

CITY COUNCIL
CITY OF NEW YORK

----- X

TRANSCRIPT OF THE MINUTES

Of the

COMMITTEE ON TECHNOLOGY

----- X

October 27, 2014
Start: 1:30 p.m.
Recess: 3:15 p.m.

HELD AT: 250 Broadway - Committee Rm,
14th Fl.

B E F O R E:

JAMES VACCA
Chairperson

COUNCIL MEMBERS:

Annabel Palma
Mark S. Weprin
David G. Greenfield
Steven Matteo
Ben Kallos

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

Nicholas O'Brien

Director of Public Affairs

Mayor's Office of Data Analytics

Albert Webber

Department of Information, Technology and

Telecommunications (DoITT)

Gale Brewer

President

Manhattan Community Board

Gene Russianoff

Senior Attorney

New York Public Interest Research Group

Dominic Mauro

Staff Attorney

Reinvent Albany

Noel Hidalgo

Executive Director

BetaNYC

Christine Berthet

CHECKPEDS

John Krauss

The Governance Lab

David McCreery

Manhattan Resident

Hacker

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

Joel Natividad

Co-Founder of Ontodia

CHAIRPERSON VACCA: Thank you for all coming to this wonderful hearing, and I apologize for being late, and if any of you know me, I'm always on time, but they also had the City University Committee the same time as this and that was across the street, so it made it a little difficult, but I thank you for your patience and I apologize again.

Good afternoon everyone; today is October 27th; I'm James Vacca and I'm Chair of the Committee on Technology and we're here today to discuss the implementation and administration of Local Law 11 of 2012, best known as the Open Data Law.

This important law required the Department of Information Technology and Telecommunications (DoITT) to create a one-stop shop for easily accessible, nonproprietary and machine readable government data that can be used and analyzed by the public at large. Since collecting and organizing all of New York City government's open data into a single portal is an enormous undertaking, this law allowed the implementation over several years, requiring full compliance by December 31st, 2018. DoITT released its annual Open Data Plan this past July, which included the release of new data

sets and an updated timeline of data releases. The Agency has also since made improvements to the site's usability and search feature. Testimony from the Mayor's Office of Data Analytics (MODA), DoITT and advocates will address this Open Data Plan and ways the portal and its data can be improved in the future. I thank all those for testifying and giving input today.

Since the last time this committee examined the implementation of this law, many changes have been enacted. Most recently, the Annual New York City Open Data Plan was released on July 15th, 2014. As part of this report, [background comments] 345 data sets were scheduled for release, 160 have been published since the prior plan and 27 sets were removed from the Plan. Among the sets released that are of significant public interest; they include restaurant inspection data, motor vehicle collision data and prison capacity data. In addition to the additional sets, the site itself has been organized in a more user-friendly format -- the search function has improved and lists of official sets aggregated by agency have been prominently featured on the site. While DoITT has made many improvements and has made

great efforts to work with Council and advocates, there is still much to be done and several issues to identify and address. There are some data sets that were identified in previous plans that were removed from the most recent plan and some sets were created and included within the last years. To ensure full compliance, this committee hopes to hear the reasoning for the inclusion of some sets and for the exclusion of others. Furthermore, some sets that were supposed to be released already, according to the timeline released in the July plan, have not yet been released. Going forward, it is also important to identify how the Administration has prioritized the release of data sets, the information included in those sets and the prioritization of further data set releases. I also want to ensure that the integrity and accuracy of available data is always paramount.

On top of the recent improvements to the user interface, I also hope to learn more about how it may be further improved, taking into consideration suggestions from all types of users.

Finally, having used the Open Data Portal myself to identify trends in my own district and throughout the city to improve constituent services

and input that I could give into policy and legislation, I must emphasize the vast possibilities available through this portal. I hope to continue working with the advocates, MODA and DoITT to ensure that the Open Data Portal is as user-friendly as possible.

With that I'd like to turn it over to Nick O'Brien from the Mayor's Office of Data Analytics, who will be our first witness. I now have to swear you in; this is something new we're doing. I've been swearing at people all my life, but now I'm swearing... [laughter] Would you introduce yourself, sir; I don't have your name, Mr. O'Brien and Mr...

ALBERT WEBBER: Albert Webber.

CHAIRPERSON VACCA: Mr. Webber. Okay, Mr. Webber, you're from DoITT?

ALBERT WEBBER: Yeah.

CHAIRPERSON VACCA: Okay, well then let's do it. [laughter] Do you swear or affirm to tell the truth, the whole truth and nothing but the truth in your testimony before this committee and to respond honestly to council member questions?

NICHOLAS O'BRIEN: I do.

CHAIRPERSON VACCA: Do you?

ALBERT WEBBER: I do.

CHAIRPERSON VACCA: Okay, I figured you did. Okay. Thank you. Mr. O'Brien, would you please introduce yourself for the record and start with your testimony?

NICHOLAS O'BRIEN: Thank you, Chairman Vacca.

Good afternoon; my name is Nicholas O'Brien, I'm the Director of Public Affairs in the Mayor's Office of Data Analytics and the Acting Open Platform Officer for the City of New York.

My roles as Open Platform Officer -- I am responsible for the implementation of Local Law 11 of 2012, the New York City Open Data Law.

Here with me is Albert Webber of the Department of Information, Technology and Telecommunications Office of Advanced Technologies.

The Mayor's Office of Data Analytics is housed within the Mayor's Office of Operations and works with the Department of Information, Technology and Telecommunications (DoITT) to fulfill the requirements of the Open Data Law. The law has the full support of the Administration and we're working tirelessly to ensure it is fully implemented.

The New York City Open Data Law was signed into law on March 7, 2012; the law was a product of a collaborative effort, including the City's technology community, the nonprofit sector, the Council, City agencies and the Mayor's Office. It has served as the basis for a number of other open data and transparency policies and initiatives throughout the country and established New York City as a leader in the field. I'm here today to testify on the progress of implementation of the Open Data Law and highlight some of the successes we've had since the last hearing on this topic in November of 2013.

The primary vehicle for delivering data to New Yorkers is the Open Data Portal, accessible at [NYC.gov/data](http://nyc.gov/data). To date we have released 1,268 unique data sets on the portal, up from 350 in 2011, when the portal was first launched; 99 of those sets have been fully automated and new automations are coming online every month. In July, we released the 2014 edition of the New York City Open Data Plan, which catalogs 345 unreleased data sets covered under the law and their planned release dates. When combined with the 1,268 already released, we have a total of

1,613 data sets, which comprises the known universe of open data sets maintained by the City of New York.

The 2015 edition of the Open Data Plan will be released to the Council on July 15th of 2015 and between now and the release of the 2015 plan we will publish an additional 100 data sets. In addition to these activities, which are mandated under the law, we are undertaking numerous initiatives beyond what is required by the law to increase transparency and facilitate the use of City data to provide meaningful insight.

Just last week we launched significant enhancements to the search and browse functionality, which Mr. Webber will detail in his testimony. These enhancements were the result of a collaborative process undertaken by the Mayor's Office, DoITT, Open Data advocates and the largest civic tech community.

In August, the Mayor signed two bills into law, Local Law 37 and 38 of 2014, to create a more transparent and effective city government. Local Law 37 requires the Law Department to publish the City Charter, the Administrative Code and the Rules of the City of New York online and update the compilation of laws within 30 days of any change.

Local Law 38 requires the City to post the City Record online within 24 hours of publication. At a public hearing preceding the signing of those bills, the Mayor also announced an innovative public-private partnership to transform 4,000 publications of the City Record currently stored in PDFs into machine readable format to ensure they can be used to understand trends and patterns surrounding City operations.

In September, the Mayor announced the winners of the 5th Annual New York City BigApps competition. This year's competition, hosed by DoITT and the Economic Development Corporation, was designed to use city data to make New York a better place to live, work, play and learn. The winner of the live category, Heat Seek, uses 311 data and thermometers placed in tenants' apartments to help ensure that New Yorkers' apartments are kept at or above the minimum temperature during the heating season.

BigApps is also a driving force in creating good technology jobs based on open data. Two examples -- MyCityWay, which received \$6 million in venture capital funding, and Embark NYC, which was

acquired by Apple in 2013, are early examples of viable businesses driven by open data.

These initiatives are just the latest embodiment of the Mayor's long-standing commitment to openness and transparency in government. This administration believes this commitment is essential to create an equitable city and ensuring every New Yorker is well-served by its government.

With regards to the data itself, we know the City's data is complex and we provide extensive background information on featured data sets to increase awareness and usability of the most frequently viewed data sets on the portal. We are currently featuring the newly released restaurant inspection data, which has never before been directly available on the portal. We're also using the Open Data Portal to show visualizations built on public data. These visualizations make the information contained in the portal more accessible as New Yorkers seek to understand their city.

Our work to improve usability and access to city data includes engaging with agencies on an ongoing basis to enhance the public's ability to access, consume, map and visualize New York City's

geographic data. By leveraging the city's GIS experience, we have enabled users to create different types of maps of more than 165 data sets through the Open Data Portal, including hurricane evacuation zones, Wi-Fi hotspot locations, parks and building footprints. This map data is not covered in the law, but we feel that is an important component of Open Data.

One of the benefits of open data we are working to quantify is the improvement of fulfilling FOIL requests. Last year the Environmental Control Board (ECB) received two FOIL requests that they were able to respond to by directing the requestors to the Open Data Portal. This not only makes it easier for agencies to respond to requests; it also provides faster, easier and cheaper access to the data, potentially reducing the need for the public to submit lengthy FOIL requests to gain access to information.

We are also focusing on measure the economic benefit of open data. We are tracking aggregate usage and have begun to catalog the applications built on top of the data. This is the first step in our effort to rigorously quantify the

economic impact of open data. These benefits come in many forms, ranging from improved agency efficiency and cost avoidance to the creation of new jobs and businesses made possible by the release of this data.

