CITY COUNCIL
CITY OF NEW YORK

----------X

TRANSCRIPT OF THE MINUTES

of the

COMMITTEE ON TECHNOLOGY

----------X

June 21, 2010
Start: 10:45am
Recess: 12:40pm

HELD AT: 250 Broadway
Hearing Room, 16th Floor

B E F O R E:
Daniel R. Garodnick
Chairperson

COUNCIL MEMBERS:
Gale A. Brewer
Mark S. Weprin

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

Carol Post
Commissioner
Department of Information Technology and Telecommunications

Don Morrow
Chairman
New York Technology Council

Aaron Brown
Senior Product Manager
Google

Andrew Hoppin
Chief Information Officer
New York State Senate

Liz Hodes
Digital Democracy

Ben Berkowitz
Co-founder and CEO
See, Click, Fix

Phillip Ashlok
Open Government Program Manager
Open Plans

Rachel Faust
Policy and Research Manager
Citizens Union of the City of New York

Sam Brookfield
ITAC, New York City Industrial Technology Assistance Corporation

Tim Hofer
Director of Operations
Manhattan Institute

Andrew Brust

26 New York

Todd Stavish
Socrata

Ray Garcia
Executive Advisor
Field Center of Executive Entrepreneurship at Baruch College

Richard Stanton
CEO of Bintro

David Weber
Senior member of the ICM

Diana Vitetti
Associate director
Common Cause New York

Tom Lowenhoff
Director
Connecting.nyc

Dylan Gelts
Roadify

COUNCIL MEMBER GARODNICK: Good morning everyone. Welcome to the committee on Technology of the New York City Council. My name is Dan Garodnick, today's date is June 21st. I am joined today by Council member Gail Brewer and Council member Mark Weprin. I want to welcome all of you to today's hearing on open data standards for New York City government.

This is, in my view, perhaps the most important transparency legislation that we have heard in the Council in years. It will encourage far greater public engagement in New York City government. Information will become more directly available to residents and to the internet developers who could bring a vast number of apps to the marketplace to aid New Yorkers in accessing that information.

The open data legislation, Intro 29, introduced by Council member Gail Brewer, requires the city to create a single internet portal from which all public data sets can be accessed in raw form via web browsers and mobile devices. In simpler terms, if government has the data and it can be made available, it should be

made available in an unlocked and searchable form to everyone quickly and completely and license free.

Let me first of all thank Council member Brewer who had the privilege of chairing this committee before me and for introducing this bill. It is a great bill and I am pleased to support it and I wanted to make sure that we gave it a hearing as quickly as possible.

Along with the availability of data through a single web portal, the bill also requires that public data sets be updated online and often enough to preserve their integrity and their usefulness. The data sets would be available without any registration, license or restrictions and would allow automated processing.

The bill also requires the city to implement a web application programming interface which would allow third party programs to request and receive information from the city's website in real time and pass that information to the users of their application. The bill mandates that DOITT establish an internet data set policy and all other agencies to review their data sets

under their control and to submit a compliance plan to be achieved by 2013.

It is contemplated of course that some data can be made available immediately and that data should be. Other data will take longer and that is contemplated by the bill giving the city more time to accomplish this.

I've always said that if you put the data out there, there is an entrepreneur ready to make an app out of it in about 30 seconds. We may not be able even to envision what they would create but let's unleash the power of creativity here. Let's give the tech world an opportunity to show government how it can work better. I believe that this bill can bring people closer to their government not only to make it more user friendly but also to be able to challenge it.

It's time to dismantle some of the barriers between New Yorkers and their information. While New York City already has made some very useful information public on nyc.gov such as the building information system, ACRIS, NYCSTAT, much of that data is available only in proprietary and/or visual formats and not all of

it is raw data.

For instance today we can only access COMPSTAT data in PDF form. Some of these databases specifically forbid automated machine processing. Let's tear down that wall and make this information available in its most raw and useable form.

The full benefit of publicly available data sets can be constrained only by the limits of our collective imagination. We look forward to hearing from the administration, Commissioner Post is here, and to all of those who are interested in testifying today. I want to note that in order to make for a more dynamic conversation between those present and those who are watching our live stream on the internet and by the way that is being live streamed at www.livestream.com/nycctechcomm. We are going to be taking questions and comments sent via email and twitter.

As indicated on the slips made available near the sergeant-at-arms, you can tweet this hearing using the hash tag INT029 or NYCOPENGOV or you can send an email to

NYCCTECHCOMM@gmail.com.

So now that we've dispensed all these ways to communicate in this hearing, I'd like to ask council member Brewer to make some opening remarks and again we thank her for introducing this bill.

COUNCIL MEMBER BREWER: Thank you very much Chair Garodnick. It's an honor to be here. I think we've come a long way in the last year, it was actually a year ago had a hearing on this earlier version of this bill and it's really great to be here with so many of you to do, just to have a discussion about a bill as the Chair indicated, that has a lot of transparency and opportunity. And I want to thank Commissioner Post also because when she first started she did publish a 30 day strategy paper which talked about the issues of transparency and the need I think for it to take place.

You know, the issue before us and Intro 29 is very exciting because I think it does a lot of, gives a lot of extraneous opportunities. You know there are perhaps 90 city agencies. Some of you know that our budget is the fourth largest

in the United States and probably larger than most nations in the world. So there are a lot of data sets, there are a lot of data. And the question is how to make it available so that it is something that can be used.

I do think that if we are able to pass this legislation in a format that is able to be worked on by both the administration and the council, there will be less FOIAing. As you know there are many many FOIAs. Some of you who many not have FOIAed may not FOIA but this is a very constant not just for reporters but also many New Yorkers.

As the Chair indicated there is a need to provide jobs in the city of New York and just having this date will enable developers and others to create jobs out of it. The MTA has done a little bit of that with the data that they have released and of course I have a very personal interest in the sense in that when the Chair and I and the Commissioner went to hear the Mayor announce the apps that he did and it was a very exciting moment for the city of New York, it was mostly and with all due respect, and these are

important things that tourists might like or people who are paying, you know, have interest in specific touristy kinds of things about New York. Parking places, to be very honest is not the number one issue on my list.

However, I am very concerned about low income New Yorkers and middle income New Yorkers and figuring out ways that some of the data that social service agencies have could enhance the lives of New Yorkers who need our services the most. So that would be another way that this data could be used.

And of course, just keeping this data updated. It is very difficult. I went on the Washington, Chicago, Boston, San Francisco websites just to see what they're doing with their data and to be honest with you, some of them have data that is outdated. Just by doing the work that we are trying to do, it has to be kept updated and I think New Yorkers would help the city to do that. So there are so many ways this data could be used and obviously in formats that make sense.

In final, I want to thank Lou

Klettner from the New York Community Fiber [phonetic] who has been webcasting this committee for many years and we appreciate it, Joely McPhee from the Internet Society and certainly the people from NYC IT division. This particular hearing is also available in the room next door. It's never been done before in the city council. And thanks Kanal Mahach [phonetic] from our office and Sam Wong. Kanal's going off to law school, so this is going to be his last hearing and Sam Wong is graduating and going off to a job so we're very upset they're leaving but we thank them for all their work and thank you Mr. Chair.

CHAIRPERSON GARODNICK: Thank you council member Brewer and Commissioner, the floor is yours, welcome and we thank you for being here today.

COMMISSIONER POST: Great, thank you Chair Garodnick, council member Brewer. I appreciate being here and being able to talk about this topic. My name is Carol Post and I'm the Commissioner of Department of Information Technology and Telecommunications or DOITT as we call it.

Joining me is James Parazzo, he's the assistant commissioner for web and new media operations for DOITT and just a pause on that introduction when I was last testifying here during budget hearings we spoke about some of the reorganizations that were occurring. James is actually the embodiment of one of those groups. He's been with the agency about eight years but since January we have repurposed him to build upon his skill set but to expand it into precisely this area to build upon the web and new media platforms and the opportunities that they represent.

I just want to take a moment to brief you on some of the advancements that the city has made with respect to open government and open data. For the past eight and a half years, the Bloomberg administration has been making New York City government more open and transparent than it's ever been. The city provides a wealth of information and data which every day is being made more accessible and interactive. A few examples include the city wide performance reporting tool, the my neighborhoods statistics feature, the New York City map, the stimulus

tracker, the NYC Scout, and 311 online. All this information and more has been made available at a portal called NYSTAT which you referred to. NYSTAT is one example of how the city is proactively aggregated disparate data and make it more accessible and user friendly eliminating the need for visitors to have to hunt and peck to find what they are looking for and in many cases there is raw data available there for dissemination by the public.

The amount of city information made available via nyc.gov today far exceeds anything previously available in the city's long history and meets or exceeds that of any other city in the world.

In the past year, the city entered the next phase of the open government movement that of open data. As the flood of social networking technology transforms government in a fundamental sense, the city must remain at the vanguard of that movement. The city will continue its efforts to develop innovative applications and to make the raw data behind these applications open and available. This is the public's

information and we want to continue making it available in as many ways as we can. Accordingly last summer we worked with the city's Economic Development Corporation on the NYC Big Apps program which council member Brewer referred to.

NYC Big Apps is a program whereby hundreds of data sets were made available to the public to create and develop new and unique applications and tools for public use. The program resulted in 80 new applications developed by the public for the public at essentially no cost to the city. These applications are now widely in use across the city and across the globe by New Yorkers, businesses and visitors.

The 200 data sets that were made available as part of NYC's Big Apps remain available at the NYC data mine which is accessible at NYC.GOV. The data mine represents data sets from nearly 30 city agencies and is searchable, sortable and free to the public. DOITT is now working with these agencies to add data sets to and improve the usability of the data mine for the second round of NYC Big Apps later this year.

