

CITY COUNCIL  
CITY OF NEW YORK

----- X

TRANSCRIPT OF THE MINUTES

Of the

COMMITTEE ON TECHNOLOGY

----- X

October 1, 2015

Start: 1:00 p.m.

Recess: 2:40 p.m.

HELD AT: 250 Broadway - Committee Room  
14th Floor

B E F O R E: JAMES VACCA  
Chairperson

COUNCIL MEMBERS: Annabel Palma  
David G. Greenfield  
Steven Matteo

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

Dr. Amen Ra Mashariki  
Chief Analytics Officer and Director  
NYC Mayor's Office of Data Analytics  
City of New York Chief Open Platform Officer

Albert Webber  
Civic Engagement Manager  
Office of Advanced Technologies  
Dept. of Information Technology & Telecommunications

John Cheney, Co-Chair  
New York City Transparency Working Group  
Executive Director, Reinvent Albany  
Presenting Testimony for:  
Jean Rushnoff(sp?) Senior Attorney  
New York Public Interest Research Group

Katherine Gray  
Vice President  
League of Women Voters

Tara Das, Intern Head  
Social Work Library at Columbia University  
Digital Social Science Center

Josh Mumm  
Outreach and Advocacy Manager  
Common Cause New York

John Crouse  
Technology Fellow  
Gov Lab at NYU



2 [sound check, pause]

3 CHAIRPERSON VACCA: Council Member

4 Gentile is here. Oh, you're ready. Okay. [gavel]

5 Everyone please take their seats. It's now 1:00 p.m.

6 on October 1st, and I would like to welcome everyone

7 to the Technology Committee of the New York City

8 Council Oversight Hearing. We'll be considering

9 today seven bills that have been introduced

10 concerning open data. My name is James Vacca, and

11 I'm Chair of the Committee on Technology. We're here

12 today to discuss the Open Data Plan of 2015 in

13 addition to a package of seven bills that aim to

14 remarkably strengthen the Open Data Law. I've

15 sponsored two pieces of legislation in this package,

16 Intro 915, which would require the timely updating of

17 certain public data sets on the Open Data Portal, and

18 Intro 916, which would implement an Open Data Law

19 agency compliance audit to the Department of

20 Investigations. Additionally, we will examine five

21 other bills, Intro 890 sponsored by Council Member

22 Cabrera in response--in relation to the retention of

23 data on the Open Data Portal, Intro 898 sponsored by

24 Council Member Gentile, who would require the

25 creation of a data dictionary for every data set.

2 Intro 900, sponsored by Council Member Ben Kallos,  
3 which aims to standardize address data. Intro 908,  
4 sponsored by Council Member Annabel Palma in relation  
5 to the review of data requested through FOIL, and  
6 Intro 914 sponsored by Council Member Torres, which  
7 would create response times for public requests on  
8 the Open Data Portal.

9 I want to thank my colleagues for working  
10 on this important legislation. It's clear that open  
11 data is a priority for this committee, and to the  
12 Council as a whole. I am sure for the  
13 Administration. Today, we will discuss how the Open  
14 Data Law has been implemented, the contents of the  
15 Department of Information, Technology and  
16 Telecommunications July 2015 Open Data Plan, the  
17 challenges and issues that have arisen, and the  
18 various ways we can all work together to solve them  
19 both administratively and legislatively.

20 The Open Data Law passed in 2012 requires  
21 DOITT to work with city agencies to post data  
22 information online in a centrally accessible  
23 location, an Open Data Portal by 2018. As mandated  
24 in the Open Data Law, DOITT must release the--an open  
25 data plan annually. The 2015 plan was released on

2 July 15th, focusing on outreach and engagement with  
3 the general public. As part of this engagement,  
4 DOITT has taken measures to increase public  
5 accessibility including a data lens featuring--a  
6 feature that presents popular data sets in adjustable  
7 charts and maps. Before we get into ways to improve  
8 open data, I'd like to commend DOITT and the Mayor's  
9 Office of Data Analytics, MODA for their dedication  
10 to the implementation of this law. Like on so many  
11 other fronts, New York City is a nationwide leader  
12 when it comes to open data, and the hard work of  
13 DOITT and MODA along with help from the Council and  
14 numerous advocates deserve credit for it. These  
15 agencies have also made considerable efforts to  
16 ensure compliance and have succeeded in populating  
17 the portal approximately 1,350 data sets with 281  
18 additional sets planned for future release between  
19 now and 2018.

20           Because the upkeep of the Open Data  
21 Portal is such an enormous undertaking requiring  
22 large staff resources, technical knowledge and the  
23 coordination of every city agency, issues are going  
24 to arise. Additionally, as I'm sure many of you  
25 agree, there are always ways to strengthen the law

2 and this committee intends to advocate for ongoing  
3 improvements. The proposed legislation before us  
4 today is meant to remedy issues that have been  
5 brought up time and time again when we discuss open  
6 data: Compliance, user friendliness, data retention  
7 public responsiveness, FOIL integration, and the  
8 timeliness of updated data sets. The bills I've  
9 sponsored are important component to--components of  
10 this packaging focusing specifically on compliance,  
11 and the timeliness of data on the portal vis-à-vis  
12 data that is on every agency's website.

13 Intro 916, which would require the  
14 Department of Investigation to conduct open data  
15 audits would go a long way to improve compliance  
16 across all agencies. In government we sometimes say  
17 that trying to get coordination across many offices  
18 is like herding cats. Imagine being DOITT, tasked  
19 with getting the data that falls under the Open  
20 Meetings Law from every city agency. There are  
21 dozens of agencies each with different resources,  
22 different types of data and varying amounts of data  
23 that's overwhelming and compliance is going to be  
24 difficult I acknowledge. Research from this  
25 committee has led to the discovery of the exclusion

2 of much data that should be included on the portal.  
3 For example, 311 referral data and complaint data  
4 from the Mayor's Office of Media and Entertainment.  
5 But this process of random discovery is not  
6 efficient. There are many public data sets that have  
7 been excluded and there needs to be a systematic way  
8 to address this. A third neutral party, DOI, could  
9 be the tool that we need to inform all the relevant  
10 entities what is missing. It's called  
11 accountability.

12           Since the Open Data Portal was designed  
13 to be a one-stop shop for government data, it is  
14 important the information on the portal does not lag  
15 behind what is available on agencies' websites. At  
16 last year's hearing I identified such a delay in the  
17 Building's Department based on what I saw in the  
18 Building's Information System, the BIS system, and  
19 this committee identified this issue with other data  
20 such as HPD violation data. My bill, Intro 915 would  
21 address that lapse requiring that data on the portal  
22 shall not be more than three days behind what is  
23 available on the agency's website.

24           I think this is a great package of bills,  
25 and open data is important to the Technology

2 Committee, and we've made it a priority. I want to  
3 thank our staff for their wonderful, wonderful help  
4 in putting these together, the Speaker's office, and  
5 I would like to thank members of the committee who  
6 will be here shortly. We do have two--two sponsors  
7 here who will speak on the bills that they are  
8 sponsoring today for hearing, and I first call  
9 Council Member Fernando Cabrera.

10 COUNCIL MEMBER CABRERA: Chair Vacca I  
11 want to first thank you for your leadership with the  
12 Open Data Legis--Legis--Legislative package hearing  
13 that we're having here today. And as the former  
14 chair of this committee, I want to commend you also  
15 for all of the leadership you're providing in this  
16 committee. Good afternoon, Chair Vacca, council  
17 members and advocates. Thank you for hearing this  
18 important bill today. My bill inter--Intro 890 calls  
19 for the creation and preservation of archives of  
20 public data on permits, licenses, and performance  
21 evaluations in New York City's Open Data Portal, and  
22 our increasingly complex and technological society.  
23 Expanding access to public data is a critical part of  
24 strengthening democracy. Opening--opening data and  
25 structuring it to improve public access is allowing

2 many government agencies around the United States and  
3 the world to improve efficiency and get solutions to  
4 long--longstanding problems. This is also helping to  
5 create more trust and transparency between citizens  
6 and government. In fact, efforts to increase access  
7 to open data in the United States have developed into  
8 a movement that has resulted in increased information  
9 on everything from a vacation to housing to public  
10 safety. City of Oakland's Open Budget--Oakland  
11 Initiative is a great example how open data can  
12 improve city government. The result of opening up  
13 its budget data to a group of computer scientists  
14 resulted in more efficient budget making policy, and  
15 eased some of the frustrations of citizens. It  
16 helped the city find ways to save money, and also  
17 preventing protests on budget cuts. The Open Oakland  
18 Citizen Coding Initiative that evolved from that  
19 experience had gone on to develop apps indicating why  
20 the properties and the status for emergency planning  
21 and for helping parents to learn about public  
22 schools, spend the money allotted to them for  
23 education improvement. Other examples abound. Intro  
24 890 in particular seeks to address the current  
25 problem of NYC's Open Data Portal. Right now the

2 portal has data on individuals and businesses who are  
3 current with licenses, permits and other documents.  
4 But we have no way to know when individuals or  
5 businesses go off that list. We cannot do a  
6 longitudinal study because the data--data from even a  
7 few years back is getting lost, which makes it  
8 impossible to track trends. The bill calls for the  
9 creation a snapshot of data that allows us to track  
10 trends over time. So for example we will have access  
11 not only to DCA's list of caring businesses and  
12 operation, but also an idea on how many businesses  
13 went out of business from year to year. The second  
14 problem is that this bill seeks--that seeks to  
15 address is that some of the data sets are so massive  
16 that they become unwieldy and take too long to down  
17 low--download. The Yellow Taxi's GPS data is the  
18 largest data set on the portal. There are over 170  
19 million entries in a single year. What's the plan  
20 when there is ten years' worth of data? We want to  
21 be sure that there are protocols in place to divide  
22 up this data into smaller data sets rather than  
23 removing the data set, which is the common practice.  
24 Data getting lost is always correctable. Ensuring  
25 the creation of preservation archives is particularly

2 important as we examine historical trends and look  
3 back to determine how best to move forward. I'm  
4 excited to be part of this movement, and I hope the  
5 Council will recognize the importance of my bill,  
6 Intro 890, and pass it into law. Thank you for your  
7 time and attention, Mr. Chair. Thank you so much.

8 CHAIRPERSON VACCA: Thank you. Thank you  
9 very much, Council Member Cabrera. I know want to  
10 call on the sponsor of another bill, Council Member  
11 Vincent Gentile.

12 COUNCIL MEMBER GENTILE: Good afternoon,  
13 Mr. Chair, and good afternoon to all of those with us  
14 today. I want to thank Chair Vacca and the members  
15 of the Technology Committee for holding this  
16 important hearing on this important set of bills. My  
17 intro, Intro 898 seems like such a simple bill, but  
18 it really isn't quite so simple and nether are the  
19 terms posted on the New York City Open Data Portal.  
20 Scientific and technical terminology is presented to  
21 the public with no explanation, which for all intents  
22 and purposes is now transparency. So in order to  
23 take the information we show past the false  
24 transparency hidden and unnecessary complication, I  
25 have introduced and sponsored Intro 898. This bill

2 we hear today will create an online dictionary of all  
3 the terms on the Open Data Portal, the general public  
4 cannot reasonably be expected to understand. It  
5 would also requiring posting of the descriptions of  
6 the relative importance of these terms as well as the  
7 measurements and data that accompany them. So I look  
8 forward to taking New York City government with all  
9 of us her today forward in its duty to be more open,  
10 and more transparent with its citizens. Thank you,  
11 Mr. Chairman.