We are confident that this analysis will demonstrate that the City's investment in open data makes economic sense and look forward to presenting the result of this effort next time we have the opportunity to testify.

I would like to thank the Council Technology Committee for providing me this opportunity to testify today and at this point I'd like to turn it over to Albert Webber from the Department of Information, Technology and Telecommunications to walk you through some of the finer details of the implementation and the new enhancements we have released in response to feedback from the community. After Mr. Webber's testimony, we'll be happy to answer any questions you might have.

ALBERT WEBBER: Good afternoon; my name is Albert Webber of the Department of Information, Technology and Telecommunications (DoITT). On behalf of Commissioner Anne Roest, I would like to thank

Chairman Vacca and the Members of the City Council Committee on Technology for convening this hearing on the implementation of Local Law 11 of 2012, the Open Data Law. Open Data is a major priority for DoITT and I'm here today to share our progress to date, highlight the recent enhancements we've made to the NYC Open Data Portal and to discuss our strategic plan going forward.

Mr. O'Brien already articulated the de Blasio Administration's strong support for government transparency and their efforts to ensure the Open Data Law is implemented fully. DoITT works closely with the Mayor's Office of Data Analytics in the implementation of this law.

Local Law 11 of 2012 requires that all public data sets maintained by New York City agencies and departments be made available on the internet through a single web portal. DoITT maintains this portal, NYC Open Data at [NYC.gov/data](http://nyc.gov/data). We currently have 1,268 unique data sets available via the portal. Each year all agencies and departments are required to submit an inventory of public data sets that they have not yet made public, along with the dates that they intend to publish them. This inventory is

combobbed [sic] by my office and published as the New York City Open Data Plan. The latest edition of the Plan, published on July 15th of this year, includes 345 data sets scheduled for future release. Since July 15th, 60 of the data sets on the Plan have come due and I'm happy to report that we have successfully published 57 of them or 95 percent. We are on track to publish an additional 100 before the next edition of the Plan, which is scheduled for submission to the Council on July 15th, 2015. As per the law, all public data sets maintained by City agencies and departments must be published by December 31st, 2018 or a reason must be given as to why that data set cannot be released and when it is expected to be released.

To enable the Council and the public to track the progress we've made, we included listings of all the data sets we've published since the inaugural release of the Plan, as well as any data sets that were removed from the plan and the specific reason why they were removed. For example, being merged with existing data into a combined data set.

We also provide a dashboard of available data sets which is updated monthly and provides a

concise listing of data sets currently available on the portal and those scheduled for future release.

I would now like to turn to some of the enhancements we've made to the portal; these improvements were made in response to feedback from users of the portal and were developed in close collaboration with the open data community.

The portal has been improved to enhance search functionality. Initially the portal relied on industry standard technology, but it was unfortunately not sufficient to navigate the thousands of official city data sets and community-generated use available in the portal. Working with the open data community and our technology vendor, we were able to refine the search function to provide much more relevant results. Referring to the slide behind me -- for example, if you would've searched 311, which is one of our most popular data sets, using the old search technology, 311 would come back on the 9th page as the 214th data set available. A user would then have to click through the pages of community-generated use to find the actual data set. With the new technology just released last week, a

search for 311 now yields the entire 311 service request data set as a top result.

While discussing the search enhancement I mentioned two types of data sets -- official city data sets and community-generated use. I would like to take a moment to explain the difference.

One of the NYC Open Data Portal's most useful features is that it allows users to create a specific view or filter of a data set and save it for future use. For example, a user may only be interested in the Department of Housing Preservation and Development's complaints in his or her zip code and create a filtered view with only this information. We feel this is an important functionality to allow users to access relevant information. However, there was so much user activity on the portal that it became difficult to find the original data set among all the community-generated use. Therefore, our release last week created a feature that allows for searching or browsing of only official city-published data sets, as you can see on this slide. While all the community-generated views are still available to users of the portal, users should now have a much

easier experience locating the original city data sets.

In addition to the enhanced search functionality and the ability to look at only official city data sets, we have also created the ability to search or browse by a specific City agency or department. For example, users can now browse and see all the data sets available from the Administration for Children's Services, Taxi and Limousine Commission or any other agency with data on the portal.

I would now like to discuss where we're going next with the Open Data Initiative. There are four areas we are focused on this coming year; they are automations of older data sets, increased community partnerships and interaction, deployment of a new open data interface and quantifying the impact of the Open Data Initiative overall.

Automation. When we first launched the NYC Open Data Portal in 2009, it was known as the NYC Data Mine. The Open Data Law had not yet been passed and all data posted to it was done so on a voluntary basis. We will be focusing on automating these older data sets to ensure they remain updated and relevant.

Partnerships. There's a vibrant open data community in New York City; the improvements we made to the Open Data Portal described earlier would not have been possible without the input and spirit of collaboration embodied by the New York City open data community. We value the feedback we receive, both on what data sets should be prioritized and the technology used to deliver them, and view our relationship as a true partnership. The publication of these data sets is useless unless we have users who make it a positive exchange for New Yorkers.

There are events nearly every weekend and evening and we receive hundreds of comments from users about City data and strive to answer every one of them. To that end, we are in the process of adding a new staff member to ensure that all comments and data set nominations receive a prompt reply and that the City's well represented at community events and that we remain engaged with this community. This representation will allow us to deepen the communication channels between the City and open data users, enabling us to not only be more responsive, but also to communicate new data and features we have deployed or that are under development.

Innovation. Last week our Open Data Portal vendor, Socrata, unveiled the next generation of its platform. This week we are working with the vendor and members of the New York City open data community to conduct usability testing on the new site to ensure that it meets all the needs of the City and our users. This will give us the chance to give feedback to the vendor and to customize the product to fit our specific needs.

The new interface will offer New Yorkers a host of new features; one being, a new online experience focused on allowing nontechnical New Yorkers to easily access information in a way that's meaningful for them, including intelligent and automatically traded data visualizations based on the data being analyzed. For example, when New Yorkers search for 311 data, the tool will automatically trade data visualizations, showing service requests by agency, a map of data by community districts and a chart showing service requests created over time. Every data set that includes location-based information can be automatically visualized by those attributes and by custom boundaries. For example, by borough, community board or zip code.

Technical users can imbed NYC data and visualizations in their own websites and access them via tools of their choice; significantly improved performance, allowing users to easily and quickly access and analyze millions of rows of data; ability to query data based on location; for example, how many pothole requests have been submitted within a specific community board; prebuilt tools to allow technologists to easily access open data based on the technology of their choice, like iPhone or Android; improved data automation tools, which will simply agency open data automations going forward.

Impact. Finally we are in the process of measuring the impact of our open data efforts; these benefits come in many forms, including better service delivery, government transparency and increased economic activity. We are now working to take the next step and comprehensively quantify the impact of open data for New Yorkers and I'll be working closely with the Mayor's Office of Data Analytics in this effort.

In closing, DoITT strives not only to comply with the Open Data Law, but to excel in creating an open and transparent government. We are

working diligently to ensure New York City remains a national leader in the open data space.

I would like to thank you for the opportunity to testify this afternoon and Mr. O'Brien and I are happy to answer any questions you have.

CHAIRPERSON VACCA: Thank you. Thank you both very much. We've been joined by Councilman Matteo.

Let me ask you some questions. How many people have used the site; how many hits do you get and... [interpose]

ALBERT WEBBER: So to date, from when we launched with the Socrata portal in 2011, we've had over 30 million page views; you can see that through an analytics page that we released to the public. We don't require registration on the portal, so in some cases it's hard for us to actually a great idea of who specifically is using the portal, but we know to date there have been over 30 million page views.

NICHOLAS O'BRIEN: And you can find that just on the drop down menu of the NYC.gov/data.

CHAIRPERSON VACCA: Are community boards using this; do you get inquiries from community boards; have you ever trained the district managers

and maybe they have a tech chairman on the community board? Because I'm not aware of community boards using this; I don't even know if they know about it.

[laughter]

To be honest with you, I don't even know... I doubt they're using it.

ALBERT WEBBER: There are actually a good number of community boards using the Open Data Portal. Maybe two or three weeks ago we met with Community Board 1 and we gave them a full tutorial on how to use the portal.

CHAIRPERSON VACCA: Well I'd like you to have a meeting with the Bronx Community Boards; okay?
[crosstalk]

ALBERT WEBBER: We can do that. We can do that.

CHAIRPERSON VACCA: I'd like that. And maybe it can be held in The Bronx and maybe the district managers; maybe you wanna do it to the Borough President's office, but I really think this is a resource, but yet no one is talking about it in my community and I don't even know if people know about it. So this is a tool that they should be using because they're in the vanguard of making sure

that people get City services and analyzing how agencies perform on a district basis. But I do think that we have to do more about engagement of the public and perhaps that can be a start.

ALBERT WEBBER: One of the things we plan to do is, in terms of bringing on a whole new staff with Open Data is having people that are focused on, I guess evangelizing open data; putting the word out. Our communications office does the job of tweeting out the new releases of the search functionality when we release new data sets, but we plan to be bringing onboard new staff members that will be doing a better job of just getting the message out there for Open Data.

CHAIRPERSON VACCA: 'Kay. Now, who makes the determination if an agency is covered by Open Data or not? Is that the call of Council; is that DoITT? Who does that?

NICHOLAS O'BRIEN: The Law Department is involved and also working with the agency's counsel.

CHAIRPERSON VACCA: Okay. Because you know why I say that; I notice that several agencies are not on there; I always question whether or not what we do here at the Council impacts NYCHA, HHC;

the non-mayoral agencies. I noticed the Rent Guidelines Board is not covered and I wanted to ask you about that, because if you go to CityStore, you can buy their proceedings, basically, or you can do that and they have their information, but they don't have it on this portal. Can you explain this to me; how do we govern quasi City agencies like that?