As transformative as these

initiatives have been and we do expect them to continue. We fundamentally agree with the city council that we can do more. And institutionalizing the unprecedented gains made by the administration will insure for future generations of New Yorkers a city government that is transparent and accountable. That said, today's proposed legislation presents a number of fiscal, operational and technical considerations that may be problematic for the city. Chief among these are concerns about establishing reasonable limits on the use of data to preserve the integrity and capacity of a universal warehousing system.

While we agree with the council that ideally every data set that does not pose a security threat, compromise public safety or contain personably identifiable information would be publicly available that is neither fiscally nor operationally feasible in the short term. To really get open data right, we would propose an approach that would seek to classify data in terms of established criteria such as technical availability, timing and frequency of updates,

costs to implement and ultimately value to the public.

We would support a clear set of standards around what types of data agencies need to publish and when with certain minimum city wide guidelines. While it's currently drafted Intro 29 speaks to these ideas in part, we believe some of it remains somewhat loosely defined to be able to move forward without revision. The administration would seek the opportunity to better survey and qualify the criteria by which agencies are required to categorize and disseminate their data.

We are now meeting with city agencies to assess in more detail the challenges and impacts posed by the legislation. It was just such a collaborative approach that enabled the Mayor's Office of Operations to develop the city wide performance reporting tool. As we continue these discussions we would like to work closely with the council to find common ground on the comprehensive open data legislation that can have substantial and lasting impact on the way city government develops and shares information.

This approach will take some time

but we hope to establish as a result what we hope to establish as a result is an achievable and realistic path by which the city can make more public data centrally accessible online. And we hope that non-Mayoral city agencies like the city council, Comptroller's office, the Public Advocate and community board offices would also classify and contribute their data as part of these efforts.

The Bloomberg administration has consistently worked at creating a new city government paradigm regarding data believing that it should be open by default unless there is a compelling reason, usually privacy or security related, to keep it closed. We look forward to working with the council in crafting meaningful legislation to that end. This concludes prepared remarks and will now be please to address any questions. Thank you.

CHAIRPERSON GARODNICK: Thank you very much Commissioner for your testimony and again for your presence here today. I have a number of questions I know Council member Brewer does too. We thank you for your general flavor of

support for the bill. Though I think that we should now delve a little bit into some of the concerns that you have specifically and see if there are ways for us to address them here or beyond this hearing.

First of all, you noted that you believe that the bill is not fiscally or operationally feasible in the short term. As you know the bill divides up public data sets into a few different categories. One is the immediate category which is public data sets that are currently maintained by an agency. Legacy, which is public data sets that are, you know that are due to their size or complexity can't be made available until January 2012 or Priority which is essentially anything else. You noted that there are impediments to getting some of these addressed immediately. That also appears to be addressed in the bill.

Why is there a problem with just putting up data which is already available and out there within 30 days for the immediate category as set forth for the bill?

COMMISSIONER POST: Well, there's

clearly not a problem with doing that, we have been doing that. I think actually Big Apps represents the most advanced step that we've taken so where data was easily readily available in the appropriate formats that's been made available, that's where we have the 200 plus data sets out there. Where it begins to get challenging is for information that is not readily accessible or readily available and that's where there are both technical and in some cases fiscal constraints to be able to make those conversions. And those are not insurmountable. I want to be clear about that, it is simply a matter of understanding what the priorities are in order to set forth how we want to approach those as I said, making sure there is a sense of, what's the public value for it as we go through because we can't turn it all on at once. We are interested in being able to convert it over time.

CHAIRPERSON GARODNICK: So the point is that it's not readily available so timing is an issue which it can, of course is also contemplated by the bill so I don't think that necessarily we are that far off. Do you, or does

DOITT or the administration have any issue with creating a fixed timeline here by which certain categories of data must be made available?

COMMISSIONER POST: I think principally, no. I think we would like to have a self-imposed timeline as well. What's unclear is what that timeline should be to be realistic.

CHAIRPERSON GARODNICK: Right, at the end of the day we don't know whether it will be self-imposed or whether it will be putting it here in the legislation but right now January 2012 is set as the date for data which is not immediately available, the 30 day information. Do you think that that is too fast?

COMMISSIONER POST: It's hard for me to comment now because the issue is going agency by agency to delve into the data sets that each agency has as we said for the Big Apps program we did that with all of the city agencies and what it revealed was what was easily accessible and readily available so we pulled all of that out. The next step then is to dive a bit deeper to pull back the skin a little bit and see what's under there and that's the effort that we

would want to take some time for.

I used by comparison the idea of the city wide performance reporting tool and this was a program that I was involved with in my prior life at the Mayor's Office of Operations that was something similar to this in that it went, the ideas was to provide for a common input for agencies to use with a universal output and a universal publication to the public and in doing so required agencies to transform a great deal of data over time but it was an exercise of literally going agency by agency into their data systems to be able to pull that out, undergo the conversion and then be able to publish it. So it is a matter of timing I think.

CHAIRPERSON GARODNICK: So that review process with the city agencies, obviously pulling back all the layers and trying to sort all of that out as we had discussed that you are endeavoring to do in past hearings, how long a process is that exactly? Can you give us a sense of timing because if we are going to endeavor to set specific parameters for time we want to make sure of course that it is done right and we have

the proper protocols across agencies to be able to do this in a comprehensive and ongoing basis. How long is it going to take for you for a review city agencies to assess the detail and the challenges that would be posed?

COMMISSIONER POST: I don't have an exact time frame for you when we did the citywide performance reporting exercise the entire undertaking took a bit more than a year. My proposal would be that we take this a little bit, in a couple of phase in working with the council to move this legislation forward. I'd like to be able to have discussions with city agencies. A little bit of back of the envelope to be able to understand the challenges and then be able to set forth with what we hope would be more exhausting undertaking.

CHAIRPERSON GARODNICK: Okay, so there's not really an answer to that question at this point.

COMMISSIONER POST: I don't have one at this time. I think we need to be able to have those discussions with each agency or at least the key primary agencies where we know some

data is locked in and to be able to have a better sense of what the challenges are.

CHAIRPERSON GARODNICK: Okay, that's a conversation obviously that we'll have to have together and think about what is the appropriate timeline if not January 2012 for the stuff that is not immediately available.

You noted about the NYC data mine which included a repository of over 80 data sets back in—

COMMISSIONER POST: 200.

CHAIRPERSON GARODNICK: Now, it's 200. Okay, so was it 80 a year ago and now it's 200 or was it always 200?

COMMISSIONER POST: It was always 200 and the 80 referred to the number of applications that were built from it via the Big Apps competition.

CHAIRPERSON GARODNICK: So you put in 200 data sets into NYC data mine last June was it when it all came out, is that right? And there were 80 applications. Have you added any data sets to NYC data mine since last June? Have there been any new ones?

COMMISSIONER POST: Have we added? I don't think there's been full data sets added. There has been updates and enhancements to the existing data sets.

CHAIRPERSON GARODNICK: Can you say a little bit more about the enhancements to the existing data sets?

COMMISSIONER POST: Well, it's mostly refining to the extent that someone had visited the data mine, if there were issues with transmitting or pulling or accessing it. We were fine tuning it and certainly updating information that needed to have frequents updates in order to be refreshed and be of value.

CHAIRPERSON GARODNICK: And who does the updating of the data sets that are currently on NYC data mine.

COMMISSIONER POST: DOITT is the steward of the data mine and it reaches out to each individual agency to ensure that they're properly updated. The updates thereafter come form each individual agency.

CHAIRPERSON GARODNICK: And how often is that data updated, is that done in real

time or is that something that's done every week for the ones that need updating. How does that happen?

COMMISSIONER POST: It depends on the frequency. The trick about the data mine from the time it was published is that it was part of a competition. So in order to ensure the integrity of the competition we needed to basically populate the data mine and then sort of close it and let it remain static during the term of the competition so that no one who entered the competition earlier would be disadvantaged by someone coming in later and having more data. So there was sort of a period where it was closed but for refinements or if there were enhancements that needed to be made to facilitate the use of the data and thereafter that was launched in November, October, November so there have been updates to this but it depends on the nature of the data so for example, school data, Department of Education data's usually just done annually whereas other data might be done on a more frequent basis.

This is actually sort of the evolution of the data mine which is to move it out

of being simply a vehicle for competition and being more of a regular ongoing vehicle for public use.

CHAIRPERSON GARODNICK: Exactly. And I think that's the real intent of the bill which is NYC data mine shouldn't be so extraordinary, it should be what we are looking to do all the time. The bill distinguishes between raw data so that which is available in a machine readable format from cooked data which has been analyzed and summarized into a report. Do you have a sense of how many data reports are available by the city, made available by the city on NYC.GOV? Do we, is that something you all have a measure of?

COMMISSIONER POST: Total reports? Or raw versus—

CHAIRPERSON GARODNICK: Well, if you can break it down to that level that would be useful too.

COMMISSIONER POST: I don't have it on hand but I'm certain that we could obtain a master list of all the data that is currently available on NYC.GOV.

CHAIRPERSON GARODNICK: Okay, I think that would be useful for us and also the cost. It's our sense that by putting data out there and endeavoring to do this it obviously comes with potential cost but also potential cost savings through the reduction perhaps in FOIA requests and things and things like that. Have you considered what the administration believes to be the cost of implementing a system like- assuming your framework where you have a little more time to go forward with it and then and put all this information out there. Do you have a sense of cost?

COMMISSIONER POST: We haven't done a cost analysis, no.

CHAIRPERSON GARODNICK: Do you have a sense about the city's budget for processing and completing FOIA requests?

