that many of my colleagues are parents, and so when we think about open data for all, one of the core components is making sure that we have events that are inclusive in making sure that



we can bring everybody in. And so, we encourage the City to explore events that are inclusive of parents, and make sure that their needs are also taken care, i.e., offer childcare at events. So in conclusion to our testimony, we have a detailed number of insights. We are seeing a handful of agencies that are really exemplary of the-kind of the technology practice as well as the open data practice. New Your City 311, New York City Parks, Department of City Planning, DOITT's GIS Division and the Taxi and Limousine Commission are—are really at the forefront of what we see where we want the different agencies to go. We think that they—that those teams continue to be supported and given the resources to lead by example. I know that some of these agencies or departments within these different agencies don't all—aren't always seen as cutting edge, and that's a bit of a frustration because we really want them to succeed, and they really need our support. For those agencies that get it, there's a huge opportunity now that the data quality is there to do more civic engagement events, which will lead to building better data guides and—and—and kind of tutorials. We think that the City is at the point where-- Well, this is kind

of like a tug of war psychologically. We think that agencies should be accountable for their own data release and that shouldn't necessarily be ultimately on the backs of MODA. So that's the direction that we would like to go, and to get there we want the City to be outlining more standards, coming up with concrete protocols, developing open source tools that will allow agencies who are at the forefront to host their data to really share their data in a--in a more inclusive way. And fundamentally, through that whole process, develop a practice, and that practice is something that we have outlined at the very end of our testimony. We're two years away from the--kind of the--the deadline of 2018 of the City's Open Data Law. We're starting to wonder about what is the future legacy. What is the infrastructure? Dr. Mashariki explained that there are three different agencies that all have different hands within different elements of the City's Open Data Law. It's great that they've been able to glue all of those resources together, and to be a well functioning, well oiled operation. It's great to see them getting more resources. But fundamentally, wonder how does this last into the next successive administrations and

Council leadership. And so, we would love to see a dedicated unit that brings together all of these different components from leadership and development of standards, technology and tools, evangelism to education and trainings, and really make sure that New York City is the number one city in the world in regards to open data. This testimony is all built with the things that we've seen from around the world, and we think that with just a little bit more of our shoulder to the wheel, we will get and be far beyond everyone else. Thank you.

CHAIRPERSON VACCA: Thank you, thank you. Thank you for your advocacy. Next. [pause]

JOEL TRINIDAD: Good afternoon. Thank you for making me speak at this Council hearing. I am Joel Trinidad (sic). I'm not a golfer, but I like that game. I represent Open data. I'm a product of the New York City's Open Data system. After winning the gaps in 2011, I quit my job and I been open data since then. I'm now part of the open data family, and I'm here to testify in position of the City Council on the need to modernize open data infrastructure in the city. New York City is a city like no other. I won't use your valuable time to go

through a long list of reasons, but I believe we can all agree that one final reason why it's so, is that it's the global center of so many industries, finance, media, fashion, information technology, advertising to name a few. Underlying all these industries is the fact that they are all knowledge based services driven data intensive professions, the technology primarily. And if New York City is to become the epicenter of the emerging urban informatics industry governing gov tech, civic tech and smart cities, I further submit that the current Open Data Portal needs to be redone to support this nation's industry. For we believe that open data is essential infrastructure, and as such, it needs to be open source, and we have proved that from the intense support innovation and experimentation, not just depending on part of them. And permit two federations so that not only city agencies can participate in growing these knowledge portals. But third parties like universities, non-governmental organizations and private sector—and the private sector can effectively collaborate with the city in creating a robust data commons open to public/private partnerships. Already New York City is home to Civic Call(sic), BetaNYC,

Big Apps and all these innovations. It's also home to two of the leading open source data portal providers in the world, New Civic (sic), which is now part of gov delivery, in your schools (sic) and for the FY recently by Open Gov, a direct part of Big Apps. We help governments around the world at every level from national governments, state governments and local governments use modern open source data portal platforms. In my mind, an open source solution allows the city to build internal capacity especially now that it's retiring staff to tap, you know, the genius of the Cloud, and the universities could tack on in the cops-CUSP. Interestingly, NYU CUSP actually stood up their C-Cam portal, and that's great that they can do that. And it also accelerates innovation and compliance to the recent amendments to the City's Open Data Law. Again, the innovation, the genius of the Cloud. There are specific things that we can actually leverage. In Carnegie-Mellon University they built a way to automatically compile data dictionaries once you upload the data. In-in-in Canada they built a way to do metadata creating, which we're extending for the City of Boston. In California, they have built a way for them to do data

set level permission at the data set level. So they can use the data portal not just for public transparency, but for sharing data with the workload component within the agencies. And these are things that are being built by governments, universities, companies like ours. So this is a real ecosystem, not just one vendor doing everything and you're trapped. You got a roadmap of that vendor. And I'm sure you're keenly aware of the urban innovation leader, and an early adopter of technologies like open data often comes the costs of creating Legacy investments that quickly get leapfrogs especially so, more so in the technology industry. So I submit that the current solution is in dire need of a rehab, and the City should consider stopping the Sole Source Open Data Contract especially now that they removed the New York City based providers in the same space. Thank you for your attention.