NICHOLAS O'BRIEN: The determination is made by corp [sic] counsel; we work with them very closely to determine who is on that list; we're happy to provide you with the full list and then any -- like RGB -- that is not on that list, we can get an opinion as to why it was not included.

CHAIRPERSON VACCA: I'd like us to fight for more inclusion. I would like to make sure that as many agencies, whether they be a true city agency or a hybrid, be included there, and why should someone have to go to CityStore and buy that type of information when it should be readily accessible, because what, in this case, the Rent Guidelines Board does is of significant impact to the people of New York City.

NICHOLAS O'BRIEN: One of the things we do is encourage voluntary release of data, so even if

the Law Department does not determine that a body is covered by the law, we will still publish their data -- for example, School Construction Authority -- it was determined they were not covered, but we do publish a good amount of data from them on a voluntary basis and we continue that activity.

CHAIRPERSON VACCA: I wanna talk about timeliness of information on the site. Now I went on the site last week and I did a search of several agencies and one of them was the Buildings Department, which I go on the BIS system three; four times a week, checking out new permits in my district, so I noticed some complaints in there and it said, basically, something to the effect that action was pending, there was no result of the complaint that was filed by citizens, yet knowing the block in my district the way I know the block, I know that those complaints were resolved, so I went back to BIS and in the BIS system there is a resolution that's about 30 days old, but it was not in the portal under open government. So I wanted to know why is there that lag; why is there, in Open Data, not the information that would be on the BIS system, for example and is this indicative of a lapse in time

that happens consistently through 311 or... just explain that to me.

ALBERT WEBBER: So with the DOB permit data set, that is a data set that we were made aware that there were certain gaps in the information that are there; we relayed this message to DOB and we're working with them to correct that automation and to make sure all of the data is on the portal. So that's something that we're actually working on. But you'll find that, like there are 99 other automations of data sets that are on the portal that we update on a regular basis and are timely maintained, but the DOB permits issue is one that we're aware of and we're working to fix.

CHAIRPERSON VACCA: And it's the 311 data site that I was referring to, more than just the BIS system; I think much of this is related to 311. Is there a lag in disposition of complaints or disposition of action that 311 has that we should be aware of? Are you in touch with 311 to make sure that there's collaboration?

ALBERT WEBBER: We're in touch with 311; that data set is updated on a daily basis. There are certain data sets that are exceptions that are left

out; usually if they deal with minors or something of that sort, but the data set as a whole is updated on a daily basis.

CHAIRPERSON VACCA: Now one of the data sets that was removed from the latest plan is called Murder in New York City; it was removed because it was a report and not data, yet this was not caught before it was included in the original Open Data Plan; do you vet the data sets planned for future release, and have you asked New York PD to supply the underlying data set behind the Murder in New York City report?

NICHOLAS O'BRIEN: So we rely heavily on agencies to do an inventory of their own data and we do give them feedback on data we believe may be missing or data that may be inappropriately included; at the end of the day it is the agency that is submitting the final list that we and DoITT compile. We to date have not requested the underlying data; we have not had an opportunity to do so since the publication of the Plan, but anything that we do identify that we believe should be included we will give that as feedback to NYPD and any other agency that we believe has data that may be missing.

CHAIRPERSON VACCA: You speak to whom in that agency, the communications person in that agency, the technology people? Who do you communicate with?

NICHOLAS O'BRIEN: Each agency has an Open Data coordinator that they have assigned and that is our point of contact and we rely on them to talk to the relevant people, be they in legal, be they in IT; be they program folks, because there's a whole host of different parts of each agency which need to be involved in publishing these data sets.

CHAIRPERSON VACCA: If there's a disagreement between what you think you are entitled to and what the agency wants to give, who is the arbiter in those cases?

NICHOLAS O'BRIEN: So currently it's my office, the Mayor's Office of Data Analytics; we do have a new chief analytics officer coming in next month and they will be the person who will make that determination.

CHAIRPERSON VACCA: So someone's coming on soon?

NICHOLAS O'BRIEN: Yes.

CHAIRPERSON VACCA: Okay. Councilman Kallos has joined us. Several data sets were added for the latest update that were not in the original plan; why were they overlooked or excluded for the first time?

ALBERT WEBBER: I think the first time around we were still feeling out the landscape of what should be in the Open Data Plan, from our end and from the agency, and we were just getting our feet wet with Open Data. We put out the initial plan and we knew that there would be some additional work behind it, some additional identification of data sets. I think each year, when you see that annual update, you'll see the maturity and how much we're coming along in terms of identifying data sets, but I think the first time around we just sort of... we were just getting an idea of what was going on and what should be included in the plan.

CHAIRPERSON VACCA: If a member of the public notices an error in an agency data set and posts a comment, how is the comment handled; is it...
[crosstalk]

ALBERT WEBBER: We receive...

CHAIRPERSON VACCA: is it relayed to the agency or?

ALBERT WEBBER: We receive comments through a number of channels, but if we see an issue with a data set, if there's some comment by the public, if we can address it; we'll address it. If it's something that needs to be specifically addressed by agency; we will relay that message to their Open Data coordinator, who will do their internal process of figuring out the resolution to the problem, relay it back to us and then we'll inform the public. We signed up to be the buffer between the public and the City agency, so we take in all of that feedback; we do the back and forth and then we bring it back to the public.

CHAIRPERSON VACCA: I wanna get on to our Borough President, Gale Brewer, who's been patiently waiting.

COUNCIL MEMBER KALLOS: May I ask a couple of questions?

CHAIRPERSON VACCA: Yes, just one or two quickly, because I started the meeting late, so in deference to Gale, I wanna get her on.

COUNCIL MEMBER KALLOS: Thank you, Chairman Vacca. I joined this Committee from the Women's Issues meeting that we both came from that also started late, so please accept our apologies if this has already been asked. Is DoITT releasing information that's being FOIed on a proactive basis, or planning to do so? Frequently... [crosstalk]

NICHOLAS O'BRIEN: No...

COUNCIL MEMBER KALLOS: Frequently FOIed documents, does DoITT have any plans to release that information on a proactive basis?

NICHOLAS O'BRIEN: If it appears on the Open Data Plan, then yes.

COUNCIL MEMBER KALLOS: To the extent it does not, would DoITT consider adding to the Open Data Plan items that are frequently requested?

NICHOLAS O'BRIEN: We give suggestions to the agencies in terms of data we believe may be missing from their plan and then ultimately it is up to the agency to include it in the plan and DoITT compiles it.

COUNCIL MEMBER KALLOS: What do we currently use for hosting our open data?

NICHOLAS O'BRIEN: Socrata is our...

[crosstalk]

COUNCIL MEMBER KALLOS: Is that...

NICHOLAS O'BRIEN: hosting solution.

COUNCIL MEMBER KALLOS: And what is the license on that; is that free and opened source or is it a different item, uh... [interpose]

NICHOLAS O'BRIEN: No, it is a proprietary system.

COUNCIL MEMBER KALLOS: Is there any concern about vendor lock-in; is there an investigation into adopting the federal software which would be free and opened source that the code has hosted on I believe a federal GitHub?

NICHOLAS O'BRIEN: There are a number of opened source solutions that are emerging in this space; they're still early stage and we're keeping an eye on them to see if they become viable solutions in the future; we do keep abreast of all the emerging technologies; we believe right now that this platform that we are using is the best one to suit our needs, but again, we do monitor the vendor landscape to see what is available.

COUNCIL MEMBER KALLOS: Do you monitor query times for average users and will there be an increase in speed and usability of the Open Data Portal and search functions?

NICHOLAS O'BRIEN: We do believe there will be an increase in speed with the rollout of Socrata 2.0.

COUNCIL MEMBER KALLOS: And for a final question, which just slipped my mind; give me one moment... Budget data; how soon before we get budget data in actual data, computer-readable data; not PDFs or reports, but actually being able to look and see exactly how much we're spending on every pen, ashtray or what have you?

NICHOLAS O'BRIEN: One of the things we've done with Open Data is make sure that data that is not released in any way to the public is prioritized and get that out first; the Comptroller has done great work with Checkbook 2.0 and so we saw that... there was some data is not available in any format that we wanna get out first before we turn our attention to making something that may have a lot of overlap and redundancy with that system.

COUNCIL MEMBER KALLOS: To the extent that will be released in the budget in open data format, I would love to adopt a federalized schema that we can use in every single city, state municipality and government at all levels. Thank you so very much for your amazing work and being leaders in the nation on open data.

CHAIRPERSON VACCA: Thank you. Thank you so much. I'd now like to call up Borough President Gale Brewer. And after Gale we'll have a panel of three and three, three people. [background comment]

GALE BREWER: Thank you very much; it's an honor to be here. As you know, I am the Borough President of Manhattan; I wanna thank Chair Vacca and Council Member Kallos and all the Members of the Committee, déjà vu; very special to be here. And I think you know; I was a primary sponsor of Local Law 11 of 2012, which we've all been discussing today, the Open Data Law and as borough president I am still paying close attention to the implementation of the law, so I'm very, very pleased that you scheduled this meeting today.

As the Committee knows, implementation of the law has been a major undertaking and DoITT, you

just heard from them, deserves tremendous credit for continuing implementation across two administrations. I am very pleased that Mayor de Blasio has announced the appointment of Dr. Amen Ra Mashariki as the City's new Chief Analytics Officer, who will be looking at Open Data and Minerva Tantoco as the Chief Technology Officer, who will be a cheerleader for the tech sector. So with exceptional leadership at DoITT, the Mayor's Office of Data Analytics, also known as MODA, and the Mayor's Office of Technology and Innovation, I am confident that New York City can continue to lead the nation in municipal data initiatives, as we have in the past.