COMMISSIONER POST: I don't have a window into the entire city I know what my agency undergoes in FOIAs and there is hard dollar costs and then there s sort of the unaccounted costs in terms of time and use of personnel and resources. But no, I don't have an exact number for you.

CHAIRPERSON GARODNICK: It seems to me that they link together, the more you make available the less people have to request. So, but I wouldn't expect you to necessarily have all that information. I thought maybe maybe there was a chance you had it today. One more question from me and then I am going to go to council member Brewer. The bill also includes an application programming interface requirement. How would DOITT do it, essentially? How would you put that into place? Is that something you would have the expertise to do in house or would you need to outsource that sort of thing? Give us a sense of how you would make that happen.

COMMISSIONER POST: Sure, with respect to what's expected from the bill, that may take a further discussion, I can tell you what we have as far as an expectation on ourselves and I can defer to James who is actually responsible for and in the process of building this out but the idea of having the skill set within the agency to offer to the city and others is the idea of being able to build applications and to be able to have common platform to be able to use this data on

ways that sort of take it to the next level, out of being simply raw data and into useable formats. So, we'd like to be able to do that internally but what we learned and I think what really revealed itself from Big Apps is that most of the great value comes from people using the data outside of the city and being able to use tools as you referred to. Someone can take the data and in 30 seconds do something really fantastic with it. So we'd like to continue to forge relationships or partnerships that can foster those kinds of developments as well and kind of have both functions available to us.

CHAIRPERSON GARODNICK: Okay, well, thank you for that. I am going to turn over to council member Brewer for some questions. I may have a few more of my own at the end. Thank you. I'm sorry I didn't note that we were joined for a moment by council member James, council member Koppel is here and now we go to council member Brewer.

COUNCIL MEMBER BREWER: Thank you very much. I know you worked hard at Operations and you're working at DOITT to get agencies to get

more consolidated in terms of some of their operations. How will this feed into that? In other words, is there more sharing of data sets? Because obviously that would be something that we want, we want less silo, we want more collaboration. So how could the project and this bill help do that and is that something that's also your goal?

COMMISSIONER POST: Well I think modeling it off to the extent that it has a comparison to the city wide performance reporting exercise which was about taking performance metrics and getting them into a more universal and consolidated state, both from the agency's perspective in terms of how they delivered that data and DOITT's perspective in terms of how it was published to the public so rather than having to agency to agency and having to hunt and peck for information it's now in a single repository and it's actually been made easier for agencies to be able to populate it on a regular recurring basis. So I think that's the model we'd like to try to follow and as I said when it's all done it looks great and it's a useful tool but it was a

relatively heavy lift to get there because the agencies have evolved organically and grown. There's a wealth of information out there but it hasn't been done on universal standards.

COUNCIL MEMBER BREWER: So I know COPIC [phonetic] years ago in 1989 said that the city should publish a data listing basically of all the databases. So does DOITT now have do you think all the listing of the databases that exist and was that one of your biggest challenges?

COMMISSIONER POST: That is definitely one of the challenges. So when I referred to being able to quantify the data that is available on NYC.GOV that's sort of square one. What's going on at city agencies is both data that is available, readily available as I said and accessible but there is a wealth of information that is used operationally that hasn't really made it to the public data base not for any other reason than it was grown organically out of the spreadsheet or maybe into an access database or something that sort of matures on its own and there is no single catalogue of all that information.

COUNCIL MEMBER BREWER: And how do you get all the data bases once you find them and the legacy systems coming in line and so on. How do you deal with the format issues? Is that also something that would be part of this effort? Because obviously for the public it's an issue for collaboration, coordination, etc.

COMMISSIONER POST: It definitely is and that's again referring to the challenges that will face agency by agency and frankly data base by data base to be able to develop a universal standard for both the agency to provide the data and for DOITT to publish it. That's one of the main challenges.

COUNCIL MEMBER BREWER: So how does one get the agencies to conform to giving. I know for instance I have friends who work for HRA. HRA has a lot of databases that you don't know anything about so, because they don't know anything about them in some cases, so how does one go about getting this to be some, is it Operations that does this, is it you that does that, who has the authority to be able to say to some of the silos, we need to—this needs to happen? How does

one go about that?

COMMISSIONER POST: Well, I think that's what we need to determine from the administration side is how to approach this programmatically and I'm sure that there are a number of ways to tackle that and we would like to be able to come back and have a solid plan for you. The idea of setting criteria for in legislation of this sort helps that in terms of being able to establish sort of a set of priorities so that we are able to approach it in an iterative way rather than saying it's all by x date which does provide, the bill does propose sets forth kind of an iterative approach but I think we wanted to add a little more structure around that so it's clear what we want by when and how to get it.

COUNCIL MEMBER BREWER: Okay, and obviously part of that would be what kind of language is it, XML or dealing with the spreadsheet challenge and so on so I assume that all of these are the challenges that you are talking about, the language, whether it's RSS and so on.

COMMISSIONER POST: Yes.

COUNCIL MEMBER BREWER: The only other question I have is the 311 data. How does that fit into all of this because it's a different kind of data set. It's obviously something that's up on the web in terms of the by borough and so on. Because there is I guess a national effort. Again New York is so huge we are not part of with all due respect to Washington and San Francisco, I understand all that. But there is a big effort to do this open 311 system back to this issue of trying to be comparable with other cities so I'm just wondering because 311 brings in sort of-- having this bill I think would help the 311 operators answer some of their calls because knowing that this has been you know a problem for a long time etc., would be something that I think would help solve some of these problems. Certainly could help the community boards I can promise you having at with them and trying to deal with their district service cabinet so how does the 311 data fit into all of this discussion if at all.

COMMISSIONER POST: Well, it does.

I'll let James answer some of the specifics that you asked he has dealt extensively with 311 and the use of its data and the transformation of its data into different forms and uses. But principally 311 is, it's a microcosm I think of what we want to achieve which is using data that is available but in ways that are kind of advancing the cause so 311 data where we have been able transform our scout data into having it be mappable is one advancement and we ant to be able to take those things further. Do you want to speak about some of the open data discussion that we have discussed with respect to 311?

JAMES PARAZZO: Sure, the 311 data as with many other data sets is included in some form in the Big Apps and would be included in the data that we give out. It's also as you mentioned given out specifically with 311 given out in accordance with the law.

COUNCIL MEMBER BREWER: By law.

MR. PARAZZO: With your law. In the format specified by that law. We are always looking for ways to give that data out within its own context and to have that contribute to other

contexts. The movement that you referred to, open 311, we have participated in conversations about that with San Francisco and D.C. and other places and we are supportive of the concept to the extent that there is a national API developed is workable for us and data will be available through that.

COUNCIL MEMBER BREWER: When I go to data mine on your site, which is a wonderful site, and the question is how does the public right now, in other words right now you have over 200 data sets and the fact of the matter is people can take those and make something of it if they want. Do you keep track, I don't know if this is possible but, do you keep track of other requests that people might come in with that could indicate that this is something that the public is interested in? People, I don't know health statistics, environmental or whatever, something that—

COMMISSIONER POST: Well we have a, as part of the data mine, we invite comment as part of the what's not there so if there is there is a particular interest in data you can submit a form right there with the data mine that says you

know what were you looking for, did we not have it, that's some of what help inform us of what we should be adding.

COUNCIL MEMBER BREWER: What agencies are perhaps right now as part of the 200 data sets and I should probably know this are the ones that are the easiest to come forward in terms of their information? Are there some agencies that have you completely given you all of their data as far as you know?

COMMISSIONER POST: I don't know if every single agency has completely given data. As I said there's literally, you know, there's some functional areas that have spreadsheets that are used and logs and those are done just for operational purposes. But the agencies have been in terms of getting through the Big Apps Stage I there was a great deal of cooperation and an openness to get there again in terms of what was readily accessible and available. There is a commitment to be able to turn that over. The next trick is to get to the more challenging information.

COUNCIL MEMBER BREWER: Well, thank

you very much. I'm very excited at the notion of working with you on this issue and as you can see from this turnout today and from the other room there were certainly quite a few tweets already, lots of emails I'm sure particularly in the city of New York there's a huge interest in this topic. And I think that in the end it would mean not only savings for the city if we could ever get those FOIA numbers but in addition it would mean jobs and opportunities that didn't exist. Of course my other prime interests are the 59 community boards and having sat through hundreds of I'm afraid meetings of community boards, clearly with the different agencies it would make a world of difference to local planning. So thank you very much. That's it for questions.

CHAIRPERSON GARODNICK: Thank you council member Brewer. I want to note that when we're finished hearing from the Commissioner we're going to hear form members of the public and we are going to have a three minute time limit because we have to relinquish the hearing room for another committee at some point, not immediately but at some point. So we are just preparing for

that.

I wanted to note we have a couple of online questions that have come in so I am going to pose those directly to you Commissioner. They are along the same lines so I am just going to give them to you together. The question is how can DOITT engage the community on open data? There's a concern that DOITT is not reaching out enough and whether there are ways you can encourage people to participate. I know you just gave an example of how you asked people at the end of NYC data mine. Are there other ways that DOITT can be taking more affirmative steps to engage the public on open data?

COMMISSIONER POST: Well, I think the answer is always yes, we can always be doing more. One of the other areas that I spoke of again, in terms of reorganizing the agency, gave a specific purpose to engaging our business community and our resident community in terms of how we can be doing better and in terms of outreach I spoke last month at the Personal Democracy forum, previously I've attended Tech Meetup sessions and the idea is that we should be

actively participating in these types of communities, sharing what our agenda is and how we want to move forward with that as well as hearing feedback from the constituents and those who can help us achieve that goal so we are very interested if there are forums, sessions, seminars that we would be happy to participate in and I'd like to hear about those.