12 CHAIRPERSON VACCA: Thank you, Council  
13 Member Gentile. I have to ask the witnesses--I have  
14 to swear you in. I'm swearing at you. I'm just  
15 going to ask you to be sworn in.

16 DR. MASHARIKI: Okay.

17 CHAIRPERSON VACCA: [laughter] I think  
18 that comes later. Do you affirm to tell the truth,  
19 the whole truth, [laughs]. I've been to a lot of  
20 these hearings over the years. Do you affirm to tell  
21 the truth, the whole truth, and nothing but the truth  
22 in your testimony before this committee, and to  
23 respond honestly to Council Member's questions?

24 DR. MASHARIKI: Yes. Yes.

2 CHAIRPERSON VACCA: Okay. I know you  
3 would anyway. Okay, would you like to introduce  
4 yourselves, and whoever you decide can start off.  
5 Go.

6 DR. MASHARIKI: [coughs] Good afternoon.  
7 My name is Dr. Amen Ra Mashariki. I'm the Chief  
8 Analytics Officer, the Director of the Mayor's Office  
9 of Data Analytics as well as the Chief Open Platform  
10 Officer for the City of New York. Here with me is  
11 Albert Webber of the Department of Information  
12 Technology and Telecommunications Office of Advanced  
13 Technologies. The Mayor's Office of Data Analytics  
14 also known as MODA is housed within the Mayor's  
15 Office of Operations, and works in partnership with  
16 Anne Roest, the City's Chief Information Officer and  
17 DOITT Commissioner, to fulfill the requirements of  
18 the city's Open Data Law, which was enacted as Local  
19 Law 11 in 2012. The Administration fully supports  
20 the law, and works tirelessly to ensure it is fully  
21 implemented.

22 The promise of open data is transparent,  
23 free and accessible data. Data is more than just  
24 numbers. It's information that can create new  
25 opportunities and level playing fields for New

2 Yorkers. It illuminates issues to invite changes in  
3 frameworks, and leads to insights that turn  
4 impenetrable issues into solvable problems. It's the  
5 invitation for more voices to join critical  
6 conversations. I'm here today to testify on the  
7 progress of the implementation of the Open Data Law  
8 via our reinvigorated interactions with agencies in  
9 our open data vision and strategy as well as on the  
10 proposed pieces of legislation by the City Council  
11 designed to enhance the Open Data Law.

12 Over the past year, MODA and DOITT have  
13 been working energetically to engage agencies in  
14 continuing to open data sets to the public. To date,  
15 we have at least 1,386 data sets this year up from  
16 1,286 last years. This includes a number of notable  
17 of new sets like trip data from the Taxi and  
18 Limousine Commission, data from City Record Online  
19 and Pre-K for all data. Along these lines we have  
20 and will continue to work diligently with agencies to  
21 assist them in the leasing their data. New York City  
22 remains the gold standard in producing open data both  
23 nationally and internationally. Our ability to work  
24 with agencies to open the data to the public puts the  
25 city far and away ahead of all cities with respect to

2 the number of public data sets available on a portal.  
3 Our open data efforts have also earned New York City  
4 a top spot on the Open Knowledge Foundation's  
5 national rankings of open data cities. While we  
6 believe this in and of itself is a strong  
7 accomplishment, the Open Data Law is not just about  
8 releasing open data sets. As stated in the Open Data  
9 Law's Declaration of Legislative Findings and Intent,  
10 open data serves to, and I quote, "Streamline  
11 intergovernmental and intergovernmental communication  
12 and interoperability, permit the public to assist in  
13 identifying efficient solutions for government,  
14 promote innovative strategies for social progress;  
15 and create economic opportunities." MODA and DOITT  
16 are partnering to realize a vision and strategy that  
17 will allow us to achieve those stated outcomes of the  
18 law. With regards to that strategy and vision, Open  
19 Data FOIL. With this initiative, we have seen that  
20 open data is an invitation for anyone, anywhere at  
21 any time to engage with New York City government.  
22 The Open Data Plan, which was released on July 16 of  
23 2015 establishes the city's vision to deliver on the  
24 promise of the Open Data for All Initiative to the  
25 benefit of all New Yorkers. This is our vision of

2 what we can do, and what we aim to achieve with open  
3 data.

4           The ultimate success of the Open Data for  
5 All Initiative will not only be measured in the  
6 number of data sets that are published on the Open  
7 Data Portal, but in the number of New Yorkers who use  
8 open data in their daily lives. And that's not just  
9 the tech savvy New Yorkers. It's all New Yorkers in  
10 all five boroughs. To deliver on this promise we  
11 will continue to release new data sets and identify  
12 opportunities to provide data for all types of uses,  
13 focus on delivering quality data, enhancing the  
14 usability of the data portal, and expanding access to  
15 open data through use. We will also continue to  
16 improve and enhance the technical infrastructure that  
17 provides the foundation for the entire initiative,  
18 and increase responsiveness to and flexibility for  
19 user needs. With these goals in mind we are driving  
20 aggressively towards building an open data ecosystem  
21 that supports expanded access to open data, and  
22 better quality of data as well as enhanced portal  
23 usability.

24           With regards to expanding access, among  
25 the primary goals of the Open Data for All Initiative

2 is to meet users where they are. This year MODA and  
3 DOITT have begun a citywide Open Data Engagement  
4 Tour. This has included visits to CUNY campuses  
5 visits to borough meetings organized by the various  
6 borough presidents, online webinars, open data  
7 tutorials, and other community specific speaking  
8 engagements. We'll be gathering feedback from the  
9 Citywide Engagement Tour to determine how to best  
10 meet the needs of all New Yorkers.

11 Delivering Quality: We are also ensuring  
12 data quality through the development of applications  
13 such as the Business Atlas. The Business Atlas is a  
14 portal that was created through a partnership between  
15 the Mayor's Office of Data Analytics and the  
16 Department of Small Business Services. It uses open  
17 data and also partners with a small start-up here in  
18 a city called Placemeter to provide market research  
19 for local entrepreneurs.

20 Enhancing Usability is important to us as  
21 well. Earlier this summer we partnered with our open  
22 data platform provider to launch a new feature called  
23 Data Lens. Data Lens utilizes artificial  
24 intelligence technology in order to display data to  
25

2 everyday New Yorkers in a way that makes sense to  
3 them, and is visually intuitive.

4 Now with regards to the proposed bills.

5 The proposed bills amending the Open Data Law are all  
6 focused on our shared goal of improving the quality,  
7 usability and accessibility of open data. MODA and  
8 DOITT have long believed that open--that the Open  
9 Data Portal should be the primary location where  
10 agency data sets should be available, machine  
11 readable and accessible. We have engaged agencies  
12 with this in mind. We understand that in order to do  
13 this, many costly mechanisms and processes must be in  
14 place. We firmly believe in working towards the  
15 goals set forth in Intro 915 to accomplish this.

16 MODA and DOITT look forward to discussing with the  
17 Council how to best put a comprehensive plan in place  
18 to achieve this public data set updating capability.  
19 We believe that in order for us to truly achieve a  
20 ubiquitous open data community, MODA and DOITT must  
21 be responsive to users as well as provide quality  
22 customer service. Timely responses to public  
23 requests on the Open Data Portal are paramount to us  
24 achieving this level of customer service.

2 Intro 914 is very much in line with that  
3 goal--is very much in line with that goal, and would  
4 enhance our responsiveness to such requests, and MODA  
5 fully supports the intent of this legislation. MODA  
6 already instructs agencies to look at data that they  
7 have released in response to a FOIL request to decide  
8 whether the released data should be considered a  
9 public data set to be included on the Open Data  
10 Portal. Commissioner Toole of the Department of  
11 Records and Information Systems is currently building  
12 an open FOIL system. We have discussed building a  
13 data exchange with the system that feeds directly  
14 into the Open Data Portal so that we can get  
15 automatic feeds of FOILED data. Such an approach  
16 would accomplish the goals set forth in Intro 908.

17 Enhancing the usability of the Open Data  
18 Portal is a key objective for MODA and DOITT.  
19 Ensuring that data dictionaries are available for  
20 each data set within the portal as envisioned in  
21 Intro 898 will enhance the user's ability to  
22 understand and use the data. This is central to what  
23 we are trying to accomplish when we say open data for  
24 all. We support the intent of this bill, and look  
25 forward to working with DOITT and the Council to

2 identify the most appropriate way to implement this  
3 capability. We believe that managing data standards  
4 throughout the portal is extremely important. We  
5 believe both high quality data dictionaries and the  
6 use of open data standards throughout the portal are  
7 achievable with agency and open data coordinated  
8 engagement.

9           Intro 900 is a great way to begin the  
10 conversation around citywide standardization of  
11 geospatial data, but would ultimately require  
12 processes that require the expertise of the full tech  
13 leadership of the city, intimate agency engagement,  
14 and the investment of significant resources. To  
15 ensure usability and access to all open data, we are  
16 going to add language to the Technology Language  
17 Standards Manual around archiving and maintaining old  
18 records instead of purging them after a certain time  
19 period.

20           Intro 890 similar--similarly seeks to  
21 accomplish the strategy for the preservation of  
22 historical records. We must maintain a balance  
23 between keeping a clean easy to use portal preserving  
24 important historical data, and the cost of storing

2 significant amounts of data that may have outlived  
3 their usefulness.

4           With respect to Intro 916, we have legal  
5 and practical concerns with the audit proposed by the  
6 bill. In particular, it appears that the bill would  
7 authorize the Commissioner of the Department of  
8 Investigation to inspect all of an agency's data  
9 records and information including sensitive  
10 confidential and privileged information in order to  
11 determine whether the information comprises a public  
12 data set. Although we support the principle of  
13 developing mechanisms to ensure compliance with the  
14 Open Data Law, we do not think the proposed audit is  
15 the best way to achieve that goal. We are open to  
16 discussing alternative means for advancing agency  
17 compliance and accountability.

18           I would like to thank the Council  
19 Technology Committee for providing the opportunity to  
20 testify today, and for your continued support of open  
21 data. At this point, I'd like to turn it over to  
22 Albert Webber from the Department of Information  
23 Technology--Technology and Telecommunications to walk  
24 you through some of the finer details of implementing  
25 Open Data Law and the new enhancements we have

2 released in response to feedback from the community.  
3 After Mr. Webber's testimony, we will be happy to  
4 answer any questions you have.

5 CHAIRPERSON VACCA: Just let me introduce  
6 members that have arrived. Councilman Steve Matteo  
7 and Councilman David Greenfield. Please proceed.