Now, this is not to say that implementation of the law has been perfect; understandably, some agencies have been slow to embrace a new way of doing business, one where the default setting for government data has switched to open, meaning that we need to make sure that we have information that's really public as sort of a way of saying that. And so for far too long many agencies operated in their own silos, as we know, and were hesitant or outright hostile to sharing their data with other agencies, much less with the public. And

I think most of us in government now realize the shortcomings of that attitude, but change and behavior is difficult. I was really pleased to hear that using Open Data has meant that in some cases, picking up on what Council Member Kallos was stating, some of the FOIL information is no longer necessary because this all was on the Open Data Portal, which is what we had definitely hoped.

So I know it's not easy, as I indicated, for agencies with limited resources to focus on Open Data, so that's why it is contingent on leadership at DoITT and MODA to make it as easy as possible to share data, including automating data outputs, facilitating data mining and creating data visualization overlays whenever possible.

To get a great sense of paradigm shift on how an agency can leapfrog others in relation to open data, we can look at the Police Department, believe it or not. After decades with reputation for being closed off and hesitant to share public data, NYPD is now looking at innovative ways to make their data publicly available and responsive to residents, balancing the needs of confidentiality and capturing operational efficiencies in the process; they've met

with me, members of the civic technology community, who are here today, and I look forward to positive results. It's not easy, because I know their lawyers are always saying don't do this, don't do that, but when law enforcement meets civic hacking, all of us in New York will be better off.

And I just wanna mention, as part of all this, somebody who helped us think about all these issues; how to automate data sets because it's good for the agencies, is Beth Noveck, who was the Chief Technology Officer for the Clinton Administration; is now teaching as a professor in New York, but she always said your best advocates are gonna be the agencies and I think you heard a little bit of that from the discussion today.

Outside of agency issues a major concern as the priority topic of today's hearing, as we indicated, is the Open Data Portal itself. From the beginning amateurs and experts alike have been frustrated by searchability and other functionality issues with Socrata's portal. One challenge during the past two years has been a difficulty in understanding if a search result is official City data or filtered view, and we know, this flaw has

cluttered the portal with multiple views of the same underlying data and I wanna thank, as you heard this today and we know for sure, that DoITT has made simple fixes to the problem. To begin with, making the default search function official data will not only eliminate much of the confusion, helping one query the most relevant data sets on the topic. So I wanna thank DoITT for doing that and making the changes, but there's still things missing. I looked today; I know that the Chair was looking for Department of Buildings, I'm focused on HPD and I was looking for 421-A; I got two sites in all of New York City. So that data set is not controversial; I think should be there, because people wanna know where the abatements are, where there's overlapping abatements; we're all trying to provide affordable housing. So that's an example of I think where you have official data; it's not there; it would be very helpful to New Yorkers.

Another issue that must be addressed is the perceived lack of responsiveness to public requests for data from agencies and DoITT. The public has lots of questions about data sets. Data integrity issues are a real concern, as are requests

from the public for high-value data sets. When we negotiated this bill -- hundreds of meetings -- we were lenient in our timeline for requiring agencies to publish all qualifying data; however, this leniency was due to our assurances that requests from the public for high-value data would be taken seriously by agencies. My understanding is that this simply has not always been the case; this must be improved for this legislation to be successful. I know there are ongoing discussions about how to improve interaction between agencies and the public and I heard today that another person will be hired just to do that and I am very, very appreciative. But if we want the technology sector and the public at large to use Open Data, we must listen to the questions of the public, address them properly and improve our performance. These are basic issues of competence and trust; without them, people simply won't use the site and neither government nor the public will benefit from the promise of open data to improve service delivery, accountability, transparency and performance.

Now I wanna answer the Chair's question very specifically. In closing, I wanna offer a

positive report on what my office is doing with the data. As borough president, one of my responsibilities is appointing members to and provide an oversight of Manhattan's 12 community boards; I know you have them in the Bronx too, but these are the ones in Manhattan. Community boards, as we all know, are New York's neighborhood city halls, but due to their limited budgets and training they have been unable to do complex data analysis. Community boards should be ideal consumers of City data, 311, 911, trends in their neighborhood, construction permits, demographic shifts, zoning and more. We all know that they play an important role in land use review. For too long they have been at a severe disadvantage when faced by complex development proposals; they have been forced to rely on anecdotal evidence often to make decisions that really should be driven by data.

So in order to equip boards with the capacity to use City data sets for decision-making, my office is training interns and they're getting paid from the CUNY Service Corps to work on Open Data core members, to work as Open Data core members in all 12 community board district offices. We expect

them to be assigned soon, because we're training them in our office; they are, as we speak, analyzing all the data, 311 data and other data sets to work on discreet data projects involving such pressing issues that we know -- UPK, Vision Zero, affordable housing, quality of life and more. In the very near future we hope to offer even more dedicated resource to the boards with many of the individuals in this room so as to expand their mapping, the boards' mapping and GIS capabilities using this data so that at every single community board meeting there's a map and a GIS and we know what we're talking about in our neighborhood.

Additionally, we have launched a vigorous and robust training and technical assistance community board leadership series; it's quite popular; 60 or 70 people are showing up at every meeting. On the data and mapping trainings offered, for example, we have partnered with the Department of City Planning to train 200 community board staff, chairs, committee chairs and members to use the community planning resources and tools that the Department has organized, and to the credit of City Planning Commission, in one community portal website,

is excellent. So my goal here is to provide the necessary training and technical assistance so that community boards can produce and submit the most data-driven, visual, updated and supported statement of district need, using as much municipal open data as possible to make their case. Couple of people from Brooklyn came to these trainings, so the ones from the Bronx are always welcome -- I had to throw that out.

Turning the Open Data Portal into everything we envisioned when we passed this law will require a lot more work; however, I am confident that we have the leaders within government, especially at DoITT, as well as from the private sector and nonprofit community; I wanna particularly single out the Transparency Working Group, to ensure that the tool reaches its full potential.

Thank you very much for the opportunity to testify about this bill that I feel so strongly about and I thank you for all the work that you, as Council Members, are doing to make sure it works. Thank you very much.

CHAIRPERSON VACCA: Thank you. I don't have any questions; I think you were great; I mean I

agree with you on so much. I thank you, first of all for your work in the Council when you were here; we miss you, of course... [interpose]

GALE BREWER: I miss you too, Jimmy Vacca.

CHAIRPERSON VACCA: miss you a lot. Also, your modeling technology for Manhattan Community Boards is something that I hope can be replicated... [crosstalk]

GALE BREWER: Yeah, we...

CHAIRPERSON VACCA: it should be replicated.

GALE BREWER: Yes, we have a little... you know, you might have to come to us to get it replicated, but yes... [interpose]

CHAIRPERSON VACCA: I agree. I'm not disagreeing. I had a community board call me recently and stated they wanted more police protection because the numbers in their precinct are not adequate. So I said, well what documentation do you have; have you looked; have you done your analysis? Really they... they just... some people just wanna go into meetings and say we need more police. But people need to... [interpose]

GALE BREWER: You can do that on a... on a wonderful GIS, on a map.

CHAIRPERSON VACCA: But you could do that with so much information that's available; we have to empower people to get that technology and then use it when you advocate to your position.

GALE BREWER: Yeah.

CHAIRPERSON VACCA: Also, when you mentioned HPD, and I know you have to leave, but when you mentioned HPD, we also found faults with HPD in so much as the Open Data Portal was concerned. We have programs like SCREE [sic] and DREE [sic] where the word really is not out there; a lot of people who are eligible do not apply...

GALE BREWER: Right.

CHAIRPERSON VACCA: and here we can get it out in another way.

GALE BREWER: But we could also map for the community board how many applications have come in from that area... [interpose]

CHAIRPERSON VACCA: And where.

GALE BREWER: and where and then you could do an overlay of how many seniors who fit are in that community... [crosstalk]

CHAIRPERSON VACCA: Yes and...

GALE BREWER: that's when you would get action. [interpose]

CHAIRPERSON VACCA: and that would tell you where in that district you have to do outreach... [interpose]

GALE BREWER: Correct.

CHAIRPERSON VACCA: where the community board should concentrate because there are gaps and... [interpose]

GALE BREWER: We'll have that done by the end of the year, just so you know.

[background comments]

CHAIRPERSON VACCA: What? [sic]

GALE BREWER: Thank you.

COUNCIL MEMBER KALLOS: May I ask a question real quick? [sic]

CHAIRPERSON VACCA: Go ahead, yes.

COUNCIL MEMBER KALLOS: I would hate to let Manhattan Borough President Gale Brewer off without a touch interrogation. So first, thank you for passing this law and I just wanted to focus on one of the things you said in your testimony, "If we want the technology sector and the public at large to

use Open Data, we must listen to their questions and concerns to address them properly and improve our performance." Is there any current legislation or anything that the City could be doing in response and... [interpose]

GALE BREWER: No, I think two things; you know the fact that you're having this hearing, I think that the pressure is on the agencies to, like I said, hire somebody who has the public in mind, which was indicated, I think before you got here, by the agencies, they're gonna be hiring somebody to do that, and I think also, you know the Mayor has hired a coupled more people to focus on Open Data, which up to now hasn't been possible 'cause the staff wasn't there. So I think these are some ways that we can [background comment] get more focus on it, and at the same time we need to work in the communities to get it used, whatever that huge number of 30 million users, people like me are going there all the time, so it needs to be unduplicated and there has to be a buzz about using the data.

COUNCIL MEMBER KALLOS: Would opened FOIL be of any assistance in trying to get users... [interpose]

GALE BREWER: I know that you and I have that bill; I think it certainly would help.

[laughter]

COUNCIL MEMBER KALLOS: And with regard to the data and mapping trainings that your office offers, would you be kind enough to offer such a training to all the Council Members at 250 Broadway?

GALE BREWER: Of course and all the staff too.