JOHN KAEHNY: [off mic] I just wanted to support—[on mic] Oh, pardon me. I just wanted to support Joel's call for a review of the open data platform, which is in our—our testimony. I didn't get to it, but in the specifics. The Socrata (sic) open data platform has some serious limitations and

it should be reassessed by the City's Open Data team and-and counsel should inquire into what their assessment would do.

NOEL HIDALGO: But fundamentally out of all the tools we've seen the certain usability of the Socrata data set-I don't want to stand-stand up here and be the poster boy or poster child for Socrata. There are some key features within the-the current open data platform that enable the city to achieve open data for all, and that is the-the fact that you can upload tabular data, and within a browser, I mean there are some caveats. As I said, you have the fast internet connection, you have a big browser. You've got to have a little bit of luck on a good day, but that-our current provider of the portal enables us to gout and to train to-to sit somebody down. Even on this particular laptop, I can teach anyone how to look at 311 data and give the context of community boards, and-and this to me there is not a-it's not an or situation. You've got to have one or the other, and I think through our conversations that we've had with the administration they understand that there needs to be alternatives. There needs to be multiple alternatives to servicing the city's open data, and

we BetaNYC we strongly encourage that exploration and that investment, but not give up what we have now because right now we get open data for all at least, a bigger portion of the pie than if we were to drop it and—and exclusively with some of the tools. But we look forward and encourage an investment in alternatives that are out there.

JOEL TRINIDAD: [off mic] This is an important piece. (sic)

[background comments, pause]

TIO CHINO: I'm a member of New York City Council Technology family. Councilman Vacca, thank you. We are Tio Chino and I'm Tio Chino, and this Daphne, and we are members of the newly created Privacy Advocacy—Advocacy group related to privacy especially related to online privacy. The majority of our members are open data, open software, freedom software, free software and transparency advocates who advocate for the sharing of all New York City's citizens data collected by public and private enterprise, true government contracted programs like LinkNYC or DOT. However, we don't believe privacy should be an excuse for advocates—for agencies to hide behind for not releasing open data.



DAPHNE: So we are here today to raise aware ness of the problems that massive amounts of public data collection can cause to the New York City--to privacy of New York City citizens. We have started an ad hoc advocacy group with the purpose of creating an official New York privacy guidelines for it. This official board could be composed of New Yorkers from various fields including communication technology, security and law experts as well as citizens who would debate implications and make recommendations to elected representatives when they discuss technology contracts that include the collection of private data, its safeguarding and potential implications. Thank you for the opportunity to speak before you today.

CHAIRPERSON VACCA: So you want to protect our privacy. Tell me where do you see--where do you see open data posing a danger?

DAPHNE: Well--

CHAIRPERSON VACCA: [interposing] Is that--is that open data of your interest or your concern or is it other aspects of technology?

TIO CHINO: It's--it's open data and privacy in general. For example, we heard that the

TLC is the fore, leading of open data--data sets, but the first data set included all the drivers' private addresses, of all the drivers' license plus where they live, which could lead to issues that we haven't even thought. So, for example, Uber has sent mailings to all the TLC drivers using those data sets. Do we--do--should the city citizens who are, for example, TLC licensees, be--their address be exposed to the general public--

CHAIRPERSON VACCA: [interposing] I-I--

TIO CHINO: --around the world?

CHAIRPERSON VACCA: --I have to tell you I did not that, and I-I do not agree that there are--our private addresses should be posted.

DAPHNE: In Addition, with regard to LinkNYC, if you look at the term--term of service, it's very vague, and specifically with regards to video, which supposedly right now is not active. But when questions are asked, there are very vague answers to some of these questions, and we would like to get better answers on that, and also find out how that would affected everyday citizens.

CHAIRPERSON VACCA: Well, if you have concerns regarding LinkNYC, I'd appreciate you

1 sending me an email so I could have those itemized  
2 concerns, and I would certainly ask the agency for an  
3 explanation, and I do think that the TLC matter  
4 should be looked into. I—I do not want people's  
5 addresses-- I mean transparency is one thing, but I  
6 do think that that is a stretch, and I don't—I'm not  
7 happy with that.