8 ALBERT WEBBER: Thank Amen and good  
9 afternoon Chairman Vacca and members of the City  
10 Council Committee on Technology. My name is Albert  
11 Webber, and I'm pleased to speak with you today about  
12 many of the improvements made to the NYC Open Data  
13 Portal since we last testified. To date, we have  
14 unlocked nearly 1,400 data sets and maps via the NYC  
15 Open Data Portal sourced from more than 80 city  
16 agencies and entities. Additionally, as part of the  
17 annual July 15th update to the NYC Open Data Plan,  
18 city agencies identified an additional 282 data sets  
19 to be released through December of 2018. We also  
20 work closely with the Mayor's Office of Data  
21 Analytics, MODA, and agencies to continue identifying  
22 data to be released or listed in the Open Data Plan  
23 beyond the yearly July update, and we make monthly  
24 updates to the plan reflecting those updates. To  
25 date, we have automated the refreshing of

2 approximately 120 data sets to ensure their timely  
3 and accurate delivery. Many of the 120 data sets  
4 automated were in high public demand, including 311  
5 service requests, NYPD motor vehicle collisions and  
6 DOHMH restaurant inspections. This year alone 18 new  
7 data sets have been automated including Department of  
8 Sanitation dial-up (sic) vehicles, Taxi and Limousine  
9 Commission for-hire vehicles and active drivers,  
10 Department of Citywide Administrative Services City  
11 Record One, Green Book, and City Store. Automation  
12 aside from the initial work done to implement the  
13 process means that no human intervention is necessary  
14 for a data set to be refreshed. Additionally, and  
15 particularly topical given one piece of introduced  
16 legislation we recently completed a proof of concept  
17 that should allow us to begin the automated delivery  
18 of geospatial data sets to the NYC Open Data Portal.

19           With regard to the 1,200+ data sets that  
20 are not automated, we work closely with agencies and  
21 their open data coordinators to manually refresh  
22 them. This process is often time consuming and labor  
23 intensive, but our team diligently works to ensure  
24 timely data. We are also undergoing an effort to  
25 consolidate data sets where possible. There are a

2 number of data sets broken down by year, borough and  
3 other attributes making it difficult for users to  
4 export a holistic view of the data they desire. By  
5 consolidating data sets, we can allow users to more  
6 easily locate and download data improving the overall  
7 user experience. This consolidation effort may  
8 decrease our data set totals, but in keeping with the  
9 themes of open data for all, that Dr. Masharihki just  
10 discussed, this approach will in the long run  
11 facilitate fuller access to data sets and maps for  
12 our many users. In regards to new additions, as we  
13 continue in our efforts to make data sets more  
14 usable, we also further the march toward opening new  
15 data. Since our last update the following new data  
16 sets have been made available: Universal Pre-K data,  
17 Workforce 1 job listings and recruitment events,  
18 civil list data, and TLC trip data. Data sets  
19 scheduled to be made available by the end of the year  
20 include bridge ratings, reportable disease  
21 surveillance data, events and cleanup costs, and  
22 school budget overviews.

23                   Additionally, earlier this summer we  
24 partnered with our open data platform providers to  
25 launch a new feature called Data Lens. Data Lens is

2 a machine learning enabled tool that allows for the  
3 creation of auto generated charts, graphs and search  
4 features so that users can easily view and interpret  
5 sets via data scientists or everyday New Yorkers. In  
6 short, this tool helps us to make the reading of  
7 large data sets less intimidating and more accessible  
8 to a broader audience. These ongoing efforts  
9 encapsulate the spirit of Open Data for All  
10 Initiative Dr. Mashariki discussed earlier, and  
11 reflect the de Blasio Administration's commitment to  
12 keeping New York City at the vanguard of the open  
13 data movement. As we continue this work with the  
14 continued support and feedback from the city's  
15 vibrant civic technology community has proven  
16 invaluable as has the support led by Chairman Vacca  
17 and the Technology Committee of the City Council.

18 A number of the bills introduced last  
19 month align very closely with the aims of Open Data  
20 for All, making open data more usable, more  
21 accessible to New Yorkers of all stripes and in every  
22 corner of the city. Some of these bills mandates  
23 efforts already underway while some others fit  
24 squarely with the de Blasio Administration's Five  
25 Borough Strategy to enhance the Open Data Initiative.

2 Taking a brief look at each in turn Intro  
3 914 would establish response timelines for public  
4 requests in the NYC Open Data Portal, and help set  
5 user expectations about when the data they have  
6 requested can be made available within the context of  
7 the city's larger open data plan. We wholeheartedly  
8 support the intent of Intro 914 and have no  
9 significant concerns about this proposal. As you  
10 know, as part of the annual update, last July we  
11 committed to updating the plan on a monthly basis and  
12 regular feedback on a request for new data sets fits  
13 into this increased engagement strategy.

14 Intros 890, 898 and 915 embody proposals  
15 that speak directly to what Open Data for All is all  
16 about, and we look forward to working with the  
17 Council on each to explore ways that our shared goals  
18 may be realized. Intro 890 relates to the retention  
19 of archival data on the NYC Open Data Portal, and  
20 rightly references the Technical Standards Manual as  
21 a place for the appropriate language to reside. To  
22 the extent that all data sets are removed from the  
23 portal without explanation, this proposal is keeping  
24 with Open Data for All. But mandating that all data  
25 sets replaced by those missing period (sic) that are

2 more accurate and complete, have to remain on the  
3 portal indefinitely is problematic. This can be  
4 confusing to visitors and runs counter to the city  
5 data making city data more understandable and usable  
6 for all. There will also be some costs for  
7 additional storage capacity as data sets like some of  
8 those we mentioned earlier containing millions of  
9 rows of data become increasingly large and complex,  
10 but we look forward to discussing with the Council a  
11 path on data set retention.

12           Intro 898 requires every data set on the  
13 NYC Open Data Portal be accompanied by a plain  
14 language data dictionary and it's aimed at exactly  
15 the right place, making what can be very bureaucratic  
16 more usable and understandable to the lay person.  
17 Data dictionaries and supporting data set  
18 documentation were at one point common on the Open  
19 Data Portal, and can be a valuable standard to more  
20 formally implement. Of course, the role of  
21 providing--the role of providing headings explaining  
22 columns and defining the technical terms and acronyms  
23 included in data sets needs to be completed the  
24 agency submitting their data to do it in MODA for  
25 posting and automation not by DOITT or MODA directly.

2 And we, therefore, agree with this inclusion as a  
3 requirement that can be made more specific as part of  
4 the upcoming Technical Standards Manual refresh. We  
5 look forward to discussing with the Council the most  
6 effective means of implementing data dictionaries.

7 Intro 915 requires that public data sets  
8 updated on other city agency websites be posted to  
9 the NYC Open Data Portal within three days. This  
10 would help ensure continuity and consistency of user  
11 experience when searching for official New York City  
12 data, which is directly in line with both the  
13 original intent of the city's Open Data Law as well  
14 as with Open Data FOIL. While DOITT is constantly  
15 working with agencies to automate data sets, there  
16 still remains work to be done. There will be some  
17 costs and resource considerations to be addressed in  
18 pursuing this effort, but we share with the Council  
19 the goal of keeping city data accurate, updated and  
20 consistent across all channels, and we look forward  
21 to working with you in making that vision a reality.

22 The next proposal, Intro 908 establishes  
23 a new process for making agency FOIL responses public  
24 accessible. The spirit of this proposal squarely  
25 aligns with the de Blasio Administration's commitment

2 to the promise of transparent, free and accessible  
3 open data. As part of that promise, we are committed  
4 to ensuring that agency responses to FOIL requests  
5 are widely accessible to the public. To that end,  
6 the Department of Records and Information Services is  
7 currently developing an open FOIL portal that will  
8 launch by the end of this year. This centralized  
9 website will empower the public to easily submit and  
10 track FOIL requests for all city agencies as well as  
11 access the vast majority of records responsive to  
12 previously filed FOIL requests even where there is no  
13 accompanying data set. Such proactive disclosure  
14 through a centralized repository will dramatically  
15 expand public access to government records.  
16 Therefore, rather than duplicating work to be  
17 completed in the coming months, we are happy to  
18 explore crafting a technical solution whereby the  
19 records that will be posted on open FOIL are  
20 automatically transferred to the NYC Open Data  
21 Portal.

22                   The final two proposals Intros 900 and  
23 916 have elements, which we agree, but also present  
24 significant staffing costs, timing or legal concerns.  
25 Intro 900 requires every data set on the NYC Open

2 Data Portal that contains address information to  
3 utilize a standard field layout and presentation of  
4 that address information, including corresponding  
5 geospatial reference data. Having a single city  
6 standard for addressing geospatial information data  
7 is ideal, but we do not believe it is a goal best  
8 issued through open data legislation. While open  
9 data would benefit from such a standard, the standard  
10 itself should be developed via a larger collaborative  
11 process whereby agencies, advocates and users can  
12 work toward a generally agreed upon acceptable  
13 standard for addressing geospatial data. Once  
14 established, the systems utilized by agency data  
15 bases feeding the NYC Open Data Portal would adhere  
16 to that standard when producing data for publication.  
17 Older legacy systems once replaced would then also be  
18 required to incorporate those standards. This is a  
19 worthwhile effort, but one that will require  
20 additional resources and time substantially more so  
21 than the 90 days proposed to implement effectively.  
22 In the interim, we continue to make all the relevant  
23 data tabular, shaped files, et cetera, available in  
24 the open data portal itself for interested  
25 individuals to use according to data sets they are

2 most interested in. This, indeed, is one of the  
3 earliest, promises of open data provided to the  
4 public and let them do with it what they will without  
5 further interference.

6 Intro 916 would require the Commissioner  
7 of Investigation to conduct audits of certain agency  
8 compliance with open data law. While we and all  
9 public officials can always do a better job of  
10 serving New Yorkers, the city's Open Data Law is  
11 among the most ambitious comprehensive policies of  
12 its kind in the world, and has set a model for many  
13 governments to follow. We are happy to work on this  
14 continued implementation, and are committed to do  
15 even better still. There are a number of steps  
16 between making the accountability improvements that  
17 Dr. Marshariki mentioned. For example, more clearly  
18 stating on the portal which data sets are either  
19 removed or delayed with detailed explanations, and  
20 requiring the Commissioner of Investigation to  
21 conduct audits in an area where the city has  
22 demonstrated such noteworthy transformational success  
23 in just over three years. More significantly, we are  
24 concerned that the proposed audits would require the  
25 Commissioner of Investigation to inspect all of an

2 agency's data including even sensitive, confidential  
3 or privileged information in order to determine the  
4 existence of public data sets.

5           We have legal concerns about the  
6 Commissioner's authority to access such information  
7 as well as concerns about the utility of such audits  
8 generally in advancing open data goals. Thank you  
9 for the opportunity to testify this afternoon. Open  
10 data has been and remains a priority of the de Blasio  
11 Administration, and we thank our partners in the City  
12 Council and across the civic, tech and advocacy  
13 communities for their continued efforts in making New  
14 York City's successful implementation even better.  
15 This concludes our prepared testimony, and we look  
16 forward to answering your questions. Thank you.

17           CHAIRPERSON VACCA: Thank you. Perhaps I  
18 should start with you, sir, first, and your testimony  
19 and--your testimony as well, sir, talks about being  
20 supportive of most of the legislation. You see  
21 problems with 916, which I can go into, but you used  
22 the word that you support the intent. You didn't say  
23 you support it as written. Now, I appreciate that  
24 you support the intent, but do you support the  
25 legislation as written in any case? Is there any--is

2 there any piece of legislation you can say now that  
3 you support, the Administration supports?