COUNCIL MEMBER KALLOS: Thank you.

[interpose]

CHAIRPERSON VACCA: I'm going to do that training as Chair of the Technology Committee, so I'm going to [laughter] initiate training for Council Members and their staff regarding this [background comment] open portal.

GALE BREWER: Great.

CHAIRPERSON VACCA: I think it's that important and I discussed it with my staff already. So Gale, we don't need you for that; we need you for everything else you do... [crosstalk]

GALE BREWER: Thank you very much; I like your...

CHAIRPERSON VACCA: which is fantastic.

GALE BREWER: Yes, thank you for much for this opportunity.

CHAIRPERSON VACCA: Thank you always, you're always welcome. I want to introduce Council Member Greenfield and Council Member Weprin, who have joined us. [background comment]

Our first panel will be Noel Hidalgo, Gene Russianoff, New York PIRG, and Dominic Moro, Reinvent Albany. 'Kay. Mr. Russianoff, why don't you go first?

GENE RUSSIANOFF: Good afternoon, Mr. Chairman and Members of the Committee. My name is Gene Russianoff and I am Senior Attorney for the New York Public Interest Research Group; I serve as Chair for the New York City Transpar... Co-Chair -- give myself a promotion there -- I serve as Co-Chair of the New York City Transparency Working Group, which Borough President Brewer was kind enough to acknowledge our organization's civic and technology groups; they work very hard on behalf of this law.

I have four brief points to make; first I want to express my appreciation to the Chairman for holding this hearing; not every law gets reviewed this closely by the Council and I think this is worth

the effort and the attention for making the City a more efficient and responsive place, so thank you very much.

CHAIRPERSON VACCA: Thank you.

GENE RUSSIANOFF: I will also say that DoITT is meeting the spirit of the open meetings law, but... pardon me... DoITT is meeting the spirit of the Open Data Law, but its efforts have been hobbled by insufficient and inadequate staff and I'm glad to hear at the hearing that they're going to beef up; I think it deserves the kind of attention. Thirdly, I'd say in the view of my organization and the coalition members, City agencies should be better following the intent and spirit of the law; you have recognized that this law can help with FOIL and make the data releases simpler. There are vast opportunities for agencies to do more with this law and we hope to see that as time passes on. Lastly, DoITT must make sure the Open Data Portal does not suffer from fundamental usability problems for long periods of time and it's been mentioned the difficulty in the search capabilities of the portal and you know I think that hobbles and it discourages the public from using this great advance. We think

very highly of DoITT's staff, particularly their earnestness and their commitment to progress..

CHAIRPERSON VACCA: Thank you.

GENE RUSSIANOFF: We hope that the bumps in the road are past and we can move forward to its realizing much of the promise offered by this law. Thank you very much.

CHAIRPERSON VACCA: Thank you, Gene..

[crosstalk]

GENE RUSSIANOFF: And thank you for this lovely computer, which I guess whoever sits in this seat gets to take.

CHAIRPERSON VACCA: Thank you. Don't take it home. Yes, sir... [interpose]

DOMINIC MAURO: Uhm...

CHAIRPERSON VACCA: Would you introduce yourself, please?

DOMINIC MAURO: Sure. My name is Dominic Mauro, Staff Attorney with Reinvent Albany. I'm good; how are you?

Well good afternoon, Chairman Vacca and Members of the Technology Committee, my name is Dominic Mauro, Staff Attorney... we've covered that part. My organization co-chairs the New York City

Transparency Working Group and is an active part of the broader civic coalition working on open data policy in New York City and New York State.

We join with our colleagues in thanking you very much for holding this oversight hearing. Council oversight is extremely important for ensuring that the hard won Open Data Law is complied with by city agencies.

Overall, we are excited about the progress of the Open Data Law, which is still in its early days. We believe the City's open data efforts are robust, healthy and improving. The implementation of the Open Data Law is fundamentally strong, though much work remains. First, here's what's working: 1. The City has five fully funded, dedicated open data staff. This is probably more than any stat or local government and gives the City the capacity to rapidly improve and expand it's open data efforts; 2. Open data has strong support from the Mayor and City Council and has a synergistic relationship with the Mayor's Office of Data Analysis (MODA) which ensures it's sustainable; 3. Open data has an expanding group of public stakeholders, including businesses, academics, advocates and

government who use the data; 4. The quantity and quality of the data available continues to improve, and with it, so does the use of that data.

We especially applaud DoITT for its sincere and energetic efforts to get agency data published. DoITT is clearly meeting the spirit of the Open Data Law and their staff has strived to make it work.

The Open Data Law calls for a new way for government to share information with the public, going forward, forever. One way to think about it is that we are in year three of a permanent change lasting for decades or centuries. Every year the City needs to get a bit better at Open Data and fortunately, we have the resources and commitment to succeed.

In the next year, we hope the de Blasio administration advances open data in four key ways:

First, the Mayor's Office of Operation needs to get agencies to put the most frequently FOIled and requested data on the Open Data Portal. Most agencies still do not understand that the Open Data Law is intended to help them reduce FOIL requests, reduce 311 requests for information and

help them get information to the public at a lower cost with less hassle.

Second, DoITT needs to create a public feedback process for the Open Data Portal which results in more and better city data being put online. When data errors are discovered and reported by the public, the responsible agency should correct those problems. DoITT should ensure that the public can track the progress of these corrections.

Third, the severe problems with the search functions on the Open Data Portal have to be completely fixed. Despite repeated and emphatic requests from the Transparency Working Group and many other open data stakeholders. It took DoITT more than two years to correct crippling flaws with the search function -- flaws that severely reduced the usefulness of the Open Data Platform and undoubtedly kept people from using the data on it. This kind of fundamental usability problem cannot be repeated and cannot take so long to fix in the future.

Four, DoITT needs to clearly show the status of data sets to be published or data sets which have been delayed or removed from the Open Data Plan. Overall, it's hard for the public to tell if

the City is meeting its own data release targets.
For example, a data set essentially disappears from
view if it was scheduled for release in the 2013
plan, delayed and then not included in the 2014 plan.

DoITT and City Hall are easily capable of
achieving these goals and we look forward to
assessing their progress with you at the 2015
oversight hearing.

Thank you.

NOEL HIDALGO: I'm gonna take the opp...

[crosstalk]

CHAIRPERSON VACCA: Thank you.

NOEL HIDALGO: I'm gonna take the
opportunity to use the computer, which is a rarity...
[interpose]

GENE RUSSIANOFF: You have my permission.

[laughter]

NOEL HIDALGO: Yeah, do I -- which is a
rarity at hearings... [interpose]

CHAIRPERSON VACCA: Can you identify
yourself for the record first?

NOEL HIDALGO: My name is Noel Hidalgo --
let's hope this has internet -- My name is Noel
Hidalgo; I'm the Executive Director of BetaNYC. If I

can figure out how to use Windows -- which is not my strong suit... [crosstalk, background comment] as I will try to multitask.

So you have several pieces of paper in front of you; the first I wanna point out is this broad sheet, tablet, which is 60 databases that our community's asking for in an updated file format. This is just the beginning of our review of the City's open data mine, but the general summation of this very fine print, which we will share to you in very large website, is the fact that we want to see the City's Open Data Portal as useful as possible, and this is just the first attempt to get there. And I'm not sure if I know how to use this computer -- there we go.

So as we are the nerds of this particular table, the technology nerds, we bring you a technology tool, and I will talk about this particular tool in my testimony.

My name is Noel Hidalgo and I'm the Executive Director of BetaNYC. BetaNYC is the city's civic technology group. Through our weekly meetings we explore how to make the city's data useful to its neighbors. Our goal is to demystify technology,

design and data to City Council Members, community boards, community groups and businesses alike. We are a community of developers, designers, mappers and policymakers who volunteer our time to improve the digital city. We are New York City's civic hackers and we love open data and open government, and because of the City's open data program and our advocacy for better data, our community has grown. In the last 12 months we are now over 2,100 members. This year three of the four BigApps winners were community members, 45 percent of the semi-finalists were BetaNYC community members. As we discuss the City's data, we discover more and more individuals who are looking to solve problems with the City data and that in turn leads us to more and more feedback about the City's open data program and the City's open government practices.

Today I wanna ask one simple question -- how does the City, together with her constituents, build the right open data ecosystem?

From BetaNYC's point of view, we need three things: adequate human resources, meaningful data and improved feedback infrastructure.

We're really excited about the appointments of the DoITT Commissioner, the CTO and the CAO; we want to thank the Mayor's Office for appointing such a talented leadership to run the City's information technology and data infrastructure and we hope that the three of them can staff up their teams as soon as possible. From BetaNYC's point of view and from previous testimony, we don't have time to lose.

As we pointed out in last year's report, *The People's Roadmap to a Digital New York City*, the City's open data program is on the verge of revolutionizing how we access municipal transactional data. We are at the point where public consumption of data is blossoming into a beautiful garden of data, utility and stories.

Second, the City's open data program is starting to breathe meaningful data into our lives. Over the last year, a few key data sets have been operationalized and placed online with daily updates. This operation allows for the public to derive daily situation reports and this is a good start, but we need sub-daily reports.

Today, in partnership with Code for America, the City of Charlotte, North Carolina and the City of Lexington, Kentucky, BetaNYC launches Citygram.NYC, that's our dot NYC domain, which is the website that's above our heads.

Citygram is a tool that converts municipal open data into meaningful notifications. Similar to Notify NYC program, you can subscribe to topics and locations. These locations will arrive via SMS or email. We wanna point out that if you subscribe via SMS that you lower the subscription to just one-tenth of a mile, because otherwise you will have over three minutes of SMS alerts coming into your phone; that's something that we're looking to fix. But as this is a free and open source tool, you can tell us what features are needed and we'll add them in the coming months.