8  
9 TIO CHINO: Do they, the TLC—I—I looked  
10 at the benefit of recent TLC data, and it's not  
11 there, but in the past like a year or two year ago,  
12 it was there.

13 CHAIRPERSON VACCA: [interposing] So it's  
14 not the—all right let me qualify. It's not there  
15 now?

16 TIO CHINO: Today's data set it's not  
17 there, but the first data set they published was an  
18 Excel spreadsheet, and everything was there.

19 CHAIRPERSON VACCA: Oh, then—then I'm  
20 glad you clarified. It's not there now. It may have  
21 been put there in error, and it may have been taken  
22 down, and I do admit that with advent of open data,  
23 there may have been mistakes, but I—I do want to make  
24 clear that I do want to respect that—privacy rights  
25 of individuals, and I—I will bring that up. We do

still have—Mr. Mashu (sp?) isn't here, and I'm hoping that he will clarify that for the agencies that have left, but certainly I want that to be noted. Okay. I want to thank the panel, and there being no other— Oh, we have one more? Oh, we have one more. Alisi Billier—Alexis—

ALEXIS: [off mic] Alexis.

CHAIRPERSON VACCA: Alexis. I'm going to write it today. Come on up. [pause]

ALEXA BILLIER: Hello, hello. Oh, that is on. My name is Alexa Billier. I am not an executive director of anything. I just—I am someone who has a user story for share of the century. I'm a CUNY Baruch College student, a mathematic major and an enormous open data nerd, and this is a really— really recent passion for me. I remember when I—way back in February when I was just hanging out on line, and I happened upon the—to see the page for New York City School of Open Data. They talked about open data, and the civic—civic—civic hacking all this stuff, and I was like sounds like really cool. Like how come I haven't heard of this before? And I went to the—I went to this conference, and I was—I was really hooked by the amazing work that BetaNYC is

doing, like Gale Brewer was doing, all--everyone-- everyone who was involved in this initiative was doing, and I started to get interested in doing this as well. So what happened is I started a blog where I would essentially just look at open data released by New York City, and I'd see like what are some things I can find in it? What are some stimulations that I can do, and big people would be interested in looking at this data, and that effort has just been initiative that has proven to be very successful. It's something that's been modeled by work-on work that's been done by other civic hackers like Ben Wellington and civic technologists like Ben Wellington and Chris Long. The website gets like 100--100,000 hits a month, and--and, you know, it's the music from New York (sic) and the Polish and Gothamists, you know, very popular music. So all--all that is--is really great, and I want to say that, you know, my--my personal story would not have been possible without the work of the City Legislature. It would not have been possible without the work of the City--City technologists. It would not have been possible without the work of--all the hard work that has been done by everyone that's been involved in

this initiative, and they have been pushing this conversation forward, and has really made New York City a leader of all of open data, and for that I think I—I really want to—before I say anything else, I really want to thank everyone involved in that and the City Council specifically. You know, I think that as I look back on some of the things that we've done, I think I kind of—I—I get excited about like where we can go from here. Plenty of team (sic), the Open Data Legislation were both taken in full force, and I think in that—by that time we'll be seeing some of these—these—these really larger really amazing data sets that right now are siloed within individual—siloed within agencies that if they could release they would have just so much value. And I know there are dozens of—of databases like these, but for the purpose of the illustration I'll pick just one, which is one of my favorite ones, which Biz Web. The Department of Buildings has—has this application for Biz Web where anyone can go on the—anyone can go online, and they can look up any information they want to know about a building that's interesting to them. There are building some other building what—whatever, and you can find out things like you can

find out permit status, vacancies, construction details, addressing, violations, signage, elevator records, plumbing, façade, boilers, cranes, everything—everything there is to know about a physical building is there. The trouble with it is that this a—it's like a system that needs resources. It was built in a different time. It was an application that was built in 2001, and it's really showing now. It doesn't run very—very quickly. It has issues. It's built on top of a mainframe, and it would be amazing if this data could be liberated, and make public, made available to everyone in an easily accessible and consumer format by open data. Because that would mean that we can now also start with the individual fillers. We can go through this data and do some really amazing things with it. Biz Web violations data could be used to find patterns of abuse in New York City building management allowing us to, for instance, practically model and send inspectors to—to areas before disaster strikes. Local legislators could use it to get weekly updates of complaints in their districts, and allowing them to get a better understanding of constituents' needs before the phones start ringing. You know, civic

activists could build a modern information dashboard so that anyone can go online and look up their building and like find out interesting information, a wealth of interesting information about the building that they live in, or some other buildings that are interesting. Private developers could use Biz Web. They could lock into data to get a block-by-block building-by-building history of construction in New York City, and to help try and figure out where it's going to go next. Construction agencies could more into-more easily interface with this data to track the permitting process for building, hoping to accelerate the bill times. And finally, other agencies that don't have easy access to this data like for instance the Department of Sanitation or the Office of Environmental Remediation who are interested in using these data streams could finally be able to go in there and use this data to investigate problems that they face themselves, and to streamline their own processes and even possibly their own-their own legislative initiatives in a way that-in ways that were never possible before. And these is all, there are possibilities that are unlocked by this one, you know, one-one example, this