4 DR. MASHARIKI: Yes, we do (coughs) um,  
5 the--the--the legislation around--[background  
6 comments]--Intro 914 we--we support as written with  
7 response to users to the portal. We support that as  
8 written. Also, with respects to legislation 898, we  
9 support. We do believe that there should be some  
10 discussions around how best to implement it, but we--  
11 we support that as well.

12 CHAIRPERSON VACCA: Okay. So you support  
13 one piece of legislation. On the second piece you'd  
14 like further discussion, but you support the concept,  
15 and the others you support the intent?

16 DR. MASHARIKI: That's right. We believe  
17 that there should be more conversation around  
18 implementation, their costs, resources, challenges to  
19 implement--in implementing these things, and we would  
20 like to have discussions with City Council with  
21 regards to that.

22 CHAIRPERSON VACCA: Okay. I did want to  
23 note regarding my legislation concerning DOI having  
24 the power to audit compliance, you did have some  
25 concerns, but most of your concerns revolve around

2 DOI possessing confidential information or privileged  
3 information that is sensitive in nature. And I have  
4 to say that that is the job of DOI every day of the  
5 week. I think DOI is the agency that probably isn't  
6 trusting--entrusted to possess that information when  
7 it arises they need it as part of an ongoing probe.  
8 So, I think to say that we don't like the fact that  
9 DOI is going to be looking at sensitive information.  
10 I think saying that is inconsistent with the mission  
11 or what most people expect the DOI to do. I also  
12 have to note very honestly that some of the things  
13 that are rated or that are classified as private--  
14 private or sensitive, make no sense to me. Two of  
15 the DOT data stats were removed from the Open Data  
16 Plan. One was a traffic signal defect response time,  
17 and the other was a pothole work order. And both of  
18 those were removed because of privacy concerns and  
19 sensitivity considerations. I don't know what's  
20 private about a pothole being fixed. That doesn't  
21 seem to me to rate as something that would be  
22 classified as private or a traffic signal defect  
23 response time. So, I say that to you because you  
24 again bring that up in relation to my piece of  
25 legislation, but in the Open Data Law there are not

2 exemptions made for privacy or sensitivity. It's not  
3 one of the reasons in the law why a data set can be  
4 excluded. So based on that, I'd like you to  
5 reconsider your--your view on this.

6 DR. MASHARIKI: Well, I'll respond by  
7 saying I think the conversation that you're starting  
8 is the conversation we would absolutely like to have.  
9 I think the agencies certainly identify data sets  
10 that they believe have sensitivities or privacy  
11 concerns around them, and we work with the agencies  
12 around it. We don't--we do--when we do speak to  
13 agencies, we do say a data set as is may have some  
14 sensitive information. We want to understand how  
15 some of that information cannot be included on the  
16 portal, but still release that information.

17 CHAIRPERSON VACCA: [interposing] Right.

18 DR. MASHARIKI: So we have those  
19 conversations all the time.

20 CHAIRPERSON VACCA: I understand, but if  
21 you have concerns about DOI when they do their  
22 audits, if you have concerns about DOI doing an audit  
23 and you have information that you think DOI should  
24 know should not be on open data, then you tell them  
25 it should not be on open data. It's not covered by

2 the law, or there is a sensitivity issue of some  
3 type. I mean I'm willing to let DOI be the arbiter.  
4 We need an agency to guarantee compliance and  
5 accountability with the law. So, otherwise--I mean  
6 do you support the right of people to sue if open  
7 data requirements are not met because we--we  
8 certainly can go that way. Some of the advocates  
9 have mentioned that, but I would like to have  
10 accountability in a different way and I--that's why I  
11 mentioned DOI. And it's not a traditional  
12 investigation as DOI often conducts. It's a  
13 compliance, an auditing aspect. Someone has to let  
14 the public know that this is being done in a timely  
15 way. Someone has to be accountable, and someone has  
16 to make a report. So, that's my view on that, and I  
17 hope you consider that in reviewing your objection to  
18 that piece of legislation.

19 I also was concerned about something  
20 else. There's a Conflict of Interest Board data set  
21 listing all of the policy makers and city agencies,  
22 and that was removed from the plan under a claim that  
23 the information was provided by other agencies  
24 despite no other agency having such a list compiled  
25 in one place. Now, open data has that opportunity to

2 be more transparent, and I think it is important for  
3 this data set to be specifically provided. Because  
4 people who work with the City of New York have an  
5 obligation to know whether or not they're a mandated  
6 reporter, whether or not they must file a COIB report  
7 or financial disclosure report. So I look at that,  
8 and even if it's removal is permitted under the law,  
9 it should voluntarily be put on the site. Sir?

10 ALBERT WEBBER: So as is the list of  
11 policy makers are compiled by all city agencies, and  
12 I think the data is compiled in various formats, and  
13 sent to the COIB. So that data is made available  
14 outside of the open data portal, but that process of  
15 taking that data, those PDFs, those Excel  
16 spreadsheets however they may be compiled, that--the  
17 data exists elsewhere, but we're not sure that that  
18 data as is, as compiled belongs on the Open Data  
19 Portal.

20 CHAIRPERSON VACCA: Okay. Perhaps you  
21 could further explore that. I'd like to see if there  
22 was a way we could include that. I think it's  
23 important. It is all over the place right now. It's  
24 not cohesive. I wanted to ask you in, um, go ahead.  
25 [background comments] I wanted to ask you also a

2 couple of things. Legionnaires' Disease is becoming  
3 an increasing concern in our city, and I wanted to  
4 know [coughs] what we can do to give cases of  
5 Legionnaires' Disease any priority when it comes to  
6 open data, getting additional relevant information  
7 from DOH to post on open data.

8 DR. MASHARIKI: think that starts with a  
9 conversation reaching out to the Commissioner of  
10 DOHMH looking through the data that they do have and  
11 understanding data could be shared with the portal.  
12 We have relationships DOHMH. hey share their data  
13 with the portal now on automated feeds as well. So  
14 we could replicate that, but it starts with a  
15 conversation with them.

16 CHAIRPERSON VACCA: Okay, who has that  
17 conversation?

18 DR. MASHARIKI: MODA.

19 CHAIRPERSON VACCA: MODA. Okay, it's with  
20 MODA. Okay, timeliness of posting information. I  
21 brought this up at previous hearings where I noticed  
22 that some city websites have the information, and  
23 then Open Data does not have it until some time  
24 after--afterward. What priority are we giving to  
25 timeliness of posting information?

2 ALBERT WEBBER: Timing it is a huge  
3 priority of ours, and we have over 120 automated data  
4 sets on that Open Data Portal. We work closely with  
5 open data coordinators around the city to ensure that  
6 data is maintained and refreshed on a timely basis.  
7 For the data sets that aren't automated, our team at  
8 DOITT and MODA work to manually do those updates to  
9 the Open Data Portal. It's a very time consuming  
10 process. That's why we try to set up as many  
11 automations as possible, but timely data, automatic  
12 refreshes of data are goals of ours and a big  
13 priority. [pause]

14 CHAIRPERSON VACCA: What do you need to  
15 get automation in place?

16 ALBERT WEBBER: Um, it's--it's--it's  
17 mostly like Amen said the initial conversation with  
18 the agency, um, and then just also our technical  
19 resources that back and forth, that exchange. Um, it  
20 also depends on how we're going about doing that  
21 automation. So it's really just opening the  
22 dialogue, and then having our technical teams speak  
23 with their technical teams.

24 DR. MASHARIKI: If you don't mind, um,  
25 there's two things that--one is it's case-by-case per

2 agency. So when we automate feeds from an agency to  
3 do its platform it--the ease at which we can do that  
4 depends on the--the status or the space that the  
5 agency's information is in. So obviously, the newer  
6 systems that becomes a less cumbersome process. So  
7 for one, to answer your question, it is a case-by-  
8 case. So it's not a general sort of planning  
9 strategies. It's case-by-case based on what they  
10 have. Two, is we do when we've investigated  
11 automating and--and upping the response time, we  
12 decided that we wanted to build sort of an enterprise  
13 architecture to not only solve instances for DOB,  
14 HPD, 311 and others, but build an architecture so  
15 that as we brought agencies on line for more feeds  
16 that capability existed for everyone else. So it's  
17 also thinking more enterprise as well.

18 CHAIRPERSON VACCA: Now, I noted on the  
19 portal in the section where the public can request  
20 new data sets, there have been around 170 requests in  
21 four years. Of those 170, six have been approved;  
22 seven have been rejected, and the rest are marked  
23 open. When will all these requests be replied to?  
24 Why do we face a situation like that?

2           ALBERT WEBBER: So the way the---the data  
3 set nominations work on the Open Data Portal we can--  
4 we moderate those requests before they come in,  
5 before we actually open them up for the public to see  
6 them. We just want to ensure that the requests are,  
7 you know, no foul language and things of that sort.  
8 So we are working with agencies every time, and we're  
9 doing more updates to the plan, but we provide those  
10 data set nominations to agencies so that they can see  
11 the data that's being requested, and they can add  
12 them to the portal. So the fact that we moderate  
13 those data sets, and we open them up, that's--that's  
14 pretty much identifying that we see that the  
15 nomination came in, and that we are working with the  
16 agencies to put that data on the Open Data Portal.

17           DR. MASHARIKI: I'd also like to add, as  
18 the Director of MODA I started last year. This is  
19 the first time I've been a part of this plan and this  
20 strategy to release the plan July 15th, and one of  
21 the things, the new things that hasn't been in  
22 existence before that we've implemented is a monthly  
23 meeting with the open data coordinators and monthly  
24 and updates. And so now we have calls, monthly calls  
25 with the open data coordinators for agencies, and in

2 those calls we will be bringing on a consistent basis  
3 these requests. Whereas before it was more of an  
4 asynchronous, now it's going to be a consistent  
5 process to engage the agencies.

6 CHAIRPERSON VACCA: You are satisfied  
7 that every agency has an open data person, liaison  
8 that you can relate to?

9 DR. MASHARIKI: Um, not every, but a lot  
10 of the--the larger agencies.

11 CHAIRPERSON VACCA: Okay. So if not  
12 every agency, why not, number one. And number two,  
13 what is your power to make sure that agencies do?

14 DR. MASHARIKI: Well, we--we are ensuring  
15 that everyone does, and we--we communicate and we  
16 reach out to not every agency. What I mean by not  
17 every agency, they don't, um, officially appoint an  
18 open data coordinator, but we have had communications  
19 and we engage with people at agencies. A lot of  
20 times it's resource availability and so on and so  
21 forth.

22 CHAIRPERSON VACCA: Well, this is  
23 something that has to come basically I think in this  
24 case from the Mayor's Office. The Mayor's Office has  
25 to tell agencies that there should be a specific

2 person doing open data. Now, it doesn't have to be a  
3 person who does open data--open data and nothing else  
4 because if the agency is small, they commissioner  
5 should have discretion on--on assigning other  
6 responsibilities. But this is something that the  
7 Mayor's Office has to make a priority namely every  
8 agency has to have a person doing open data. It's  
9 the law and you don't have the power to tell agencies  
10 what to do. You can request, you can ask, but if  
11 they don't want to do it--Are there agencies that are  
12 also giving you problems when it comes to posting  
13 information on open data? Have agencies been  
14 resistant to posting what you've determined they  
15 should be posting based on the law?