CHAIRPERSON GARDONICK: Okay, so we've heard from the Commissioner, invite her everywhere because we do want to make sure that you get the feedback from the community out there because they will be the best guide in telling you where there are gaps or where there is an opportunity for more openness.

So let me go to a couple of final questions from me. In your testimony you noted that there were fiscal, operational, technical considerations and you say chief among these are concerns about establishing reasonable limits on the use of data. Can you give us a sense of what you view to be reasonable limits on the use of data and what you propose because there is a concern about whether that could have a chilling

effect on the use of the data and reasonable is in the eye of the beholder if you will. Can you give us a sense of what you believe to be reasonable limits?

COMMISSIONER POST: Well, reasonable is, I think it’s the appropriate term and where we define it is with these further discussion I would welcome other agencies and in terms of establishing some of these parameters within ultimately the legislation. The idea though is I think twofold. One is to first and foremost, protect the integrity of the data that is not and should not be made available to the public. At the end of the day the city is the steward of very crucial and critical information an data and I think we want to be careful never to err on the side of overexposing and disclosing information that would both be damaging to those who would be the subject of it and I think damage the credibility of what we’re trying to do so the reasonableness of ensuring that protection of the integrity of that data.

The other is a little more from the technical perspective which is ensuring that the

methods and means by which we make the data available have appropriate limits is not really the right term but appropriate mechanisms to ensure that no single entity can sort of tap in and sort of use the capacity for purposes that limit any other ability to tap in. It's a, to the extent that it's a broadband issue, it's a capacity issue. Being able to ensure that we package the data in a way that everyone can access it as they need to and no one is sort of kept off the wayside while others are tapping in. I think it's both the integrity and the capacity issues.

CHAIRPERSON GARODNICK: So if I'm understanding you correctly, it's not really reasonable limits on the use of the data some of it is really the question of what data in the first instance is put out there at least for problem one as you described, to protect privacy interests the city might have. Is that right?

COMMISSIONER POST: That's right.

CHAIRPERSON GARODNICK: But once it's out there are you talking about limits on the use in some form or another and once it's acquired and once you get over technical hurdles that you

cited and pronged to is there any reasonable limit on the use that you anticipate?

COMMISSIONER POST: Well, again, I think I want to be able to preserve an opportunity for agencies to opine about their respective data, any particular data set that may have particular issue or matter where a reasonable limit makes sense. I don't know that there is one but I think it's of importance that we're mindful of that and I think more so is the fact that once data is put out into the public realm that there is an understanding after that. We can't necessarily vouch for the integrity of that once it's been manipulated in ways we want to be able to ensure that there is an integrity to what was published and thereafter it's on its own after that. And there's concerns about that.

CHAIRPERSON GARODNICK: Of course once you put it out there and allow individuals to double check and make sure that all eyes are on it and it does have some element of protection as well but it sounds like sitting here today you don't have anything in particular in mind about what might constitute a reasonable limit on use

but you didn't want to foreclose the possibility of an agency identifying something to you that you have not yet considered. Is that correct?

COMMISSIONER POST: Yes, that's correct.

CHAIRPERSON GARODNICK: Okay and then you also noted that among the—in the approach that you recommended about how to classify data, technical availability, timing and frequency of updates, costs to implement and ultimately you say value to the public. Question of course is another one along the lines of reasonableness. Shouldn't we be concerned when it's the agency that's making the determination on value to the public? Might that not ultimately restrict data because of the bureaucratic absence of imagination perhaps? How do we address that?

COMMISSIONER POST: Well, I think in the remarks that I made. It's not intended to imply that we would be reserving the ability to make those judgment calls. I think each of those factors should be brought into consideration and the idea is again if there are tens of thousands of data sets that have potential for publication

and we can't do them all at once, which ones do we want to do during this iterative process and if the ultimate determination is which is going to bring the most value I think we want to be able to have that as part of the dialogue. I mean simply stated we don't want to I think we don't want to be expending efforts, resources, time and money to put data sets available that may be of little value. The obvious question is, who's to say? I think it's just a matter of having a broad set of criteria that can all come into play and into consideration.

CHAIRPERSON GARODNICK: Okay, we take your point on that for sure. With that, we see no other questions coming to us via the email address we set forth so we will thank you for your testimony and your presence and your support and we look forward to having further conversations with you and the administration about how to put together all of the right nuts and bolts here to be able to get this bill passed. So we thank you for your testimony.

Now we're going to turn to the public session and I'm going to invite up our

first panel which will be Andrew Hoppin, Don Morrow, and Aaron Brown. And I just want to remind you gentlemen that we're going to have to put you on a three minute limitation. My apologies in advance but again we have to relinquish the room at some point. It's nice to see you all and welcome. So why don't we just start right over here. Go right ahead.

DON MORROW: Thank you. My name is Don Morrow. I'm the founder and Chairman of the New York Technology Council, a trade association whose mission is to help make New York a world recognized center of excellence for technology. Founded in 2009, NYTECH boasts over 250 member companies and is proud to include among its founding sponsors Google, Verizon, Information Builders, Citibank and others. I would like to thank the council for allowing me to speak today on this important topic.

As I'm sure everyone on this council is already aware, open data initiatives are taking hold across the country. From the California open government directive S2009 to the federal government's data.gov's data portal,

governments are beginning to realize the societal benefits of open data standards. Indeed New York City as Commissioner Post has indicated, New York City has already begun to dip its toe into the oceans of data rightfully belonging in the public space. Last year's Big Apps competition invited entrepreneurs from around the city to develop software applications based on publicly available data sets from New York City agencies.

The competition was a huge success garnering dozens of submissions, 80 apparently, and paving the way for the recently announced New York City entrepreurial fund and a $300,000 investment in one Big App company. There is no debating the positive impact of this program.

To paraphrase perhaps the biggest business leader of our time, the pointy haired boss from Dilbert, this bill is low hanging fruit and a win-win. It will empower synergies, shift paradigms, develop core competencies and encourage out of the box thinking. In short, this bill is a good idea.

This should not be a contentious bill. No one will lose a vote, no one will lose

an endorsement, no one will lose a dollar of financing by supporting this bill. What you will gain is recognition from the community that your affirmative role will open doors for enterprising companies to develop new and exciting ways to experience New York City.

Without this law, left to their own devices, some city agencies, such as those already participating in Big Apps would no doubt take the initiative and release valuable data sets for public consumptions. Others however will be less than cooperative. A city mandate to publish these data sets would serve to overcome the petty bureaucracies and misguided excuses that frequently mire such programs.

Government regulation frequently lags technological changes. By enacting this legislation, you have the opportunity to break that trend and ensure that great New York City takes the lead in recognizing the power of technology to build trusting government, foster innovation and improve the society with which we live. I encourage the full council to pass this important piece of legislation soon. Thank you.

AARON BROWN: Good morning Chair Garodnick, council member Brewer and the rest of the committee. Thank you on behalf of Google for the opportunity to be here and to testify on this important issue. My name is Aaron Brown. I'm the Senior Product Manager at Google based here I New York City. As you know, Google has a major presence in the city with over 2000 employees, our second largest world wide office here. We are very excited to be part of the city community and to participate in important discussions like this.

At Google we are very familiar with the power of data. We use it extensively to build our own products and perhaps more importantly for this discussion we build and make available platforms and tools for others to make data more available and more transparent whether it's through Google Maps or Google applications or purpose built data sharing platforms like Fusion tools and public data explorer.

We do all this because we recognize the importance of openness, transparency and broad based innovation around data. We want to help further promote those goals. Our perspective has

always been that citizens should be just one search away from all online public information and we've worked with many state, local governments over the past years to help encourage those principles.

For example, we partnered with state governments to help them expose more data behind web forms and firewalls to search so that constituents can find it. We recently partnered with the state of California around their apps for California's initiative where they published a number of electronic data sets on our platforms that have made it easier for constituents to build innovative apps quickly on top of it. We worked with the Census Bureau, the CDC, and others to create interactive visualizations that bring population data to life and shed new light on inequities and opportunities in that data.

We've also seen many cases where public access to data has helped everyone from journalists to businesses to academics and grassroots organizations to take public data and quickly create compelling presentations. So all of these projects were made possible by the open

electronic availability of public data and by the innovation of entrepreneurs that comes when they have access to it.

There’s an important pattern here. One that we think is important to encourage where open data is driving innovation and as such we are very supportive of the goals of Intro 29 and look forward to any opportunity we might have to work with the city should it pass to collaborate on making New York City a beacon community in its support of availability of open data.

I would like to comment before wrapping up on a few of the provisions in Intro 29. Based on our experiences, technologists are particularly critical to its success. The requirement that data be available in form to support automated processing, the requirement for API access to the data and the requirement that public data sets be accessible to search. We recognize that the cost to implement some of these provisions can be non-trivial. We believe it’s important to set goals to achieve them even if they’re phased over time. And we’d encourage the department to leverage the private sector where

possible to accelerate this process perhaps working with Google and others in the industry that already have scalable open platforms with APIs to make data available.

So to conclude I'll just reiterate my support for Intro 29 and I hope the perspective I've provided today has been useful. Thank you for allowing me to testify.

CHAIRPERSON GARODNICK: Thank you. Right on the buzzer. Thank you. Go ahead, Andrew.

ANDREW HOPPIN: Hi. Thank you for having me here today. So in the New York State Senate the number one—

CHAIRPERSON GARODNICK: Why don't you actually identify yourself first.

MR. HOPPIN: I'm Andrew Hoppin, Chief Information Officer for New York State Senate. The number one project that has delivered value out of our working technology over the last two years in the New York State Senate in my opinion has been our open data effort. So with open data at NYSenate.gov we publish all of our administrative data: how much I get paid, how much

money we spend on everything in a searchable sortable downloadable mashable form.