Biz Web agent. There are dozens of applications like this. Like for instance CompStat 2.0, which was mentioned by John Kaehny, which are—have a wealth of—a wealth of knowledge or wealth of opportunity within them, and the only thing—the thing that we need to to—to like use that is like make it public, make it available, and a lot of—a lot of people use that. And so what I think is most important how is—I think that, you know, City—City Council has been the—has been the leader in this field. The City council has pushed—pushed the open data legislations, pushed these—pushed these things that make this possibility possible, and now it's up to us as a community, government technologists, private developers, legislators, everyone—everyone that's involved in the Open Data Initiative to help push this conversation further and to, you know, get more research that's available to bring online these huge data sets. To Bring online more—more—more information, more ideas to do the hackathons, to do the idea sessions, to do—to get the—get the juices flowing that help make open data conversations that have made New York City a leader in open data so far, and that will become ever more important as we get closer to that critical

deadline when in-December 31, 2018 when-when technically speaking all readily available data sets should be available in the portal. And so these are proven point things. And I think that I really want to thank the City Council for the amazing work that's done so far, and I really want you guys to continue doing because this is-this is really important and really-really essential. It's a huge opportunity for social communication. I think it's a logic opportunity. And 2018 is fast approaching. We will face this pledge to do no less than to bring an entire city's worth of resources online, and let's make it happen. Thank you.

CHAIRPERSON VACCA: I thank you. It's what you really-you really hit something on the head. Let me say a couple of things. I use the Biz system almost everyday, and I've been using it since 2003. It's even before because I was a district manager to a community board before I came here. I have found Biz to be fantastic. I can check on demolition permits, alteration permits, new permits, stop work orders. So to me it's great, but I do think it has to be updated, and many of the purposes that you spoke about should be included in a revised Biz

system. Now, I do understand that the Buildings Department is looking at that. I'm not part of those discussions. However, but I do know that they're looking at redesigning the Biz System, and as someone who believes that this has been an--an unbelievable asset, could you send me an email with your suggestions, and then I have some--some of my own, and I'd like to put together a position for the commissioners for review based on my experiences.

ALEXA BILLIER: Absolutely

CHAIRPERSON VACCA: Isn't this great? I think it's been belated. It's--it's slow.

ALEXA BILLIER: Yeah, that is.

CHAIRPERSON VACCA: I know, I know, I know.

ALEXA BILLIER: If you want--if you want information on these--

CHAIRPERSON VACCA: [interposing] The government is slow, too. So, therefore, I'm kind of used to this stuff, but otherwise--

ALEXA BILLIER: [interposing] Yeah.

CHAIRPERSON VACCA: --it's--

2 ALEXA BILLIER: [interposing] If you want  
3 information on an individual building, this web is  
4 amazing--

5 CHAIRPERSON VACCA: [interposing] Yes.

6 ALEXA BILLIER: --but it also, you know,  
7 the interface is a little hard to get around, and  
8 it's really hard to get information on--

9 CHAIRPERSON VACCA: [interposing] Well, I  
10 have to tell you one thing I'm trying to do is I've  
11 been trying to do almost since I'm here in the  
12 Council, but that's something called a bad actor's  
13 bill, and that is how do we trace bad actors who  
14 don't maintain their buildings, yet they go and seek  
15 permits to build new buildings on top of the building  
16 that they already have, but they don't maintain, and  
17 there is not that ability to cross-section that. And  
18 that's something that I'm looking to do because why  
19 someone be given new permits to build more when they  
20 cannot maintain their property they have based on the  
21 amounts of violations they have be they through HPD  
22 or through--will though the Buildings Department.  
23 They should be told no new permits until you  
24 straighten out your act. Get the violations taken  
25 care of, pay the city the fines the City is owed,

and then we will consider a new permit for a new building. Right now, we don't have that capacity across agencies, and I know you alluded to that in your reform suggestion. So I'd appreciate something to let me try to be helpful.

ALEXA BILLIER: I've got that in here.

CHAIRPERSON VACCA: Thank you. Thank you very much. Okay, there are no further speakers. It is now 2:35 and I want to thank everyone for attending hearing of the Technology Committee. This hearing 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 September 30, 2016