16 ALBERT WEBBER: I think agencies for the  
17 most part have been very cooperative. You know, in  
18 the early days of open data I think there was a  
19 culture change just the--just the shift to--to  
20 putting that data out there. But for the most part, I  
21 think agencies have been very cooperative. It's come  
22 to the point where now we have a team that's doing a  
23 lot of automation with the publish--with our  
24 publishing data that we have to--I'm not saying that  
25 we're turning away data, but we have to push agencies

2 a little further back in the queue because of the  
3 demand for agencies to get data on the portal. So  
4 I'd say for the most part agencies are--are very  
5 cooperative.

6 CHAIRPERSON VACCA: Okay, and now do you  
7 see my point about there being an agency in the City  
8 of New York that is above approach that can make sure  
9 that there is compliance. That's my point. Here we  
10 are going back and forth, and you're counting on  
11 voluntary cooperation from other agencies. You've  
12 really been in a position where you probably should  
13 have asked the Mayor's Office some time ago to  
14 intervene, and that request probably was not made. So  
15 therefore we're in a situation where not every agency  
16 has an open data person. So I want to make sure that  
17 the law is complied out and, your--your inability to  
18 tell other agencies what to do means that you cannot  
19 ensure compliance.

20 ALBERT WEBBER: With regards to open data  
21 coordinators, (sic) I'd also like to point out that--

22 CHAIRPERSON VACCA: [interposing] You--  
23 you have to depend on good will and compliance, and  
24 you know what it is with some commissioners, they  
25 think that their agencies is their own bailiwick.

2 Not just in this Administration. I'm talking  
3 historically. You have commissioners who think that  
4 their agency is their agency, their bailiwick, that  
5 they're going to make decisions, and there are ways  
6 to implement the law and there are other ways to  
7 implement the law. Well, no, the public has a right  
8 to know. We have specific legislation we enacted in  
9 the Council, and there has to be a check. There has  
10 to be the accountability aspect.

11 ALBERT WEBBER: I mean with regards to  
12 the--the open data coordinators and the positions  
13 that are there in agencies, every agency may not have  
14 an open data coordinator but most do. And in some  
15 cases where we have the agencies where an ODC hasn't  
16 been officially appointed, we have multiple points of  
17 contact who--who really serve in that role of being  
18 that ODC who--who plays that position. So the ODC  
19 may not exist, you know, by name, but there's a  
20 person there that does that similar work. They just  
21 may not have been officially appointed as that open  
22 data coordinator.

23 CHAIRPERSON VACCA: Uh-huh. I--I think it  
24 says something to the people of New York City that  
25 there is a person that's known as the open data

2 person. So I would like that. I understand your  
3 point certainly. Okay. I think you and do we have  
4 questions from any of the members?

5 COUNCIL MEMBER GENTILE: [off mic] I do.

6 CHAIRPERSON VACCA: Okay. Council Member  
7 Gentile and then Council Member Cabrera.

8 COUNCIL MEMBER GENTILE: Thank you, Mr.  
9 Chairman, and thank you for your--both of you for  
10 your testimony. Mr. Webber, first of all, I'm glad  
11 to hear that you've both agree that 898 and data  
12 dictionaries is a good idea, and--and you approve of  
13 it the way it is written. But I'm curious. In your  
14 testimony, Mr. Webber, you said that--that it's  
15 really up to the agencies to submit their data and  
16 their definitions of technical terms to MODA and  
17 DOITT. It's not MODA and DOITT's job to do that.  
18 So, should--and this almost follows up on Council  
19 Member Vacca's question. Should they fail to do that  
20 it sounds like you throw up your hands and say that's  
21 it. We can't--we can't force them to do it.

22 ALBERT WEBBER: Well, I guess in terms of  
23 the data dictionaries, the reason we say it, we  
24 believe that they understand their data more than we  
25 would. So we would expect them to put together those

2 data dictionaries to explain to us what columns, and  
3 field names and records of that sort mean. We do  
4 feel that data dictionaries are useful. We do think  
5 that that's something that could go into the  
6 Technical Standards Manual that agency open data  
7 coordinators review, and it should be that standard  
8 when it comes to open data, that the data  
9 dictionaries are actually completed. So we do think  
10 the Technical Standards Manual would be a good place  
11 for that to reside.

12 COUNCIL MEMBER GENTILE: Because  
13 ultimately you, um, you're responsible for complying  
14 with the law, correct? So, and you mentioned,  
15 Doctor, about these monthly meetings with the agency  
16 open data representatives.

17 DR. MASHARIKI: Uh-huh.

18 COUNCIL MEMBER GENTILE: Is that an  
19 appropriate place where you can get this kind  
20 information should it be missing?

21 DR. MASHARIKI: Absolutely, absolutely  
22 and that's what--that's what we will do. And there's  
23 actually--we can put together subcommittees from  
24 those monthly discussions to discuss exactly the best  
25 way to automate data dictionaries. Data dictionaries

1 COMMITTEE ON TECHNOLOGY 49

2 doesn't have to be a laborious, onerous process. It  
3 can actually using technologies, can actually be  
4 quite simple, and agencies use them all the time. So  
5 this is not the first time that we've gone to an  
6 agency. I know under the work that the Mayor's  
7 Office of Data Analytics does, we require agencies to  
8 share data dictionaries with us when we get their  
9 data to do analytic work for them. So agencies are  
10 used to this, and so we can use that platform to  
11 engage them smartly about this.

12 COUNCIL MEMBER GENTILE: Great. Okay,  
13 thank you, Mr. Chair.

14 CHAIRPERSON VACCA: Council Member  
15 Cabrera.

16 COUNCIL MEMBER CABRERA: Thank you so  
17 much, Mr. Chair. Welcome. I have a couple of  
18 questions. I just want to know as a general policy  
19 do you agree that we want--we don't want data to be  
20 erased or removed from the Open Data Portal once  
21 posted?

22 ALBERT WEBBER: We have concerns about as  
23 data is growing just keeping records of all of the  
24 data that's in the Open Data Portal. We have feeds  
25 that vary in their refresh rates, daily, weekly,

2 monthly. So we do have some concerns about storage  
3 costs, about retention and things of that sort, but--

4 COUNCIL MEMBER CABRERA: So talk to me  
5 about the storage costs. How much are you talking  
6 about?

7 ALBERT WEBBER: We'd--we'd have to do a  
8 best cost analysis to see exactly what that would be,  
9 but we just released--

10 COUNCIL MEMBER CABRERA: [interposing]  
11 But you mentioned it as a concern.

12 ALBERT WEBBER: Yeah.

13 COUNCIL MEMBER CABRERA: So if you  
14 mention it as a concern that you tell me that  
15 somebody already evaluated what the cost analysis  
16 would have been while you're telling me that you guys  
17 are guessing that it's--I mean how much are we--  
18 what's the range here? What are--what are we talking  
19 about, or anybody who you brought here do they have  
20 the cost analysis?

21 ALBERT WEBBER: So, I mean we have our  
22 current contract in place with our current vendor.  
23 Um, but when it comes to these storage costs, I mean  
24 this is something that we have to do some real  
25 analysis on. Our data sets are getting larger and

2 larger. Our daily automations are capturing millions  
3 of new records every single day. So there are--there  
4 are--obviously there are concerns about, you know--

5 COUNCIL MEMBER CABRERA: [interposing]

6 But we knew this was going to happen. Remember when  
7 Council Member Gale Brewer and I both talked about it  
8 (sic) and I was the chair. We--we knew this back  
9 then. I hope somebody in the agency foresaw that  
10 this was going to happen, right? I mean this is not  
11 a new use that we were going to archive and it was  
12 going to be a greater cause. When we had last year  
13 \$3 billion surplus, we were trying to figure out  
14 where to put money. I mean did anybody request for  
15 more funding from the agency to--to make sure that we  
16 could have additional storage space, capacity?

17 ALBERT WEBBER: We have a number of open  
18 data lines that are available. With regards to our  
19 current, um, contract with our vendor, that's on a  
20 yearly basis as of right now. And we're not opposed  
21 to having archived data. You can go and look at it  
22 on the Open Data Portal. We do have a number of data  
23 sets that are archived. There are just concerns when  
24 you have that data that's refreshed as frequently as  
25 daily. Keeping those snapshots of ten plus reading

2 records of data. You know, by the end of the month  
3 you're looking at 30 million records of data, and  
4 that's from one data site. So those are where the  
5 concerns are.

6 COUNCIL MEMBER CABRERA: Do you have the  
7 capacity do it if you want to do it, and there's--  
8 there's people that can see the value of it.

9 ALBERT WEBBER: We--we--we'd have to put  
10 together a better plan in terms of like archiving our  
11 data.

12 COUNCIL MEMBER CABRERA: But--but when  
13 you do, I'm just wondering if you can answer the  
14 question, you do have the capacity to do it if you  
15 really want to do it, right?

16 ALBERT WEBBER: Can I do it overall? Yes,  
17 DOITT can do that? [laughs]

18 COUNCIL MEMBER CABRERA: Okay, so you can  
19 do that. Okay. So this is not an issue of capacity,  
20 this is an issue of whether you want to or not, you  
21 see a value to it?

22 ALBERT WEBBER: No, no, no.

23 COUNCIL MEMBER CABRERA: No? Go ahead  
24 and explain where--where you have the capacity to do  
25 it. DOITT has the capacity to do it. So tell me

2 where's the gap here. If--if you don't see--you see  
3 the value.

4 DR. MASHARIKI: If you don't mind, I can  
5 give you--

6 COUNCIL MEMBER CABRERA: [interposing]  
7 Please.

8 DR. MASHARIKI: --I can give you  
9 incidents.

10 COUNCIL MEMBER CABRERA: Please.

11 DR. MASHARIKI: We do see the value in  
12 archiving. We do archiving now. You mentioned Taxi  
13 and Limousine data.

14 COUNCIL MEMBER CABRERA: Right.

15 DR. MASHARIKI: We work diligently with  
16 our vendors to ensure that we could not only store,  
17 but also provide high performance to users coming  
18 onto the portal gaining access to that large data  
19 set. With respect to archiving, what I mentioned in  
20 the testimony was that we do have to strike a balance  
21 between a clean and usable portal. And so, we don't  
22 want to have this portal become some cumbersome and  
23 so unwieldy with data sets all over the place over a  
24 specific period of time. I'll give you an example.  
25 If an agency is providing data to us in an automatic

2 feed now, and let's say that agency just for  
3 operational reasons are not storing a particular data  
4 set, maybe three or four fields out of that data set,  
5 so they are no longer sending that to us because they  
6 are no longer tracking those--that data set. But we  
7 archive that and it remains on the portal as is, but  
8 now we're having a new feed with less fields. Then  
9 there is going to be some confusion in terms of users  
10 saying well why can't I get that data? I used to get  
11 it last year. I'm not getting it now so on and so  
12 forth. So we're not saying that that should stop the  
13 process of archiving. We're saying that we should  
14 have a fuller conversation really and work out  
15 exactly what the strategy is to ensure a clean usable  
16 and not a messy portal, but also smartly archive this  
17 information.