Today Citygram.NYC starts with NYC 311 data and NYPD crash data. We must not that the NYC 311 data set is one of the best. The NYC 311 provides an unparalleled and timely view of the city and we hope that many other data sets learn from 311. On the other hand, the NYPD data is better than it was last year, but it still needs significant

improvement. NYPD data is wildly incomplete and contains notable discrepancies, and I have quotes from transportation alternatives on that in the testimony. That being said, we're excited to work with these two data sets and make open data meaningful.

One year from now we will live in a city where we can receive push notifications on NYC's public meetings and procurement notices. Currently BetaNYC is working with DCAS and DoITT to put the City Record and its data into machine readable formats. For this to become a reality, we need a sustained, open effort from DoITT and DCAS as we build a data format that maximizes utility.

You were wondering about community boards, well I can say that Community Board 6 in Manhattan has taken the City's Open Data Portal and embedded it into their website. On the CB6 website we see contextualized NYC 311 reports Road Repair Requests, Broken Parking Meters and Noise Complaints and this contextualized data actually puts CB6's website as one of the top referrers to the City's Open Data Portal and maximizes the public's understanding of issues related to their community.

Just to talk about one other story is the fact that there was a blog that was launched in February called I Quant NY. I Quant NY is third behind Gothamist and Al Jazeera in pushing traffic to the City's data portal, I Quant NY is the tenth highest referral to the City's Open Data Portal and it beats out the City's own open data success stories blog.

I Quant NY refers more traffic to the City's data portal than other data-friendly media outlets, like WNYC and TechPresident and it also refers more traffic than social sharing sites like Reddit, GitHub and Facebook.

I Quant NY is so good at humanizing the City's open data story that Ben Wellington, its editor, has been featured on a number of press outlets recently and since launching in February, his site has received over half a million unique visitors and 1,500 Facebook likes and this is about as many Facebook likes as the New York City's Department of Consumer Affairs.

I wanna read this quote from Mr. Wellington.

"There are two sides to the coin.

Anytime you point out something, it could go either way. If you tell the Department of Health that there's something wrong with the rating system, they could say, "Wow, let's look into that" or they could play defensive. Generally, agencies are defensive, but there's also not a good mechanism for them to take in information like this. They get caught off guard. I hope in the coming years they build in ways to reach out like this. If there were a liaison I could reach out to, maybe I would go that route. But right now, the only way to get attention is through the media. Unfortunately, that creates an adversarial relationship, which I think is the wrong way to look at open data and I really believe that if you empower people, you'll be able to get much more than you will through criticism."

That being said, the effective way to build the right open data ecosystem leads the City and her constituents to develop a shared understanding of possibilities and collaborate on shared outcomes. We need the City to finish hiring its open data team and we need to have a shared view of that particular garden.

After two years of planting data seeds, it's time to harvest the garden and build the type of data ecosystem we all want. BetaNYC is honored to represent the open government data users and we're ready to work with the City to build an equitable and just 21st century. And attached to this you will find more or less all of the data that we're asking for to be improved and we look forward to doing that.

CHAIRPERSON VACCA: Alright, thank you. Thank you all. Councilman Greenfield has a question.

COUNCIL MEMBER GREENFIELD: Thank you very much. I want to thank you all for the work that you're doing and for your advocacy; I appreciate it. What was that website that you mentioned, by the way? I just haven't heard that before.

NOEL HIDALGO: I Quant... [crosstalk]

COUNCIL MEMBER GREENFIELD: Can you spell that out?

NOEL HIDALGO: I Q u a n t N Y.

COUNCIL MEMBER GREENFIELD: Got it. So you mentioned before; I think you used the term somewhat loosely, that you guys are "data hackers," right; is that what you said?

NOEL HIDALGO: Civic hackers.

COUNCIL MEMBER GREENFIELD: Civic hackers; it's a good thing, not a bad thing, right?

NOEL HIDALGO: Correct.

COUNCIL MEMBER GREENFIELD: So can you tell us a little bit about some of the data that you found, 'cause you've been pouring through a lot of this data; anything of note that sort of jumped out at you that was instructive or helpful?

NOEL HIDALGO: As I said, the NYC 311 data is really, really helpful, it provides a really unique view to the City. The Department of Education last year released its schools in a machine readable format that enabled -- we have a few different tools or a few ways to access now that telephone book that students and parents would get on what to access and that's been a really amazing resource that the Department of Education has now been promoting like 15 different websites that's out there. In addition to that, after-school resources has been released, which is an amazing way for students to find out what are the community kind of services that are offered next to their school; traditionally that too existed in a telephone book-like pamphlet and it was really

hard to search. Those are just a few of the highlights. I will say that our advocacy with John Krauss, who's in the audience, Transportation Alternatives, who couldn't be here, really fought hard to get the crash data into a more useful format. We're happy to have that data that's out there; we know that that's gonna help save lives, but we need that data to be better and that's purely dependant upon police officers inputting that data accurately. And we're waiting for that day that the NYPD goes out and educates police officers, so that way we can know how dangerous and what are the incidents that cause crashes in our neighborhoods, so that way we can build safer streets for everybody.

COUNCIL MEMBER GREENFIELD: Excellent point, especially as how the Mayor signed our legislation today to lower the speed limit in New York City to 25 miles an hour, effective November 7th.

Let me ask the panel just a broader question about this issue in general. How do you feel, from your experience; -- talking about the state and the federal government -- how do we match up New York City with other cities and with the state

and federal government in terms of our open data policies and more than the policies, 'cause of course we have policies, but then again, the implementation of those policies aren't necessarily recognized in the actual policy, so what's your view on that?

DOMINIC MAURO: Sure. Well as regard to the federal government, their open data portal is set up in an interesting fashion; they don't centralize all their open data publication through a single federal agency; there's no federal IT agency tasked with this. Knowing the federal government, there's about four agencies they could, but they're not. So essentially the federal portal knows where to look for data sets that get published by agencies and so agencies are responsible for publishing their own data in a format that the portal can go out and find -- they call it harvesting; you can call it federating. So the federal government avoids a big bottleneck that affects you know, state and city portals where there's a single agency that has to sort of clear all this data for publication. And as far as the state's open data portal, I mean it runs on the same software that... it's the same practice. You know, Andrew Nicklin, who used to work for DoITT,

is now running the state open data efforts in the same way, so I think they're roughly comparable. I don't know off the top of my head how many data sets we have compared to them, although I expect that I'm about to know. [laughter] But I think we're doing great, because agencies in the City, they're really committed to this, so we're very happy to see so many agencies, you know, participating voluntarily, like earlier testimony, the uh; it was the School Construction Authority, I didn't know they weren't covered because they're on the Open Data Portal, so it's great to see those sorts of things happening.

GENE RUSSIANOFF: And this is really an answer to your first question; we're really not civic technologists; what we are is trying to dig up the information for people who can make good use of it. So we focus a lot of our attention on the public authorities, particularly the Metropolitan Transportation Authority...

COUNCIL MEMBER GREENFIELD: Sure.

GENE RUSSIANOFF: and five years ago they were suing people who wanted their data, trying to prevent its disclosure and they finally realized that they could gain huge advantage by having people who

are smart and had the time for free make applications for them. But it's a slow process and I could tell you that, for example, we tried to get them to release their complaints from the public and we'd be interested in visually mapping where people are having particular difficulties and we're now in year two and they're still resisting, although they did make their lost and found database available and it's quite unusual what people lose on the subways.

[laughter]

NOEL HIDALGO: I'll just point out another website. So the Open Knowledge Foundation and Code for America and the Sunlight Foundation has put together an interactive census of municipalities; New York City just recently took the number one spot in regards to its data; that does say that we're at the very top of the list, but sometimes I would say that we're not number one. When you look at actually how the data is released and the usability of the data, that's where things fail. We're proud and honored to be the number one spot and to have that much data, but we really need it in a better format. I was glad to hear Gale Brewer's testimony that the Manhattan Borough President's gonna be reviewing

those data sets; this is something that we were thinking about launching in February as part of one of our community events, is to go through every single data set and look at the exact structure of those data sets and how good they are. But from our particular point of view, New York City really is the best because of this well-developed ecosystem inside a government and outside a government.

COUNCIL MEMBER GREENFIELD: Great. Thank you all so much.

CHAIRPERSON VACCA: 'Kay. In your testimony you spoke about documentation; can you go over that; how can we improve documentation and what kind of documentation are you speaking about?

NOEL HIDALGO: Well that's kind of a big issue. In general, if you go to the City's Open Data Portal, you're essentially presented with an Excel spreadsheet and if you're a savvy or you know, a subject matter expert, you can kind of read into that data immediately. What we don't have are read me's or how to's to kind of explain the data. There is a button that you can push that's kind of hidden that will talk about the data description, but in general, in our world we talk about schemas and we talk about

just, you know general read me and documentation and that's something that's lacking; the federal government does a better job of displaying what's in the data, kind of what are the data columns, what do they mean, where are they coming from, how frequently is the data updated and that's something that I wish that New York City did a better job of.

CHAIRPERSON VACCA: And basically you're talking also about simplifying access, simplifying procedurally how people can get the information.

[background comment]

NOEL HIDALGO: I mean in some ways it's explaining what's... [interpose]

CHAIRPERSON VACCA: Yeah.

NOEL HIDALGO: in that data... [interpose]

CHAIRPERSON VACCA: Right.

NOEL HIDALGO: and if we had the data in a more usable format -- one of our other calls is that we want all data that can be geo-tagged, which means have a GPS coordinate, more or less, attached to it. The City last year produced an amazing database that translates all of the different definitions that the agencies use to mark up the City into a universal format; it's an amazing tool; I mean

it's absolutely amazing; can't thank the GIS Department at DoITT... thank them enough for creating such a database and that was some of the insight that was derived from Sandy. But we need to have that API, more or less, applied to the data catalog as those data sets are coming in so that way we can immediately turn it around and make that data more useful.