Similarly our work product is legislature. All of legislation is now available for free as feeds also with the what we think is a fairly intuitive interface and with an application programming interface so that third parties can build applications on top of that data. And we were able to accomplish all of that in less than year and actually we're able to net save money on all of our technology spending and bargaining just because of the way that we did it in concert with other organizations and enterprise IT.

So, all of this to say that we care about this deeply and we believe in it. And I think the relevance of that experience to New York City, which is admittedly a much larger and complicated set of data to consider publishing in this matter, is that you can start small and you can start modestly and you can do this work efficiently if look to the best practices that are being created by our peers and other governments elsewhere. This work is happening all over the country right now and New York would not be alone

if it had the opportunity to lead we would also able to leverage off of a lot of work going on within government and within the private sector. And so I think that the complexity and the costs should not be impediments to going down this road because I think you'd get a lot of help in doing it.

Also, another plug for government, government's as a customer, governments outside of New York City, governments need to collaborate effectively across all levels of political geography, local, state and federal. And I think one of the greatest benefits of opening up New York City's will be an enhancement of your ability to collaborate with your peers at other levels of government and certainly we count ourselves on the state level as being very interested stakeholders and potential collaborators of that data product.

And then finally in terms of process, I would just encourage you to minimize the temptation to adjudicate what's going to be of value and what's not going to be of value because really the point of pushing this data out in a away that other people can leverage it is that

they will come up with the innovation that otherwise civil servants would have to come up with and we'll never have all of the smartest brains within government, we'll never be able to hire enough people to do all of the work that we might like to do so as much as possible really focus on getting the data out in a mashable form. And so I think that having an API for city data would be really exciting and I think it would be great to collaborate with other government entities and the private sector in developing that. Thank you very much.

CHAIRPERSON GARODNICK: Thank you. Let me just follow up on that point, Mr. Hoppin, on the subject of minimizing the temptation to make the decisions ex-ante for the public about what is or should be made available. You heard the Commissioner talk about reasonable limits and also value to the public. Can you envision any reasonable limits that could or should be placed on these data sets? And she testified that there was nothing in particular that she had in mind but that agencies might come up with reasonable limits.

From where you sit and I'll pose the same question to the other gentlemen, is there such a thing as a reasonable limit here?

MR. HOPPIN: Two things. One, I think certainly when an individual citizen information comes into play there are reasonable privacy and security questions that need to be answered but the lion's share of the data that I've seen governments publish effectively that's delivered value does not pertain to individual citizen information or records. In terms of, the way that I think you can address the obvious need to start with the low hanging fruit and to start somewhere and not let the perfect be the enemy of the good is to make decisions about what to focus on first as part of an open collaborative process. You know, use the same process and you hope the data will ultimately be used within meaning if publish the catalog of the data that is available take public input on what is of value will start the process of people thinking about and weighing in on what might be valuable in that manner rather than having that process go on behind closed doors.

And then the other thing is I think to build your systems that are going to publish this data, you know build your API in a way that anticipates that everything is ultimately going to be published. And so you build that infrastructure that supports you going to that goal over time.

MR. BROWN: I'll just reiterate the point that I am not letting perfect be the enemy of the good, we have seen in our experience that getting data out even if it's not in the most perfect form with all the API's available starts the process of innovation and provides a lot of insight into how to drive it forward in the long term. So again, design for the future but don't let that hold back from the initial exposure of data to the community.

MR. MORROW: Clearly once the data has been vetted for privacy and security concerns and it's put out in the wild people have the tendency to use the data to tell the story they want to tell and we need to differentiate between the creative use of data to perhaps tell one side of a story as that could still be a proper use of data so differentiating between a truly improper

use of data and just creatively massaging data perhaps to tell a story that might not be a very popular story we need to be careful to differentiate between those two.

CHAIRPERSON GARODNICK: Council member Brewer.

COUNCIL MEMBER BREWER: Thank you. I'm impressed that the Commissioner is still here and Hoppin, the New York State Senate gets a bad rep sometimes, I think they should just mention you and they'd all be happy.

The issue of Google and the MTA, how could it provide an example of what could go on in terms of what we talked about today because people are very pleased? Well, the MTA data.

MR. BROWN: Well, that's to me you said. I've actually not been involved with the MTA's project and Google's involvement there so I can't speak to that directly. What we have seen with other data sets is that putting them in platforms that are web accessible and friendly to consumers and the small organizations like Google's has been very energizing to the community and [crosstalk].

COUNCIL MEMBER BREWER: And in places like large states like California where you mentioned in your testimony, how again does that serve as an example as to what is or isn't impossible in New York just generally.

MR. BROWN: I think the great example of California is that California started by putting data out 400 or so data sets and did not have API's available and so there were challenges in getting the pace of innovation going. By working with us they were able to basically simply transfer that into our platforms which already provided open API's and as such they didn't have to invest in the creation and building of those API's. We just essentially gave it to them for use and so citizens could then move much more quickly and point and click visualizations without programming without a lot of the work that would be happening. They'd have to build that out themselves so it just accelerated its innovation.

COUNCIL MEMBER BREWER: Okay, Andrew, is the state doing anything close to what you're doing? You're working on the Senate side legislation and so on but is the State doing

anything to the best of your knowledge.

MR. HOPPIN: Not in a comprehensive across the board way in the executive branch. You know I sit on the statewide CIO Leadership council and we were having a lot of conversations about this and I think one of the exciting things is if New York City moves aggressively in this arena you're going to be moving in parallel with a lot of peers at other levels of government and again that's where the real upside of massive efficiency and potentially even massive cost savings can come from freeing this data because you're talking about unwinding silos here where you may have had a city department not being able to readily share data with another city department. You unwind that and inherently if you publish this data openly you're also breaking down the wall between government and on-government. But I really think breaking down the wall between local, state and federal has tremendous potential upside and it's hard to measure that until it's done, right? But it's a really timely time to be looking at this because of what I think will also happen eventually also at the state level.

COUNCIL MEMBER BREWER: And congratulations on all the work that you've done to put together everyone but the question I have is, do you get in some of your discussions around the table and around the forums, people who state if I only had this data I could do this and this in terms of apps or is it, obviously 200 of these are available but are they other discussions about other apps do you think?

MR. BROWN: I get that in a general sense but I don't get a lot of that type of discussion around city data in particular I can tell you that there was overwhelming good experience with the Big Apps competition. I heard nothing but positive things, nothing but positive outcome and I think that was the genesis of this whole movement.

COUNCIL MEMBER BREWER: Thank you.

CHAIRPERSON GARODNICK: Thank you council member Brewer. One last question from me for Mr. Brown. You noted that there is potential delay associated with mandating APIs from Day 1. Obviously having them in place is a great appeal and a great usefulness. How concerned do you

think we should be about that and what is your recommendation as to whether or not we should have the firmest most restrictive put it in right at the beginning. What would you say to that?

MR. BROWN: I recommend a phased approach much like is described in the Intro 29 for data availability itself so such that there's a grace period before API requirements are in place but that those requirements are defined and make clear goals by a certain time period, one year, two years, three years, whatever makes most sense that those API's are required and available. My point is with the testimony is basically not to halt the release of data while a long drawn out process as this happens to determine a standardized API's.

CHAIRPERSON GARODNICK: And you think the time necessary to put them in place is measured in months or year or two years as opposed to anything longer than that?

MR. BROWN: It depends on when API's are chosen. For example, the Google fusion tables, our publishing product has open APIs and if those were adopted as the state of California

started to do and they could be available I a matter of hours frankly.

CHAIRPERSON GARODNICK: Well, we thank you for that and we thank all of you for your testimony today and we're going to move on to our next panel so we appreciate it. The next panel includes Ben Berkowitz, Phillip Ashlock, and Liz Hodes. So come on up and join us. And I'm sorry if that's Hodes as opposed to Hodes, I have some relatives who are Hodes. Alright, are we missing? Alright, let's get right started. Ms. Hodes you can start us off and we'll see who we're missing.

LIZ HODES: My name is Liz Hodes and I work for Digital Democracy, a non-profit based in New York looking to empower marginalized community with digital technologies. The decisions made here have an impact on our schools here and overseas and I thank you for taking the time to listen.

Tomorrow's youth need to have the skills of 21st century citizens. Imagine a city where students can learn about their environment by getting data in real time in their classrooms.

Students would have the knowledge of what's changing right outside their window, in their parks and on their streets. This increasing information would spark interest locally and teach them if they can make an impact on a local level they will ultimately contribute their efforts around the world to enact environmental change on a global scale. Young people would be engaged in meaningful ways with the world around them.

If New York City is to continue to be a competitive global center we need to be able to react and respond to our changing times to this post industrial revolution, a transition due to the development of a global community online. In schools, students are learning the skills they need in industrial societies that favors education offline. These methods of educating are becoming increasingly outdated as data becomes available by the minute and as today's youths spend more and more time being connected.

Youth are consuming all available information while crucial government data is still locked away. Why not harness this passion online as time spent on facebook, producing content for

youtube and on mobiles connecting with friends for a positive engagement in the classroom, meaningful connections with the community and ultimately for effective global change. Otherwise we run the risk of their minds becoming obese with unhealthy information as their bodies would from unhealthy foods.

I ask you, what would you change and improve about your community? I'm sure that many of you have an answer to this question. Why not ask our kids and in doing so give them the information they need to not only answer this question but make the goals for their community a reality. Why not give them the information about where our city's best water supply is, public safety information and places to volunteer. Can increasing the prevalence of park benches decrease local crime? Can making data available about the nutritional value of our food decrease rates of obesity. By having this crucial information, youth can find answers to these questions and can be ambassadors for change in their communities. They can see what will affect real change and know whether their community can have the resources to

make these changes possible. On Saturday June 5th we worked with 120 students with the Future Now program with the Department of Education. Using limited data we helped these students to see what changes they would like to see take place New York City by 2020. We have seen open data work wonders in other countries. In Thailand we saw data supported Health Care systems that are cheaper healthier and more efficient than ours. In refugee camps there are internet connections enabling young adults to come attend classes at NYU.