18 COUNCIL MEMBER CABRERA: You--you have a  
19 lot of smart people in your agency. Probably some of  
20 the brightest among agencies. Some agencies might  
21 debate that. I think you guys could figure that out.  
22 I think that whenever you have areas where you have  
23 scenarios like the ones you mentioned that--that  
24 somehow that's indicated that that information could  
25 be indicated. I just, you know, and obviously I'm

2 addressing the issue here of Intro 890. You know, I  
3 don't think costs should be an issue because it was  
4 mentioned. I just--I don't [coughs] it's shouldn't  
5 be an issue especially we don't know how much we're  
6 talking about. I think with the economy of scales at  
7 which you're working with and with new contracts, I  
8 don't think that should be a stumbling block to  
9 making Intro 890 a possibility. Let me ask you  
10 something related to Council Member Vacca's comment  
11 does--does every agent--is there--is there an agency  
12 that have an open data coordinator vacancy at this  
13 moment?

14 [pause]

15 ALBERT WEBBER: Probably. I think I have  
16 to get back to you on that one.

17 COUNCIL MEMBER CABRERA: Can you please?

18 ALBERT WEBBER: Sure.

19 COUNCIL MEMBER CABRERA: Okay. The  
20 second question is are there agencies that--this is  
21 kind of a dumb question, but are there--are there  
22 agencies that are like way behind providing, um, the  
23 info that you are requesting?

24 ALBERT WEBBER: Um, so what we do--we  
25 monitor everything based on that open data plan. You

2 can see the status of the open data plan on the open  
3 data dashboard. In terms of way behind, I would say  
4 no. There are agencies that looking at their dates  
5 for this current year may be a little behind, but  
6 we've had regular engagement with them to get that  
7 data.

8 COUNCIL MEMBER CABRERA: Well, how would  
9 you qualify a little behind?

10 ALBERT WEBBER: There are data sets that  
11 we're doing the last month where we're having  
12 conversations and we expect them up in the--in the  
13 next few weeks.

14 COUNCIL MEMBER CABRERA: Okay, do you  
15 have--do you guys have a rating system that your  
16 provide back to them and say hey this is how  
17 effective you have been? That way the commissioners  
18 can see. It's very possible the commission is doing  
19 it now and the information has been related late  
20 right?

21 ALBERT WEBBER: We do not have a rating  
22 system, the dashboard is the--is the method in which  
23 we monitor progress of the agencies.

24 COUNCIL MEMBER CABRERA: Yeah, but I'm  
25 sure the commissioners and the other agencies, I mean

2 they're not going there to see how--I mean, they're  
3 dealing with basically with the staff. Maybe it's  
4 something for you guys to consider that information,  
5 this type of information. It would be directly  
6 related to the commission and say look this is how  
7 effective your coordinator, you open data coordinator  
8 has been in relaying the information to us.

9 ALBERT WEBBER: Absolutely.

10 COUNCIL MEMBER CABRERA: Okay. Thank you  
11 so much and I hope you will reconsider your--again  
12 the value with the value of Intro 890. I think it's  
13 necessary. At the end of the day, we're going to be--  
14 --we're going to be confronted as to how much we're  
15 going to archive, what's going to be made available,  
16 how it's going to happen. I know it's a judgment  
17 call at one point, but I think the original intent of  
18 the law as I recall with Council Member Brewer and  
19 all of the hearings we had back then was to do  
20 exactly that. To make sure that as much information  
21 that was available is made available, and that we had  
22 the historical context to that. Thank you so much,  
23 Mr. Chair.

24 CHAIRPERSON VACCA: Thank you. I wanted  
25 to ask on Intro 915, what do you believe is the

2 appropriate update frequency for a data set when the  
3 data was also available immediately on the agency's  
4 website?

5 [background comments]

6 ALBERT WEBBER: Okay, yes. So I believe  
7 this is in reference to if data is made available  
8 somewhere aside from the Open Data Portal.

9 CHAIRPERSON VACCA: Right.

10 ALBERT WEBBER: Okay. Um, I--I think  
11 three days is a little bit aggressive. What that  
12 would entail is whether it be Web Master from  
13 NYC.gov, our open data team in working with those  
14 agencies. We'd have to get a better understanding  
15 of that process at other agencies tied into our  
16 process. Three days is a bit aggressive. We do  
17 think we would need more time, but in terms of  
18 keeping that data fresh and making sure it's  
19 consistent, we're definitely on board with that.

20 CHAIRPERSON VACCA: Well, if you don't  
21 agree to three days, there's got be an outside--there  
22 has to be a parameter and, um, it sends a message  
23 about how useful you view the information also if you  
24 don't have a time table. Then agencies will take it  
25 upon themselves. So your agency has to really

2 establish a time table for these agencies to comply  
3 with. You can just say three days is not enough, but  
4 that's okay, but then don't you want it in seven days  
5 or six day? I mean, there has to be a firm guideline  
6 so that we can implement them more successfully. So  
7 I'd like you to take that back. Okay. Oh,  
8 Councilman Kallos has joined us. Do you have  
9 questions? Oh, wait, let me ask Councilman Kallos to  
10 first, of course, speak on his bill and then if he  
11 has questions.

12 COUNCIL MEMBER KALLOS: I just want to  
13 drop by, and thank Council Member Vacca, Chair of the  
14 Technology Committee for his leadership on technology  
15 issues, for putting together an amazing package of  
16 open data bills to take the package--the Open Data  
17 Law to its next iteration. Um, that's what we do in  
18 the technology fields. We improve and improve and  
19 improve over and over again, and just want to thank  
20 him for his generosity. Not to show too much  
21 transparency, but sometimes billing legislators will  
22 work and do all the work themselves and create a  
23 package of legislation, share them with their  
24 colleagues based on their interests. And so I'm  
25 grateful pairing the GIS information bill. I

2 understand that there are some concerns with it, but  
3 ultimately as long as we are sticking with the  
4 geospatial recognized by the consortium whatever the  
5 standard is we should just have it and adopt it.  
6 Camel(sic) is one that is easy to use and plain text  
7 accessible. Which might be one especially for the  
8 open data sets because otherwise you need to have  
9 references to external data sets. So Camel is one  
10 that I would recommend, but either way, having a GIS  
11 standard that uses longitude, latitude as well as a  
12 geospatial markup would be amazing. And I just want  
13 to thank Jimmy for his leadership on this and remind  
14 Dr. Mashariki that there's an outstanding challenge  
15 between us that must be settled.

16 CHAIRPERSON VACCA: Thank you, Councilman  
17 Kallos, and we will now thank you, thank this panel  
18 for their testimony. We have one more panel I think  
19 or two? Two panels. I'd like to now call up Tara  
20 Das, Columbia University, John Caney, Reinvent  
21 Albany. I wish you would. Katherine Gray. Please  
22 come up. Katherine Gray is from the League of Women  
23 Voters, Transparency Working Group. Thank you.  
24 [background comments, pause] Who would like to go  
25 first? Okay, sir, identify yourself for the record.

2                   JOHN CHENEY: I've done this before. I'm  
3 John Cheney. I'm the co-chair of the New York City  
4 Transparency Working Group and also Executive  
5 Director of Reinvent Albany, and I'm here today also  
6 presenting this testimony on behalf of my co-chair  
7 Jean Rushnoff(sp?), Senior Attorney for New York  
8 Public Interest Research Group who was unable to come  
9 today. First, I just wanted to thank you, Chairman  
10 Vacca and the other members. This is one of the best  
11 hearings I've been to, oversight hearings. I know  
12 you're a master of this, but I will compliment you.  
13 The questions that you and the other members asked  
14 were directly on target today. They're exactly the  
15 concerns we have. So it's really gratifying to hear  
16 you guys be so well informed, and so persistent and  
17 just ask the perfect questions and I appreciate that  
18 tremendously. So thank you. Your oversight is very,  
19 very important and I just have brief written verbal  
20 comments. And we submitted a whole ton of written  
21 comments on the legislation that the Council  
22 introduced. I will say we--we strongly support the  
23 intent of all of those bills. We really haven't had  
24 time to work through a lot of the details on them  
25 like how many days should this take or that take.

2 But again, these bills are addressing the problems  
3 that we as data users and advocates are facing every  
4 day with open data. I was very lucky to be one of the  
5 public stakeholders who helped negotiate the New York  
6 City Open Data Law in 2012, and I'm just thrilled to  
7 see how far open data has come and how much we've  
8 achieved here. And I'm about to criticize the Open  
9 Data Law and the Open Data Initiative, but it's all  
10 within the context of the fact that we passed a law  
11 that is truly of global importance. My group had  
12 visitors from all over the world asking us how did  
13 you do it? How did the City Council do it? Why is  
14 New York so special. So before I beat the heck out  
15 of the law, I just want to say, you know, that we  
16 love the progress that's been made, and we're very  
17 proud as New Yorkers in what's going on here, which  
18 includes your oversight.

19 So that said, we're three years into the  
20 open data process and we're starting to see some very  
21 serious problems with how data is made available.  
22 The usefulness of the data that is being made  
23 available, and the use of the usability of the data.  
24 And we recognize the sincere efforts by MODA and  
25 DOITT and the folks that they had here today are true

2 experts, and committed to things. But the cold fact  
3 is this: Agencies are not complying fully or  
4 usefully with the Open Data Law, which you recognized  
5 in your questioning. The reality is that MODA and  
6 DOITT do not have the authority to compel them to  
7 comply with many, many aspects of the Open Data Law,  
8 and this has emerged as a major problem for the  
9 public, and created a tremendous amount of  
10 frustration among advocates, neighborhood groups,  
11 community boards, folks on City Council. Basically  
12 all the public stakeholders of open data, people who  
13 use open data are saying hey it's great But, you  
14 know, it's time for some big changes. And really,  
15 what I'm here to do today is to say to you and  
16 Council and really to the Mayor's Office is we think  
17 you should aim high here. Aim higher than you  
18 perhaps are, and maybe than you think the political  
19 process is ready for. But we want you to aim high.  
20 What do we mean by that? We want you to think about  
21 open data as you do the Freedom of Information Law.  
22 We want the public to be able to petition judges to  
23 make agencies comply with the Open Data Law, and we  
24 want the Open Data Law to be looked at as the other  
25 side of the Open Government coin from FOIL. And

2 that's what we think is going to make open data  
3 really work. So you were getting at this in some of  
4 your lines of questioning. But, um, but what we want  
5 is a private right of action, which is the public's  
6 ability to go to court and get a judge to say hey  
7 agency you have to comply with the Open Data Law just  
8 as you have to do with the Freedom of Information  
9 Law. Because we think ultimately that the Open Data  
10 Law is not going to work nearly as well as it could  
11 unless the public has that--that right and that power  
12 to--to move agencies on that. And we know you're  
13 trying to get at that in other ways there, but we  
14 think that ultimately we have to get there through  
15 private right of action.

16 Now, a couple things. The City Council  
17 has passed many, many ad hoc data reporting bills,  
18 taxi crashes, bike crashes, youth in foster care,  
19 Hepatitis, student demographics. The list goes on  
20 and on, but really the Council cannot pass thousands  
21 of bill requiring agencies to post specific data  
22 sets. You just can't. So just a couple interesting  
23 facts here. We looked at Legistar. Since 1998, the  
24 City Council has passed 398 laws requiring agencies  
25 to report on specific activities, reporting bills or

2 laws. Only five of those 308--98 laws deny the  
3 public the right to go to court, deny a private right  
4 of action. Only five out of 398 reporting laws  
5 passed by the Council. One of those is the Open Data  
6 Law. That's entirely ironic to us. So, open data,  
7 Freedom of Information law, two sides of the same  
8 open government coin. Really, we think the last  
9 three years is pointing to the fact that we do need a  
10 private right of action. We do need forceful action  
11 as some of the bills that you and your colleagues  
12 have introduced call for, and we have to get that.