CHAIRPERSON VACCA: Councilman Kallos.

COUNCIL MEMBER KALLOS: This is for the entire panel; all of you brought up an issue regarding searchability for two years on the Open Data Portal and NYPIRG made a very strong recommendation and in fact, it seems like the panel members feel like it may have been longer than two years; if somebody could just elaborate on the issue, how it was corrected and any additional issues that remain.

DOMINIC MAURO: So as regard to the search, yeah, since we can tell from day one, the issue affected the panel. Essentially, when you searched for something like 311 on the portal, you saw him do his presentation; the official 311 data set didn't show up until the 214th listing,

apparently. And that's because when you were searching for 311, the portal can't search for information just in the title of the data set or the description of the data set; it's searching the entire contents of every data set simultaneously. So any school that has exactly 311 kids, any police precinct where there've been exactly 311 assaults or something like that, those come back when you search for 311. So this is... it happens a lot. So you saw again, there were nine pages of useless results. And so now the search has been fixed a little bit. So what happens is; it's still searching through the entire contents of every data set simultaneously, but now 311 that appears in the description or the title of the data set is more highly prioritized. And so you're still going to get hundreds and thousands of results for 311 on the Open Data Portal, but now those show up at the end of the list and the official 311 data set shows up at the stop. And the other improvement that goes along with this is; now you can only view, if you so choose, you will only see official data sets and not those filtered views. So people have... you know, community boards have created filtered views for their specific region of the 311

data sets and now you can just see the actual 311 data set instead of someone else's filtered view of it.

NOEL HIDALGO: I just wanna take one moment and say that there's also an issue of resilience, the reason why it took so long. So some estimates it's kind of three years that search was broken since the City implemented the Socrata platform and you know, for some particular reason there was just not an effort or kind of a desire to fix this, though it seemed like there was a desire to fix this. And when the last administration was leaving office there was the key people who were in charge of helping make some of these decisions; they left, and so there was a deficiency in talent. And so this is one reason why we're really excited to have more people in the Open Data Program, because that way you have this resilience within DoITT, but in some ways we kind of lost a year, maybe two; maybe three, because of just the fragile nature and the limited resources that the NYC Open Data team has and so hopefully, you know like any good practice, you're developing muscle and that muscle is the actual talent and the NYC Open Data, because it's so

popular, we really need more resources, more bodies that we can depend upon to make all of this more useful and more resourceful. This unification of the search bars between type, category, agency and source and then search field, we just moved them all together, because that was like a logical thing to do and it was amazing that it took us two years to finally sit down with DoITT and have that conversation and just pull up the website and be like, look, this doesn't work; how do we get this fixed? And then they went through and said hey, you know, Socrata; can you please get this fixed? And it was remarkable how quickly they were able to get that fixed, but it was remarkable how long it took us to have that conversation, that simple, basic conversation.

COUNCIL MEMBER KALLOS: Do you believe that this administration, with a progressive City Council a the progressive Mayor is being more responsive to those conversations, where you feel less inclined to have to go to the media in order to get a response or what are your findings and sorry to ask such a difficult question? [interpose]

NOEL HIDALGO: Yeah, I'm not sure if this comment has been approved by my sponsors. I think that this administration, as it has staffed up, has been remarkable, and have to give them compliments for being open and welcoming to our ideas in conversations. And so it's kind of amazing to hear; they may say, we don't have the resources to do that, but at least we're putting it down on an issue queue and so that's something that I think we just didn't have enough people to do under the previous administration.

GENE RUSSIANOFF: And you know, we make our judgments of the Council and the Mayor as we go along and see how they focus on various policy issues, open FOIL and making the process more open, transparent and relevant to the public got some hot press [sic] and so far, you know we could be doing better.

COUNCIL MEMBER KALLOS: Gene, in your testimony it said if DoITT cannot get a problem fixed by a vendor; that vendor should be fired; we strongly recommend that DoITT begin vetting other vendors so it is not locked into a non-responsive vendor that treats New Yorkers with such contempt. Is that so...

GENE RUSSIANOFF: It's between too tough and over the top, but it may be because my home address begins with 311, [laughter] but there's a lot of frustration in the civic community about how long this has taken and how fundamental it is to the workings of the portal. I mean, you know, I have a bad search engine on my office computer and it just drives me wild, so this has driven a lot of the interested public pretty crazy.

COUNCIL MEMBER KALLOS: Google Drive is amazing, so I guess the question you'd like to see is; you would like to see DoITT opening up the Open Data Portal contract for bid and even looking at whether or not we can build something free and opened source in-house or even using the federal government... [interpose]

GENE RUSSIANOFF: Correct.

COUNCIL MEMBER KALLOS: And then, just wrapping up with BetaNYC, in your testimony; I just wanna make sure the record notes, you have a list of questions for the administration and I would just ask that in the administration's response to those questions that my office be copied and I'm assuming

the Technology Chair would also like to be copied on that response.

CHAIRPERSON VACCA: You assume? I'm the Chair of the Committee; I would, so there's no need to assume.

COUNCIL MEMBER KALLOS: Perfection. And then with regard to the 60 data sets, I had asked about the budget, but how many of the 60 data sets do you believe at PDFs versus just raw data, and then I see a preference for APIs, but is an API to exclusion to have the flat data dump, as it were?

NOEL HIDALGO: Can someone take note with all those questions? Uhm so... [interpose]

COUNCIL MEMBER KALLOS: Sorry.

NOEL HIDALGO: let me try to address them one by one or... yeah, and remind me which ones I don't respond. [interpose]

COUNCIL MEMBER KALLOS: You got it.

NOEL HIDALGO: So we have a list of recommendations in our testimony; I will flag some particular parts of the... In regards to the questions, because there wasn't a venue, an open venue where we could talk about usability; you know, we're fortunate enough to have access to have dialog amongst

ourselves around usability of the website and have access to the City; you know, we're not the only users in this; in our community of 2,100 members, they had a lot of feedback and a majority of these questions come from them and so we stood up our own little working group to address quality of data and additional questions that we wanna ask around of the Open Data team. And so there is a Google group that you can join and your staff can join and interface directly with the public and ask them questions of things that you want to see done with the Open Data site and we've had conversations with DoITT and they're open to joining that as well and having that conversation in open, but we will gladly share any responses with you.

In regards to data, you know, flat data in a file, sometimes as great, as transparent as it is, is still unusable; it's great that the City's vendor Socrata has a native API that once you put the data into that system we can build applications on top of that. Citygram is powered off of the Socrata API because of that; is that correct?

MALE VOICE: Correct.

NOEL HIDALGO: Okay. And so, you know, whatever data goes into the City's data portal automatically has an API that we can access. But there are systems that just may not work well for flat data into the Open Data API or you know, the Socrata Portal API. As I talked about the geo location and you know, that's an excellent example of -- you're trying to triangulate a lot of different information to come back with coordinates. Another example of that may be the City Record and we're trying to explore that conversation right now of procurement notices and public hearing notices and all the other things that kind of go into that, and the City Record is kind of this weird, unique beast for the fact that its workflow pulls in PDFs and so we're not sure if the Socrata tool will elegantly support that and so we're exploring different options around that.

In regards to the 60 data sets, I wanna point out; beyond the crash data, which is something that we pulled out, is NYCHA data; it's appalling at times to read kind of what New Yorkers who are most vulnerable have to live in; it's appalling that NYCHA is roughly the size of the city of Oakland and we

have zero transparency in regards to the issues and problems that those individuals have. The Comptroller's most recent report referred to data that was two years ago and you know, we're civic hackers; we wanna improve the city, we wanna improve the city for everyone; we really need access to NYCHA data and I don't know how to get to it if the Mayor's Office keeps on saying that that's out of our jurisdiction, but you know, that's a really tough data issue and the 60 data sets kind of go into that; they talk about police and crime and public safety data that routinely isn't normally shared, but we need that data to make the environment more useful.

CHAIRPERSON VACCA: Well I thank you. I did bring up these issues in the beginning of the hearing today; whether or not NYCHA and HHC and Rent Guidelines Board are included or not, kinda becomes a superfluous issue because if the Mayor's Office tells them you're included, they will be included. So more than looking for a legal interpretation, I would expect that since we're going into this age of transparency, that all agencies be compliant. I want you to know I've been reading the City Record for about 25 years every day. I read the City Record

every day. Now, most people will be bored to tears when they read the City Record [laughter] every day, [background comment] but do they have the right to access? Of course they do and I think putting it online is a good thing. And those will really be kind of people that relish knowing everything about everything. But that's what we're here for..

[interpose]

NOEL HIDALGO: Thank you.

CHAIRPERSON VACCA: to empower them.

Thank you very much. [background comments] Our next panel. John Krauss... oh, Christine Berthet, David McCreery and Joel Natividad. [laugh] Christine, why don't you go first, okay?

CHRISTINE BERTHET: Thank you.

[background comment] Yes; one second, here it is.

Well dear Chair Vacca and Committee Members..

[background comment] good... thank you for holding this important hearing. CHECKPEDS is an advocacy group for pedestrian safety in Manhattan. The Vision Zero initiative is attempting to focus on the root cause of the crashes and fix them in order of priority. Unfortunately, the data collected and reported by NYPD is so incomplete that I don't think it will help

Vision Zero. For example, 45 percent of all collisions are untagged for boroughs, so which means that each borough has inaccurate information. Forty-five percent; I'm talking about enormous proportion. How is that possible; how is it that an officer doesn't know in which borough he his? I mean something is wrong with that, right; why isn't it tagged; that's a very basic, basic data issue.