Here in New York, Digital Democracy has been creating a free and open source educational platform, Roebling, which facilitates digital literacy and technical skills.

We target students using mobile phones, computers and other devices to share photos, videos, maps, blog post and homework assignments. This positively impacts student's academic performance and prevents drop outs by engaging them through participatory education and enable teachers to track academic performance through quantitative analytics and qualitative

data. We support this legislation. It enables us to begin implementing our cutting edge educational programming which fosters positive engagement between students, their government and community. Thank you.

CHAIRPERSON GARODNICK: What perfect timing. Nicely done. Go right ahead.

BEN BERKOWITZ: Thanks. So I'm Ben Berkowitz and I'm representing my company, which is See, Click, Fix. I'm the co-founder and CEO and I'm also representing an already open and public data set in New York if you go to seeclickfix.com/newyork so what I will speak about is specifically I guess 311 data in New York and what our tool does, we created our tool as a way to provide open and collaborative communication with City Hall and about three years ago when I think only Gail Brewer was talking about this as far as I know and specifically you can go online you can post something like a pothole publicly. We will send alerts to City Hall, City Hall, the council, your neighbors anyone who is interested in being part of the solution. And your neighbors can come on and support your issue thereby having

an open and collaborative way to discuss the problems in our municipal fabric and talk collectively about resolving them.

So currently the 311 system in New York is really a one to one system where data flows from a single citizen to a single call taker and then and then I guess I don't have a full understanding of this but at the end of the month there are data sets that are put out but not in real time and not in a real way that can encourage participation.

So what we are proposing, we support this bill obviously because the open data would be something we could display on our site but I should add a caveat that it really is only half the battle when it comes to 311. You're talking about the read side of the API and the write side is equally important and if we can't contribute our data set to 311 that we already have, I think there is a shortcoming that exists but this is definitely the first step and we want to see it happen.

So some quick benefits I think of what you would get from open 311 data is one,

empowerment. Providing that quick real time feedback would to citizens that they're issues were being responded to and seeing other citizens' issues being responded to will provide a necessary first level of engagement and a positive feedback for citizens to engage further in improving New York City and being an active citizen.

Efficiency, obviously crowd sourcing citizens and developers and the media to get the data to citizens is something that's really going to save costs as we get more 311 calls going on the web and off the phone lines.

And entrepreneurialism, obviously there are going to be other businesses like ours that can develop in New York City and can create jobs and benefit from this open data.

CHAIRPERSON GARODNICK: Thank you very much. That is certainly a loud signal. You can't miss it. Alright, thank you. Go ahead.

PHILLIP ASHLOK: Good morning, Chairperson Garodnick and thank you for having me here testify before you today. My name is Phillip Ashlok and I am the open government program manager for Open Plans, a non-profit civic

technological organization here in New York City. Much of the work that I do at Open Plans directly relates to this bill in that I work with cities to establish open standards and best practices for municipal technology. One example of that is the open 311 standard that was discussed earlier that involves See, click, fix in several cities in this country and others.

So Intro 29 is a very important piece of legislation which I believe can have a profoundly positive effect on the city. However, rather than starting off by going into depth about what is good about Intro 29, I'd like to provide some context in which the place is built relative to precedence in New York City government current state of open data and open government practices internationally. Section 1062 of the New York City charter requires the New York City Commissioner on Public Information and Communication to publish a public data directory describing the computerized data sets maintained by city agencies.

This is the first public data directory published pursuant to their requirement.

Publication of this first edition represents an important step towards fulfilling the goal to improving public access to information about the wide variety of computerized data maintained by the city. Information maintained by the city agencies is increasingly being stored into computers. Until now however there has never been a single source of information available to researchers, community groups, businesses, and other members of the public regarding the types of electronics data kept by city agencies much of which is required by law to be accessible to the public.

The New York City Commission on Public Information and Communication is new city agency established by 1989 amendments to the New York City charter. The Commission is chaired by the President of Council and includes public members as well as representatives of the Mayor, the city council and a number of city agencies. In addition to publication of the public data directory, the Commission's responsibility includes education and outreach to assist the public to have access to city information and

developing new strategies for use of new communication and technologies and improve access to and distribution of city data.

This public data directory represents the joint efforts of the members of the Commission and particularly the staff of the Mayor’s office of Operations and the Law Department. The Commission also wishes to acknowledge the valuable assistance provided to the agencies themselves in preparing the directory. For each agency, the directory provides a brief description of the agency’s mission, the names and the phone numbers of the public liaison available to assist members of the public and brief descriptions of the contents of the databases.

User notes contain important information on methods of access, legal restrictions, and access to certain records and other information, so I can keep reading that but that was from the introductions to New York City ‘s first public data directory published in April of 1993. And let me reference another document dated April 30, 1993 this is from SERN [phonetic]

signaling the release of another data directory. This is SERN's [phonetic] main declaration of the world wide web and is essentially the web's birth certificate.

I draw these parallels for historical context both were released in April of 1993. New York City has the earliest and most comprehensive open data policy of any city or government that I'm aware of and it's written right into the city charter. But since this policy predated the birth and current ubiquity of the web, it has largely fallen into obscurity and has been treated as nearly irrelevant and the rest of this is on datanyc.org.

CHAIRPERSON GARODNICK: Thank you. You know we do not have copies of your testimony. Do you have copies with you?

MR. ASHLOK: Yeah.

CHAIRPERSON GARODNICK: If you could provide those the Sergeant we'll pass those around and we'll have questions for you even before you go anywhere but let me start with one for you Mr. Ashlok. On the subject of timing and the need for agency to do and the patience that the

Commissioner encouraged us to have today. Do you have any comment on that as to whether the time lines set forth in the legislation are reasonable or whether we should be waiting for a more deliberative agency consultation process. And of course this is a question that's open to any of you.

MR. ASHLOK: Let me first start of with the reason I began my testimony with reciting the 1993 public data directory because that's 17 years old and it hasn't been maintained for 17 years. It was only published once as far as I'm aware. So I think 17 years is probably a pretty good amount of time to get at least the directory of data sets that the public should be able to access.

CHAIRPERSON GARODNICK: Okay. So that's for a directory and I guess really what I want to understand from you is putting the data right out there. The bill says some categories must go up within thirty days, other categories have to go up by January 2012, others by 2013. 17 years obviously is well beyond the contemplation of this bill and yes we have waited too long and

we are going to move much more quickly here so my question for you is [crosstalk] specifically on the actual data sets.

MR. ASHLOK: I don't see any reason why I think it's an additional three years for all Legacy data. I don't see any reason why that should be, I think that's completely feasible in fact if I think there's cost savings in doing so it involves getting rid of Legacy systems which often cause governments undue or extreme expenses where they wouldn't otherwise.

CHAIRPERSON GARODNICK: Anybody else want to address that one? Go ahead.

MR. BERKOWITZ: So I spoke a little bit with DOITT about the 311 systems specifically and I think there are some hurdles in New York that other cities don't have, one being that there are more Legacy systems that 311 is or should be interacting with than in most cities as opposed to having one work order system there may be many. And that does provide some complication but that being said we've brought around CRM's online integrated spitting out open data and collecting open data in under a month. I think you should

set aggressive deadlines and it's time for this to happen.

The other pieces that as Phillip just mentioned in regards to some of the data sets like 311, there's already a standard that's being worked on by other municipalities. We're connected to San Francisco and D.C. and so by being just a few steps behind you can really use some of the resources. We're obviously here to help and I think, set aggressive timelines and we'll help you get moving.

CHAIRPERSON GARODNICK: Mr. Berkowitz, let's go back to one of your comments in your testimony. I just want to make sure you have a chance to put a little more meat on the bone here. You talked about read side versus write side and what's being contemplated is only half the value. Can you just say a little more about what that means?

MR. BERKOWITZ: Sure, well I think if look at democracy it's not just about listening it's also speaking up and participating. And specifically what read looks like is a list of the potholes. What write looks like is you

communicating that there are potholes and yeah you concurrently you can do that through the phone by dialing 311, by going on the web, or by going to the iphone app. But if the city allows for developers to build alternatives that can create and contribute to that data set you will find that you will reach a lot more people.

Concrete example would be Washington D.C. has done this where you can now report potholes through other applications including ours, See, click, fix is actually embedded into the post as a place where you can report potholes. That to me seems like a great way to reach as many people on the web in Washington D.C. as possible. It exists, it’s not a fairy tale. It’s something that could be done right here.

CHAIRPERSON GARODNICK: Council member Brewer.

COUNCIL MEMBER BREWER: Very quickly when you say real time for your students what does that mean to you and improving the education for the students and how would this legislation help in that sense?

MS. HODES: Well I think that if the

data is going to be up as quickly as we'd like it to be going up the students can access it in their classrooms online and they can see by the week, by the month what's going on in their communities what's changing, what funding their communities are getting, are park benches going in, are trees being planted, and they can be involved as closely as possible or as quickly as possible with the data in what they want to see happening with communities and localizing that with their local government.

COUNCIL MEMBER BREWER: Okay, Copec, I feel like when I die somebody's going to say Gail died to try to get COPEC going because I was there in '89, I wrote that report, I've been part of the discussion for, me and Jean Wistonoff, so you just mentioned the word COPEC I get epileptic but it is a law every single year, COPEC is supposed to publish a book I don't remember if it's in the law but there's a certain timeframe and that's not being followed so it's an issue we will be addressing. Okay.