13           Two other quick major issues with open  
14 data, data quality, and you're going to hear on this  
15 from other folks, too. It is incredibly vexing that  
16 there's no way for the public to report errors,  
17 receive responses or track fixes made to--made when  
18 they report data problems. One of the basic ideas  
19 behind open data is that the public is going to  
20 report problems with government data, and that the  
21 government is going to fix them. And essentially,  
22 the government is acting as free data detectives. So  
23 this is almost like a free service that the public is  
24 giving back to their government to help them clean  
25 data. We all expect data to have big--government

2 data to have problems in it, and one of the ideas  
3 behind open data was to created a positive feedback  
4 loop where we're continually correcting those data  
5 sets, and it's just not happening right now. So,  
6 this is a major, major issue. Some of the  
7 legislation that you introduced speaks to this, but  
8 we want you to keep hammering away on it.

9           The other issue--big issue is that of  
10 data usefulness, and by that I mean is the data up  
11 there, um, as presented useful to say community board  
12 members, and the answer is maybe, maybe not. And we  
13 want you to keep pushing the Administration to be  
14 asking data users, in particular community board  
15 members, City Council staff and non-profit groups  
16 that are serving neighborhoods, at the neighborhood  
17 level what exactly they need. Because we think  
18 there's a lot of frustration at the grassroots level  
19 when people actually do look at the Open Data Portal.

20           The other thing we'd note under data  
21 usefulness is that strangely the City of New York  
22 seems to be keeping three data sets, three copies or  
23 slightly different copies of data sets. One for the  
24 agency, which originates it; one for MODA, which  
25 takes the data and cleans, applies a uniform

2 geospatial or, you know, address field to it; and one  
3 for the Open Data Portal. And those data sets are  
4 different. So there are three different data sets  
5 being used, and that is--that's very, very expensive  
6 and very, very silly. So we would like to see city  
7 agencies start using data in the Open Data Portal.  
8 One data set. That's--we should be seeing the same  
9 data that agencies see minus the privacy or security  
10 fields in it. But something to keep an eye on there  
11 is that why are we keeping three data sets there.  
12 And then just lastly, you know, the bills that you've  
13 done we look at as a huge opportunity for kick  
14 starting this entire process, and really, really  
15 bring the Open Data Law to the next level. So thank  
16 you for all of that and much appreciated, and we hope  
17 to be part of the conversation.

18 CHAIRPERSON VACCA: Certainly, and I  
19 thank you for your support and your very nice words.  
20 Would you identify yourself, please?

21 KATHERINE GRAY: My name is Katherine  
22 Gray. I'm one of the Vice Presidents of the League of  
23 Women Voters. As multi-issue non-partisan political  
24 organization we encourage informed and active  
25 participation in government work to increase

2 understanding of major policy issues, and influence  
3 public policy through advocacy and education. The  
4 League in New York City is over 96 years old, and has  
5 been interested in government operations and how it  
6 affects every person in the community. The path to  
7 the information is now electronic data, and it  
8 assets--access is paramount to the success--to the  
9 successful operation, oversights and transparency in  
10 a democracy. We appreciate this opportunity to make  
11 comments. We're here basically to support the  
12 Transparency working group. The League supported  
13 reform to the Open Data Law, Local Law 11 of 2012  
14 because--because after three years it has become  
15 apparent the law needs adjustments, as we can see in  
16 the seven open data bills before us today. I would  
17 like to reiterate the three key factors that  
18 transparency working groups identify, there is no  
19 legal mechanism to compel agencies to publish data or  
20 keep their own data publishing schedule. Some of the  
21 data sets have been found to be riddled with errors,  
22 and there is no formal method of reporting errors so  
23 those errors can be fixed. Three, not all data is  
24 presented--presented in a uniform and useful manner.  
25 This addresses Intro 898 by Council Member Vincent

2 Gentile requiring a data dictionary for every data  
3 set on the Open Data Portal. The data dictionary,  
4 according to the law, accompany each data set or a  
5 link to such data dictionary. I feel this could be  
6 improved by not saying or. I do believe both. I  
7 think the definition of the data set should be with  
8 the data and in the dictionary. I think that word to  
9 both would improve the law because the idea of  
10 something between database is cumbersome. The  
11 concept of every data set having the same format for  
12 dates, addresses and those addresses being geo coded  
13 with latitude and longitude, community board, council  
14 districts and even census districts would be  
15 revolutionary in the use of the data. We at the  
16 League of Women Voters taking election districts and  
17 census data and we're trying to put them together so  
18 you can understand populations of the non-voters.  
19 The Open Data Law would greatly benefit from all six  
20 council members, members of the Open Data community  
21 and the U.S. Office of Data Analytics working  
22 together on a comprehensive legislation that would  
23 make the Open Data Law more convenient to users and  
24 city agencies. The League applauds the council

2 member and are grateful for the opportunity to speak  
3 to you.

4 CHAIRPERSON VACCA: Thank you. Thank you  
5 for your suggestions and thank you for your support.

6 KATHERINE GRAY: Thank you.

7 CHAIRPERSON VACCA: Would you identify  
8 yourself please.

9 TARA DAS: My name is Tara Das--[off mic]

10 SERGEANT-AT-ARMS: Turn on the mic,  
11 please. [pause]

12 SERGEANT-AT-ARMS: [off mic] The mic.  
13 There you go. it's on.

14 TARA DAS: Good afternoon.

15 [background comments]

16 TARA DAS: [laughs] Good afternoon  
17 everyone. My name is Tara Das. I'm the Intern Head  
18 of the Social Work Library at Columbia, University.  
19 I'm affiliated with the Digital Social Science Center  
20 in the Libraries, which supports data intensive  
21 research by acquiring data collections and managing  
22 data in its spatial and numeric data catalogs. We  
23 provide research consultations to students, faculty  
24 and other university affiliates. My responsibilities  
25 also include overseeing the university collections

2 and government information issued from the federal,  
3 state and city government. With the majority of  
4 government information now in electronic format  
5 including open government data, concerns arise with  
6 preservation of this information so that it remains  
7 publicly available over time. For example, a  
8 government agency may have decided to remove a  
9 publication from its website. If they did not  
10 transfer a copy to an archive or a library, users  
11 will find it difficult to access that publication.  
12 With this in mind, my colleagues and I decided to  
13 select data sets from NYC Open Data and maintain them  
14 in our catalogs at Columbia. I took a much closer  
15 look at the data sets on the portal under the  
16 assumption that if we were going to store this data,  
17 it needs to be shareable and visible by others for  
18 research. Users cannot analyze data if they do not  
19 know what it means, how it was collected and cleaned,  
20 and what was the purpose of data collection. They  
21 need to be equipped with the same information that  
22 the data creators had in order to conduct research  
23 with the data and appropriately evaluate city  
24 services. Close examination of NYC Open Data  
25 revealed that it was not usable due to poor data

2 documentation. For example, data sets were missing  
3 definitions of column names and values and  
4 descriptions of how the data was collected and clean.  
5 I present three cases in my testimony to explain what  
6 I mean, and I'll just hit on the first two at this  
7 hearing. So the Department of City Planning has a  
8 data set related to Hurricane Sandy, damage by Sandy  
9 by age of building. And the columns included in the  
10 data set are year built, inundation area destroyed by  
11 storm, red tags, yellow tags. So how are we to  
12 analyze this data when we don't know what the columns  
13 mean, how the data was collected and the percents in  
14 the data sets were calculated.

15           The Department of Homeless Services has  
16 two data sets on NYC Open Data that overlap with  
17 their daily report, which is available in PDF format  
18 on their website, but they're not all exactly the  
19 same. The time and subject coverage vary across the  
20 three sources. In addition to lack of data  
21 documentation there are other issues with this  
22 report. One of the ones that's on NYC Open Data the  
23 Daily Report of single adult and family intake has  
24 one day of data from January 28, 2013, and the other  
25 one DHS Daily Route is routinely updated, but does

2 not contain columns of the first data set as the  
3 agency website report do. As I hope some of these  
4 cases illustrate, open data is not just a matter of  
5 making data available. Since most of this data comes  
6 from the internal business operations, there are  
7 abbreviations and/or agency jargon in a lot of data  
8 sets. Truly open data, data that is usable by others  
9 requires data about the data or meta data. People  
10 less familiar about the idea of meta data can  
11 understand it in terms of data documentation, data  
12 dictionary or the data code book. That is essential  
13 to understanding the data set that one did not  
14 create, and to interpreting any analysis that used  
15 that data set. In research communities, it is  
16 expected practice to include meta data that describes  
17 the data collection background, instruments, data  
18 definitions, and methodology when sharing data sets.  
19 Given the limitations in NYC Open Data, our immediate  
20 objective shifted to acquiring data dictionaries for  
21 data sets. Since data dictionaries are not required  
22 for NYC Open Data, my first step was to contact the  
23 data set owners through the portal via the about tab  
24 in each data set to see if they can provide this  
25 information, and no one replied. I reached out to the

2 Mayor's Office of Data Analytics to see if they could  
3 assist with providing agency contacts or in general  
4 with acquiring data dictionaries. There was one  
5 conference call with them where these concerns were  
6 expressed by no follow-up action from them. I then  
7 reached out to the New York City Transparency Working  
8 Group. Don McNay (sic) or Reinvent Albany responded  
9 and provided me with a list of open data coordinators  
10 for outreach. So since early this year, I've reached  
11 out to coordinators with questions about specific  
12 data sets. This is a time intensive activity. I've  
13 created one data dictionary and it's included in the  
14 appendix of this testimony for the ACS Community  
15 Partners data set. Since data dictionaries are not  
16 legally required, relying on government officials to  
17 voluntarily provide this information is not the best  
18 way of make open data usable for research. If I need  
19 to contact an open data coordinator about multiple  
20 data sets, putting all my questions regarding data  
21 definitions and data collection methods in one long  
22 email is not the best way to get a response. This  
23 means I ask a few simple questions first, see if I  
24 get a response and then follow up with phone calls or  
25 emails to ask more questions in order to created data

2 dictionaries. So completing just one data dictionary  
3 for one data set on NYC Open Data required sustained  
4 communication within the agency. While I received at  
5 least one response from some open data coordinators,  
6 I didn't receive any response other including OMB,  
7 Department of Correction and Department of  
8 Transportation. Some agencies do not have an open  
9 data coordinator to contact as of February 2015.  
10 That was the list I was using. But I had provided  
11 data sets to NYC Open Data. These include the Civil  
12 Complaint Review Board, the Fire Department of New  
13 York, New York City Housing Authority and Health and  
14 Hospital Corporation. Moreover, government officials  
15 aren't always equipped with definitions and  
16 descriptions of how data is collected and what each  
17 category means. This data comes out of routine  
18 government services and programs and wasn't  
19 originally collected for public release. Sometimes  
20 the people on the ground collecting this information  
21 for internal purposes don't have themselves written  
22 documentation for what specific categories mean. I  
23 encountered this while speaking with DHS about their  
24 daily report. Other times I was referred to laws to  
25 assist with understanding data sets, and while

2 helpful, I'm not a lawyer and couldn't be confident  
3 that I was mapping the appropriate value to the legal  
4 definition.