Second, for 60 percent of all city collisions, the contributing factors, which is really the thing that tells you, well somebody was not yielding or somebody was going too fast, etc., is not reported by NYPD; I mean 60 percent is not reported by NYPD. So we are making decisions with less than half the proper data. In Manhattan, more specifically, which I looked at in detail, NYPD captures contributing factors only in 42 percent of the crashes with pedestrian injuries, but for motorists injured, the information capture increased to 54 percent. Why is that? Very interesting.

So we urge you to direct the NYPD to tag 100 percent of their collision by borough and to report contributing factors in 100 percent of the crashes where we have injuries or fatalities; not the

other crashes, right? After all, those kinds of crashes represent only 20 percent of the totality of the crashes, so it's kind of a small proportion, but they are the ones that need to be really tagged properly.

At the same time, NYPD could reduce their involvement into the 273,000 monthly collisions when there is only vehicular damage and where they collect 49 percent of the information. Why is NYPD collecting information about vehicular information, vehicular crashes without injuries? I mean are we doing the job for the insurance industry? I don't think our taxes should pay for that.

The City Council and Mayor de Blasio have voted stronger laws for drivers who injure or kill pedestrians and I applaud that. I would also recommend that for each collision with an injury or death the police collect and publish new information, which is the summonses or the penalties meted out to the responsible party; that information is not published and we have a brand new law for that.

So I want to thank you for this hearing and these are my suggestions.

CHAIRPERSON VACCA: Sir.

JOHN KRAUSS: Hi. Good afternoon, Council Member Vacca. I'm here representing The Governance Lab or GovLab at NYU Poly, as a Research Fellow there. Our mission is to improve people's lives by changing how we govern.

We were founded in 2012 with funding from the MacArthur Knight Foundations and we bring together thinkers and doers who design, implement and study technology enabled solutions that advance a collaborative network approach to reinvent institutions of governance. We think that open data is a critical element for improved governance; the current Director of GovLab, Beth Noveck, was co-architect of Data.gov and co-author of the U.S. Open Data Policy and advisor to NYC on the Open Data Law. In April 2013, we released the first comprehensive study of companies that use open government data, the Open Data 500.

From our experience and research, we can say that an engaged community is a key element of the successful open data strategy; the community can provide capacity, feedback and also data itself back to the city. In our view, it is difficult for the NYC open data community to have their questions about

NYC data answered or track the progress of those requests within the City's Open Data Plan. We understand that this is a hard problem, but we have ideas about how the City could improve its facilities, leverage the community and maximize the value of its open data.

A successful community engagement strategy comprises active involvement in both physical and online communities from City Open Data officers, in addition to improved idea exchange and accountability mechanisms, and we've numerated a few.

First, a solid tool where it's easy to find data sets, easy to report feedback or issues about data sets and easy to suggest new data sets. For example, right now it can be difficult to quickly query for existing data sets, establish whether or not the City's even publishing the data that you are querying for or to understand whether returned data is relevant to the search. Despite the fixes we have seen, often doing integral [sic] search restricted to the portal leads to faster and sometimes more relevant results than the built-in search feature.

Second, a clear and transparent process to report issues, track progress and understand

outcomes and timelines. For example, the City removed many data sets in the 2014 updated Open Data Plan with minimal explanation; for example, did not qualify as data and without tools in that publication for tracking changes versus the prior data plan; there are free and open source tools available to facilitate tracking changes and comments and we would be enthusiastic about the City adopting those tools in future publications of data plans.

Third, a real human being you can talk to who's accessible either via direct email or during face to face meetings unless accountable. For example, in Chicago, open data officers regularly attend hack nights regularly, which facilitates communication with the community and we're glad to hear that the staff at DoITT and MODA are being beefed up to enable that sort of participation, note that they are enthusiastic about the community here.

Fourth, conversation tools for the various parties. For example, city, business, civic hackers and the community at large, thus the community questions. There are such forms already in existence which can be created using free and open source tools, but integration of those existing

discussion tools would be essential in allowing feedback to take part in the actual Open Data Portal.

Fifth, a feedback loop mechanism for open data consumers to report successes, which would motivate City personnel at agencies into doing more work on open data. The current system imposes a wall between the agencies publishing and producing open data and the consumers on the actual website.

In terms of what GovLab could contribute to improving the Open Data Portal, we'd be happy to draft a report on best practices for community engagement around open data and we're also located on the same floor of Brooklyn MetroTech as DoITT, so our door's always open and we are also happy to walk across the hall and share some of our research and expertise with the City.

CHAIRPERSON VACCA: Thank you so much.

[background comments] Sir.

DAVID MCCREERY: Thank you and good afternoon. I am David McCreery, a resident of Manhattan and a hacker. This year I took part in the New York City BigApps competition. One team I led won a grand prize, a second team I created is still waiting for data from the City we requested via FOIL

over six months ago; we were eventually outright denied by the City. The data in question is not controversial; its release will not embarrass anyone; in fact, it is something that could prove essential to the City's goal of economic development in the outer boroughs. I'm speaking about the Walk NYC mapping GIS data; this is the data used to create the five-minute radius maps found in Citi Bike stations, subway stops and bus shelters. The data should be fully available on the City's Open Data Portal. Rather than open the data up, I've been informed that the Department of Transportation is taking steps to close off this data. The DOT is seeking protection under Copyright Law to prevent anyone from making maps similar to their Walk NYC maps.

I am especially concerned because the Walk NYC program itself incorporates locations from open data maps created by DoITT and other City agencies. We may face a situation where the DOT uses Copyright Law to shut down hackers who are creating innovative apps using copyright-free DoITT data. As a matter of public policy, it is better for the City to encourage entrepreneurship and job creation around open data instead of looking for ways to restrict and

license its data. I hope the City officials here today will fully consider my concerns, remain focused on creating the greatest amount of openness possible for City data. Thank you.

[background comments]

JOEL NATIVIDAD: Honorable Chairperson, I am Joel Natividad, Co-Founder of Ontodia, an open data startup borne out of open data ourselves, after winning NYC BigApps. I'm also a proud member of BetaNYC, the largest Code for America Brigade in the nation.

I strongly support the City's open data efforts and acknowledge that the City has made great progress thus far, but I know we must do more to realize the full economic potential of open data. In a study released last October, McKinsey estimates that the open data industry has the potential to unlock \$3-5 trillion in economic value per year in the U.S. alone. In New York much focus has been placed on hackathons to unlock that potential and I'm a product of that.

But beyond that there is limited operational usage of this great resource that some have come to call the new oil of the information age.

Last year, as an experiment... [lights went off]
[laughter] sorry... [laughter, background comment] I
guess that took you off... [crosstalk]

CHAIRPERSON VACCA: Somebody thought this
meeting was over already?

[laughter]

JOEL NATIVIDAD: Last year, as an
experiment with... [interpose]

CHAIRPERSON VACCA: No lights means no
transparency; you know that.

JOEL NATIVIDAD: No sunlight.

[background comment] [laugh] Last year, as an
experiment with DOE, iZone, we helped them release
open data about the high school handbook, which Noel
mentioned earlier; we helped them refine that data,
taking spreadsheets primarily used to produce the
PDFs and printed phone books, which nobody hardly
reads, except some people here and... [interpose]
[background comments] [laughter] That was meant as a
compliment. [background comment] [laughter] But as
you know, need to modernize that and we need to
produce clean, ready-to-use data supplemented with
additional data, which technologists call metadata,

so people can build great innovations with it, using an application programming interface.

Since we helped DOE, several organizations have used that data to create high school choice engines, with WNYC just releasing one earlier this month, the SchoolBook project. Even then, WNYC stated that News Team had to spend several days manually rekeying in some data, even with this great resource, as most of the City's open data is still in the former PDFs and raw data dumps. What if these other data sets had an API and additional metadata; will it make the process of updating the handbook much easier? We believe so. Will it make the data more current; instead of being six months betaed [sic] and allow parents and their children to choose high schools the same way they do everything else online? We believe that this is achievable and NYC is uniquely poised to achieve this vision of government as a platform.

As it is, New York City is the center of several information centric data-driven industries and we feel that we can become the center of this API-driven innovation with your support. Right now I am a proud member of the City Record Online Working

Group, made possible by the City Record Online Law passed by the Council this August. We are working shoulder to shoulder with DCAS in a landmark public-private partnership to unlock the City Record. At the moment the scope of our involvement is limited to making sure that the beta be machine readable and available for bulk download by August of next year. We believe that we have a golden opportunity to create something more than machine readable data available for bulk download, as the law prescribes. If we took the added step of creating a City Record API, we're convinced that not only will we fulfill the goal of the law; it will create a foundation for high impact innovation that will truly set NYC apart.

My co-founder and I quit our jobs to work in this new field because we strongly believe that NYC will create yet another data-driven industry and will be the center of civic tech innovation. It will not only create new jobs; it will also make way for a happier constituency as they see city government keeping pace with all the private sector data-driven innovations that is now the norm that they expect in their daily lives. I ask that the City Council expand the DCAS mandate to create the City Record

API. We've been meeting with DCAS on a regular basis and the feedback we've been getting is, even if we offer to help to create this modern interface that that's not what they were told to do; that was not the mandate of the law. You can be sure that Ontodia and as a proud member of BetaNYC we will create unparalleled innovations by the City Record Law's first anniversary next year if this is done. Noel just showed you a preview of City Plan; we want to have alerts from the hearing to the procurement awards, everything; if we can have that without me having to read through the City Record, which if you've seen it, is something that will really put you to sleep. [background comments] Yeah. Thank you.

CHAIRPERSON VACCA: You make some very good points. Thank you so much and thank you all for your testimony today; thought the hearing was very, very good. I think the Chairman did a good job of running the hearing today; [laughter] I have to say so myself. But I thank you. There being no further speakers; no further business; it is now 3:10 p.m., I thank you all for coming and this hearing on the Technology Committee is hereby adjourned.

[gavel]

C E R T I F I C A T E

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



Date October 31, 2014