CHAIRPERSON GARODNICK: Thank you very much to all of you and we're now going to

move on to our next panel which will include Rachel Faust from Citizens Union, Sam Brookfield, and Tim Hofer, Manhattan Institute. And if you have copies of your testimony please provide them to the Sergeant and he'll have them in front of us. You want to go right ahead and get started?

RACHEL FAUST: Sure. Good morning, Chair Garodnick. My name is Rachel Faust and I'm the Policy and Research Manager for Citizens Union of the City of New York, an independent non-partisan civic organization of New Yorkers that promote good government to advance political reform in the state. We thank you for holding the hearing and thank Council member Brewer for continued leadership for increasing public's access to government information through technology. We continue to believe that it's critically important for the city to take major steps that are outlined in this legislation and as we testified before to this committee we do believe similar to others tonight, not tonight this morning that posting this information online proactively can help to eliminate some of the costs associated with FOIA so I'd just like to put

that out there. We've testified in favor of the previous version of the bill so I'd just like to outline some of the changes and what our thoughts are around those.

We continue to support this bill as it creates a single data portal but I'd just like to note that the previous legislation included publications other than data like reports, files, accounts and records. And I understand this could be quite a lot of information but we think ultimately there could be a more unified approach to the city to releasing this information on a single site. I think one thing that we're interested in is publications that would include data like annual reports of agencies that might be cooked or processed if you will that could also be included in this to provide similar types of information.

We also support the bill's effort to have web syndication technology because we think it's important for the public to have up to date information of what's going on. It's up to them to be able to weigh in on the decisions that are being made while they are happening and not after

the fact.

And regarding the standards, technical standards policy in here and the development of them, I think one thing we'd like to add that we'd recommend being added to this bill is that there be an opportunity for the public to comment on the standards similar to regulations that are developed in the city. There should be a public comment period for that.

Regarding the agency compliance plan, the start date of this is July 5th. I think that's a little bit too soon. That's the only timeline in the bill and I think Citizen's Union finds problematic and I'm sure you do as well is introduced earlier so just a technical change. We are pleased the bill's been changed to requiring the agencies to detail the reasons why records are classified in particular categories and I think this will give information both to the council and to the Mayor about technical limitations but also give the public more ease to know that the information is not just being classified in those categories to prevent the release of it. I think that's an important change in the bill.

We also, we'd like to recommend that DOITT publish an annual report on the implementation of this website and we think that possibly there could be a public hearing on the implementation of the website as well. And that's most of my major points.

CHAIRPERSON GARODNICK: Great, thank you very much. I will note that you are correct about the dates in the bill they have been there for some time. I don't think anyone anticipates requiring DOITT to do it in two weeks but we got you. Thank you very much. Go ahead.

SAM BROOKFIELD: Good morning Chair Garodnick and Council member Brewer. Thank you for the opportunity to testify this morning. My name is Sam Brookfield and I work in the technology department at ITAC, New York City Industrial Technology Assistance Corporation. We're an economic development corporation with 22 years experience helping New York City small businesses grow and create high value jobs. We're funded by New York State Foundation for Science Technology and Innovation as the designated regional technology development center for the New

York City region. It's also a manufacturing extension partnership center under a nation wide national institute of standards and technology program. We're one of three centers in the state funded to assist small research and development firms apply for small business and innovation research program. Funding from 11 federal agencies over the past five years, ITAC clients have reported over a billion dollars in economic impact and 4,500 jobs created or retained.

We also run sponsored programs for city companies such as city council funded Move Smart Stay Lean Grow Fast program and NYSERDA funded NYC energy tech program. And on behalf of my colleagues in the tech department, Colleen Gibney and Franklin Madison as well as the President of ITAC, Sara Garrison, I'd like to thank the city council for your consistent and generous support of the Move Smart Stay Lean Grow Fast program.

ITAC supports the community efforts to make city data openly accessible to businesses and individuals alike. We work with numerous technological companies that would greatly benefit

from access to data. We see an excellent opportunity for three sectors in particular. The first deals with education which Ms. Hodes just spoke about. So I will not reiterate that point.

The second is dealing with software development which several people have spoken on also already. To that point I'll just say access to city data would provide software developers an affordable path to bring top notch products to the marketplace. This could have a positive impact on small businesses and innovation research grant applications as well as meaningful data sets and highly desirable to create competitive proposals. As ITAC is one of three NYSTAR funded SPI regional specialist centers we are committed to raise SPI winners coming from New York City.

The third deals with supply chain transparency. Opening up city data to public access would allow local manufacturing and technology firms to see what the city is buying and from whom. In other words it would make the supply chain more transparent. This information would greatly beneficial to such firms because it would provide essentially free market research and

data to companies to which this could be prohibitive.

Such an understanding of the marketplace would allow firms to better prepare themselves for future growth and expansion and would be an especially significant development for young start up companies that may not have the financial resources to conduct market research on their own. We'd like to see the city work with the New York Public Library's Science industry and Business Library as well as local universities to make access and comprehension of this data as simple as possible. Thank you for the opportunity to testify before the council today.

CHAIRPERSON GARODNICK: Thank you. Please.

TIM HOFER: Thank you for inviting me to testify. My name is Tim Hofer. I'm the Director of Operations at the Manhattan Institute's Empire Center for New York State policy. The Manhattan Institute is non-partisan not for profit think tank and the Empire center is their Albany based project that focuses on New York State policy.

Promoting better transparency and accountability in one of the Empire Center's major ongoing priorities. We take a strong focus on ensuring public access to government records and so I would like to begin by commending the Chairman and for your very important work on that accessibility. About two years ago we launched our own open government project. The website known as Seethruny.net, the site gives the public unrestricted access to millions of pieces of public information including searchable databases of state and municipal employee salaries and pensions, collective bargaining agreements, state legislative and expenditure data, expenditures and a benchmarking feature to compare local government spending. To gather this information we filed over 250 freedom of information law requests during the two year process, two year period.

During the process we heard many different explanations or excused for failure to comply fully or on a timely basis with the State FOIA law. Some high profile government entities including the city of New York often complain to us that agency resources are strained by the

necessity of replying to numerous FOIA requests by the public and the news media. Today's technologies, specifically the internet presents a solution to that problem.

We believe that all public information should be more actively disclosed on the internet starting with expenditure, budget and payroll records that will give taxpayers a clearer view of how the bulk of their tax dollars are being spent. This would also free agencies of the time consuming burden of processing multiple FOIA request for different slices of the same material. It's a win win for citizens and for government alike.

A few agencies are already pursuing this strategy. Last year for example as we already hear the State Senate are posting and updating its payroll every two weeks and in a format that's accessible even to those who are not computer savvy which we think is very important. They also began posting their biannual expenditure reports in electronic form. Both of these things are things that we post on seethruny so as you can imagine we're a minimum of less than three FOIA

requests from the Senate every year.

We believe the Senate majority took the initiative in this case and as Andrew said in a short period of time without significant expenditure they were able to develop, implement and format a simple yet effective means to make the data available.

While we commend the bill we have five suggestions which I am probably not going to be able to get to but we believe that you should standardize all public available data in the most simple commonly used electronic formats. We think we should post the records for downloading in a simple webpage link probably to the existing agencies websites and not to one master website. We'd like you to require the immediate posting of all newly generated public records in the same simple formats as a matter of routine. We believe that the updated records for financial transactions, contracts and payrolls should be done as soon as possible.

CHAIRPERSON GARODNICK: That's okay. I'll just put it to you in the form of a question. What are your other recommendations?

MR. HOFER: I just had one more. The other was that non-electronic records should be posted as they are FOIAed starting immediately to take the burden off of posting some of those records that we believe would create the bigger burden on the agencies.

CHAIRPERSON GARODNICK: These five suggestions you believe are not already included in this bill?

MR. HOFER: To some extent we think they are and some they are not, basically the general overall is that we think it is overly complicated, that posting the data from the agency to an agency website will prevent you from having to build a more complicated and timely and consuming data warehouse.

CHAIRPERSON GARODNICK: On the FOIA question, your point was that you think that if a FOIA request is made and if it is a non-electronic record that it should be posted online as it is delivered to the FOIA requestor, is that right?

MR. HOFER: Right. Currently as it stands if you request, the agency that is FOIAed has to deliver you whatever you are FOIAing in the

format that they have so if it's not going to an electronic record you get a photocopy. They are not required to scan it or do anything extra. So the theory is that if you already have to handle this document instead of photocopying it putting it in an envelope and mailing it, why not scan it to your database.

CHAIRPERSON GARODNICK: Okay, council member Brewer?

COUNCIL MEMBER BREWER: For ITAC, please give Franklin our best amongst others, when you mentioned the supply chain transparency, that stuck out as incredibly important. Databases would allow local manufacturing and technology firms to see what the city is buying and from whom. How do you get that information now or is that not accessible? The reason that I ask is that's a constant question for not just the committees but people trying to the research and calling us and so on. How does that information, if at all get transpired to your companies now and I know you work really hard to grow these companies.

MR. BROOKFIELD: As of now,

companies do their own market research, they try I mean it’s very difficult especially for the small start up ones and other than that we work with local manufacturing residents. We have a manufacturing residence that helps facilitate that connecting businesses with suppliers and supply chains.

COUNCIL MEMBER BREWER: So what you’re saying is that even for the, we think of as the obvious savings for the economic development but this would be an extra one where the market research could be something that could be done much more easily.

MR. BROOKFIELD: That’s correct.