5           So in sum, obtaining data documentation  
6 voluntarily and retroactively is just not realistic.  
7 In light of these experiences and given the critical  
8 need for preservation and data documentation for NYC  
9 Open Data, I strongly support bills Intro No. 898,  
10 914, 915 and 916. Along with my support, I have  
11 comments for each bill, but I'll just focus on Intro  
12 898 on the data dictionary and Intro 915 on timely  
13 updating. So regarding Bill No. 898 in relation to  
14 requiring a data dictionary for every data set on the  
15 Open Data Portal. So I was very happy to see this  
16 bill proposed because it addresses many of the  
17 concerns outlined in my testimony, but I do have some  
18 comments. In addition to explaining and I quote from  
19 the bill, "Any relationship or connection between a  
20 column or other columns within a data set,  
21 connections across data sets in NYC Open Data and the  
22 NYC website should be explained. This would help  
23 clarify the confusion across a similar report  
24 provided by DHS as presented earlier. All data sets  
25 should have the most recent upload and generation

2 date posted, not just the data sets updated less than  
3 daily as specified in the bill. Daily data sets may  
4 not include a data column with their information, and  
5 even if they do, there may be differences across the  
6 event, date of upload and date of generation. The  
7 dates should also be posted on the portal perhaps in  
8 the about tab where there are also dates of update  
9 and creation related to the portal itself, which  
10 makes it confusing, and not posted in the data  
11 dictionary as suggested in the bill.

12           Regarding the statement in the bill, the  
13 data dictionary may also include any additional  
14 information or description that can provide context  
15 to the data. The data dictionary should include this  
16 information. So that would be nice if that were to  
17 be amended in the bill. Description of the purposes  
18 and methods of data collection is essential for  
19 research and analysis as discussed earlier. Data  
20 dictionary should follow a standard template and be  
21 located in an easy to find location, and external  
22 review of data dictionaries is also recommended. It  
23 is essential that acronyms and technical terms are  
24 explained as the bill requires but government  
25 officials close to the programs that generate open

2 data may not realize what an outside person would  
3 consider a technical term.

4           Regarding bill--Intro No. 915 in relation  
5 to the time of the updating of certain public sets in  
6 the Open Data Portal, yes NYC Open Data must contain  
7 data that is synchronized as possible with data  
8 published on data websites. Lack of updated data on  
9 NYC Open Data not only lowers its utility, but it  
10 frankly lowers its prestige and value. And, in fact,  
11 at Columbia University we encourage researchers not  
12 to rely on NYC Open Data and to consult agency  
13 websites for the most recent data. In looking  
14 forward, the operations underlying NYC Open Data need  
15 to be less manual for updates to happen seamlessly.  
16 And really could consider following data.gov and  
17 harvesting data automatically from agency websites  
18 for new and updated data sets without having to  
19 consult with them one on one and go case by case as  
20 we explained earlier.

21           In conclusion, if the primary intent of  
22 NYC Open Data is make government more transparent and  
23 accountable to New Yorkers. That goal cannot be  
24 easily achieved if the public is not also provided  
25 with sufficient information that makes the data

2 understandable and analysis meaningful. Simply  
3 transforming government information into open data  
4 format and releasing numerous open data sets does not  
5 add this value. The proposed bills, particularly the  
6 bill requiring data dictionaries are essential to  
7 making NYC Open Data usable for research and  
8 analysis. Thank you.

9 CHAIRPERSON VACCA: Thank you. I want to  
10 thank this panel, and we have one more panel. John  
11 Klaus and Josh Mumm. [pause] Mr. Mum, do you want to  
12 go first?

13 JOSH MUMM: Sure. I'd love to.

14 CHAIRPERSON VACCA: How are you?

15 JOSH MUMM: Good. How are you?

16 CHAIRPERSON VACCA: All right. Thanks  
17 for staying.

18 JOSH MUMM: Yeah, it's been a long one,  
19 but I will be very brief. My name is Josh Mumm. I  
20 am the Outreach and Advocacy Manager for Common Cause  
21 New York. We provide a voice for citizens. It's  
22 more of an open and honest and accountable government  
23 at all levels. So thank you for holding this  
24 hearing. There's a lot of bills on the table today.  
25 So I'm just going to comment on three of those very

1 quickly like I said. New York--New York City's Open  
2 Data Policy was made law at the end of 2012. It's  
3 already been implemented under two different mayoral  
4 administrations. Common Cause New York recognizes  
5 that a lot of work goes into coordinating the city's  
6 numerous agencies, and we'd thank the Department of  
7 Information if they were here still, the Department  
8 of Information Technology and Telecommunications and  
9 the Mayor's Office of Data Analytics for their  
10 continued hard work in wrangling information out of  
11 HC silos. I know it's a very long and difficult  
12 process. However, the law still has a long ways to  
13 go in terms of ensuring universal compliance. Common  
14 Cause New York supports Intro 916 of 2015, which  
15 would mandate the Department of Education--  
16 Investigation to perform an open audit on nine  
17 departments including the Department of Corrections  
18 and the NYPD. Conducting an audit will not only be  
19 helpful in unraveling the confusion that an  
20 individual agency may have in complying with the law,  
21 it will also be a concrete way to delineate future  
22 compliance with other agencies. We strongly suggest  
23 that the New York City Economic Development  
24 Corporation and the Department of Small Business  
25

2 Services, also subject to FOIL, should be included in  
3 the agency whose compliance is included in the audit.  
4 Thus far, there has been no comprehensive open data  
5 audit, and it is now long due as we are almost four  
6 years into the implementation of the law. We are  
7 also in support of the spirit of Intro 914 of 2015  
8 though we are unsure of how it will be carried out.  
9 The lack of responsiveness to questions posed by the  
10 public in the Open Data Portal is a real cause for  
11 concern, and there must be a way for concerns and  
12 queries to be answered in a systematic interactable  
13 format.

14           Additionally, we want to voice our  
15 support for Intro 908 of 2015, which would place data  
16 sets that have been successfully requested by FOIL  
17 into the Open Data Portal easily allowing other  
18 members of the public to view FOIL information is not  
19 only a sensible move from a transparency and good  
20 government perspective, it will also drastically cut  
21 down on repetitive FOIL requests and thus save time  
22 for agencies and FOIL officers.

23           We share the concerns of our colleagues  
24 at Reinvent Albany for the enforceability for the  
25 Open Data Law. While we are supportive of the idea

2 of creating a private right of action to gain  
3 compliance, we are mindful of the burden and costs  
4 associated with litigation, which makes it a seldom  
5 used tool in these situation. Perhaps an additional  
6 enforcement mechanism might be to empower the  
7 Department of Investigation upon request to determine  
8 whether there has been a failure of compliance and  
9 issue such a finding in order to foster agency  
10 compliance. As for the other bills being heard  
11 today, Common Cause New York supports their passing  
12 as they will bring added clarification to the law in  
13 terms of mandated compliance. In conclusion, we echo  
14 our colleagues' testimony for the need to improve  
15 overall data quality that is made available to the  
16 public, and we hope that more resources will go  
17 towards staffing the Open Data Portal. Thank you.

18 CHAIRPERSON VACCA: Thank you. Your  
19 testimony was very concise. Thank you.

20 JOSH MUMM: That's how we like it.

21 CHAIRPERSON VACCA: Yes, me too. Go  
22 ahead, Crouse.

23 JOHN CROUSE: Oh, there we go. Thank you  
24 for questing that I testify today about these  
25 proposed amendments to these proposed amendments to

2 the Open Data Law . My name is John Crouse and I'm  
3 the Technology Fellow at the Gov Lab at NYU whose  
4 mission is to improve people's lives by changing the  
5 way we cover them including through the increased  
6 availability and use of open--of data. I'm also  
7 active with the Transparency Working group and in the  
8 data community in New York City. Within that open  
9 data community there's been agreement for years that  
10 unanswered questions and unresolved problems in the--  
11 in the data posted on the portal impede the use of--  
12 To understand this issue, we downloaded all 293  
13 comments from the portal from 2011 to the present and  
14 analyzed them. So I wanted to speak to kind of the  
15 results of that analysis, and I think it provides  
16 some interesting evidence as to the deficiencies in  
17 the way the process works right now.

18           So we determined there were four  
19 categories, one of which was problems need a fix such  
20 as this data set 70 poll sites have blank entries for  
21 the location field. There were 70 comments along  
22 those lines. Questions that needed an answer such as  
23 will this be updated regularly, which was 106  
24 comments and then two other categories, comments that  
25 didn't need a response like I don't trust this data

2 or responses such as thank you for the feedback and  
3 that was another 114 comments in those categories.  
4 So if you actually begin to look at these comments  
5 and try and figure out what happens to them after  
6 they're made, you'll find that slightly less than  
7 half of the questions, 51 of 106 actually were  
8 answered. And on average about half a year, 180 days  
9 elapsed between the question being posted and the  
10 response being posted. The average unanswered  
11 question on the data portal is over a year old  
12 already. Comments that raised actual problems with  
13 the data as opposed to questions that needed to be  
14 answered didn't even worse. About two-thirds or 46  
15 of the 70 issues that were raised in comments haven't  
16 yet been fixed. On average these unfixed problems  
17 were reported over a year ago. Over half of these  
18 unfixed problems, 28 of the 46 never even received a  
19 confirmation that the problem was being looked into  
20 by anyone in the Open Data Portal. And of the  
21 unfixed problems that were confirmed, 18 of the 46 on  
22 average there was--on average over a year has elapsed  
23 since that confirmation with still no fix.

24           What we really need are people who can  
25 respond to comments by actually fixing the issues

2 brought up in them, and we need people to fix this  
3 data, not necessarily more requirements that it be  
4 fixed. The current arrangement where DOITT manually  
5 ferries reports of problems to agencies that may or  
6 may not have resources or people to resolve them has  
7 not worked. We need open data expertise on the front  
8 lines responding to these problems and empowered to  
9 fix them. This means funding for staff and resources  
10 within the organization dedicated to open data, which  
11 unfortunately DOITT is not. It also means a private  
12 right of action, which was previously raised by open  
13 data users to put them in the same class as FOIL  
14 applicants whose requests must be responded to. I  
15 want to emphasize that DOITT is doing their best to  
16 resolve questions and concerns on the Open Data  
17 Portal. There are deficiencies in the portal's  
18 design itself, which make it difficult to keep track  
19 of these comments, and without a clear mandate to fix  
20 the problems at the source. Which DOITT does not  
21 have, often times there is nothing more that they can  
22 do than tell the user who noticed the problem that  
23 the agency has been informed and nothing more. Which  
24 understandably is very frustrating and I think to the

2 detriment to the portal user experience in useful of  
3 open data in New York City. Thank you.

4 CHAIRPERSON VACCA: Thank you. I want to  
5 thank you both for your testimony, and I want to  
6 thank everyone for coming to the hearing today. We  
7 have no further witnesses. It is now 3:40 p.m., and  
8 this hearing of the Technology Committee is hereby  
9 adjourned. [gavel]

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

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 October 5, 2015