CHAIRPERSON GARODNICK: Do any of you want to weigh in on the question about the timeline by which we should be putting this all of this into place. Obviously there are some technical constraints, the bill itself anticipates different categories of data some of which might need to be transferred to the appropriate format some of which are already in that format. Do you have any comment or testimony on the speed or the timeframe that’s either contemplated in the bill

or any of your own? Go ahead.

MS. FAUST: I just have one small thing that I didn't get to in my testimony and that is that for the technical standards manual the bill has DOITT publishing it and at the same time having the agencies do their own compliance plan. I would think it would makes sense to give agencies a little more time after reviewing the technical standard policy to develop their compliance plan so I think there could be a little bit of a buffer there.

And as far as the roll out period for the different years I think it's important to set dates for it given that you know as was mentioned earlier we have some pieces in the law right now regarding COPEC and accessibility that aren't being met so a date is an important way to track and make sure things are happening but I don't think we have a particular sense of what's appropriate for city agencies giving technical concerns but I think the idea of it, the concept of it is sensible.

CHAIRPERSON GARODNICK: So you would agree with the Commissioner that give the

technical standards, give them a chance to weigh in, give the public a chance to weigh in on the technical standards as well but set firm time lines whatever they are and perhaps may not be able to be resolved at this hearing.

MS. FAUST: Yes, I do think that sums up our sense of what would be appropriate for the bill.

CHAIRPERSON GARODNICK: Thank you. Anybody else? Go ahead.

MR. HOFER: Yeah, I think form our perspective we tend to look at records as electronic and non-electronic and for whatever's available that electronic at this point, I can't imagine why each individual agency wouldn't be able to post it in some format within a twelve month period. And by some format I think we're talking about more publicly acceptable things like CSV and text files and I'd hate to see you getting caught up in creating this unified system where all city data is in because that's where I think it becomes burdensome and time consuming.

CHAIRPERSON GARODNICK: Thank you very much and with that we appreciate your

testimony. We're going to call up our next panel which will include Ray Garcia, Andrew Brust, and Todd Stavish. Gentlemen come on up and we will get you started. Thank you very much, welcome. You can go right ahead. Please introduce yourself.

ANDREW BRUST: Very good. Thank you. My name is Andrew Brust and I help run a consulting firm, 26 New York, here in Manhattan. I'm also a technology columnist and blogger and serve on the New York Technology's council advisory board. As I've explained before in previous testimony to this committee, I'm a life long New Yorker and began my IT career in the employ of the government of the city of New York. And I'm going to excerpt myself so I can fit in the time limit. I've testified to this committee before voicing my support for open government data. I'll reiterate today that I feel the benefits of publishing data from all city agencies are huge.

One thing I'll say is the prospect of opening each data stream in each agency might seem daunting to city IT professionals. I would

encourage DOITT and the individual agencies to conceive of the requirement with the right mind set. Data feeds are just software services and good software is built on the premise of designing a service layer at the foundation. So rather than taking the approach of building closed systems and opening them up the agencies should premise the architecture of their systems on building the services and feeds first and then layering the application logic and functionality on top of them. With this approach open data would become byproduct of normal software development rather than a burdensome discreet step. Ultimately the thing to remember is that data is raw material which the city government can refine only to a certain extent. Making the raw material available to the public allows a far greater amount of refinement of value to be added to that data than can be had by keeping it sequestered within the agency that has collected it.

The city can directly benefit from its own open data and that’s because integrations of systems between agencies will be much better facilitated through a normal data sharing regime

than customized point to point data interchange that will enable streamlined construction of numerous systems. For example the Mayor's management report should be much easier to produce and the notion of a general inquiry system across all agencies for 311 becomes compellingly feasible.

There are also nice possibilities for an enterprising data warehouse, scorecards and so forth. I'm going to skip ahead with about 30 seconds left. Just to make you aware of something from of all people, Microsoft. They've created a framework called the open government data initiative and it was actually built by the organization that works with developers specifically in the U.S. public sector and federal, state, local government.

It's already carrying data from the Bureau of Labor Statistics and other agencies. It allows for download in CXV, excel and KML formats and it's also Section 508 compliant. So look into that and I'd be happy to answer other questions when the other testimony's done. Thank you.

TODD STAVISH: Hi, my name is Todd

Stavish, I represent a company named Socrata, I'd like to thank the council for letting me speak today. Just a few key points of my testimony, we definitely agree with disseminating public data. It's the right thing to do. Doing so hold the government accountable, improve efficiency and reduce costs and ultimately stimulates economic growth as some of the other testimony has shown.

There's no need to build an open data solution from scratch and in our mind Socrata offers a purpose built open data platform empowering governments and other organizations large and small to share their data in the widest array of data consuming audiences.

We've proven that for major cities like Seattle, Chicago, as well as some federal agencies as well as some states and counties. Socrata delivers a configurable, customizable platform as a cloud based software as a solution. We are a market drive service provider and each organization invests a fraction of the cost to deploy a platform that represents that particular area of the government.

Organizations benefit from this

evolving platform as a monthly service subscription and our plans range from hundreds to thousands of dollars, definitely much less than some of the projections and costs that some city governments have published. And that's basically it. Thank you.

CHAIRPERSON GARODNICK: Start you fresh, don't worry. Go ahead. The button on the microphone. Perfect.

RAY GARCIA: Thank you for having this session today and allowing us to speak to these issues. My name is Ray Garcia. I'm the Executive Advisor to the Field Center of Executive Entrepreneurship at Baruch College so I have an interest in the entrepreneurial side of this and what it can do to enable developers to access this information and provide for economic development in New York so I'm just going to get to the recommendations, the justifications and rationale for it is in the testimony and I won't go into the detail of that. So the first is that I suggest to the city council to consider mandating a vendor neutral standard for all documents including spreadsheets, presentations, graphics, video,

sound, and all forms of electronic media. This is something that countries have done in response to in reaction to proprietary formats if the city were to adopt it to make it much easier to put that information online and make it searchable and to allow for processing. Second thing is to adopt an open standard for data such as the resource description format schema which expresses an ontology for data which makes it useful for information processing.

Data.gov is a good example right there on the first page, they say the format that they've adopted is a very good example of what can be done with the data. I think that New York City shouldn't reinvent this stuff they should just look at other examples and maybe consider mimicking them.

Third point is to mandate all government agencies to provide interop loads between their systems following the standards. This is a point that other people have made. Speaking I think while the emphasis is to make the data accessible to the public if the data was accessible between the agencies then as a

byproduct would be accessible to the public it makes sense that should be part of the focus.

Fourth point is to require all software vendors providing solutions to New York City to support an open data software format standard and open source. I won't go into to this too much but I don't think you can separate open data from open source. I think these two are one and the same. This hasn't been, I know this is not the focus of this particular session but I suggest that you consider it for future adoption the concept of how does a software, an open software interface and why is it important for open data.

The last point is to set up a registry of people accessing and using the documents and data sets so the city can easily and automatically communicate when updates are available as well provide a directory for citizens and businesses to find providers who have enhanced the data. Thank you.

CHAIRPERSON GARODNICK: Thank you very much. I am going to go back to Mr. Brust for one moment because I think you were just finishing

up a thought about the Microsoft data systems software. I'd like you just to finish your thought on that so we could just understand it.

MR. BRUST: Sure. Again, it wasn't built by a product team in Redmund it was actually built by the developer group that works with U.S. public sector. It's section 508 compliant and it works with an open standard believe it or not, even though it's from Microsoft called OData. The data's published as XML but it's fully queriable, both readable and writable so the discussion you were having before about the read side and the write side that all comes with it and what you get is both the machinery of all data format and the user interface that is section 508 compliant.

It's a starter kit so it's not something you simply deploy as is, it's actually open source as .net open source that then can be modified and implemented as you see fit.

CHAIRPERSON GARODNICK: Council member Brewer.

COUNCIL MEMBER BREWER: Mr. Garcia. My question is I know your incubator program well and how now do some of your companies or the ones

that you're incubating manage to use data? What are they looking for? What would be helpful, etc.? You listed some of the ways it should be conceived but I just want to know what goes on now.

MR. GARCIA: The data's accessed in whatever format's provided and then the developers struggle with trying to understand exactly what that data means and then matching that up with proprietary data that they receive. The more that the data is provided in a format that is similar and the more about information that is the data so it's not just the data that's the data about the data or the information about the data. The meta data is as important as the data itself so the more that is provided the easier it is for [crosstalk]—

COUNCIL MEMBER BREWER: Because they have to understand where it's coming from and the source and so on.

MR. GARCIA: Right.

COUNCIL MEMBER BREWER: In order to do the work that they are doing.

MR. GARCIA: Sure the data

represents the context in which it came to being and without that context we can't carry forward the semantics of what it means is very difficult to access and infer without a body of knowledge and without then going back to the source. It's good that it's a start for the city. And all the people are advocating this, to put text and throw up documents and throw up data but frankly I think the city's going to throw up an enormous amount of data that's not going to be useful to entrepreneurs unless they can make sense of it. Alright, so it's important that the formats be self describing or the data about the data or the information about the data or an ontology more accurately is provided so that there's some sense that can be made.

COUNCIL MEMBER BREWER: That's why it's helpful to hear this because I think different people have different uses from the entrepreneur makes sense as we heard from testimonial earlier to get it up there no matter what. So it's good to hear both sides. Thank you.

CHAIRPERSON GARODNICK: And before

we left you guys I just want to note to a point made by Mr. Garcia. The bill also does require to be in a non-proprietary format, the readability so I just want to make that point. Well, thank you very much to all of you and now we'll call up Richard Stanton, Diana Vitetti and David Weber. Welcome to all of you. Thank you. Sir, would you like to get started since you're settled already?

RICHARD STANTON: Sure, that will be fine. My name is Richard Stanton and I'm the CEO of Bintro. We're an aggregator of classifieds and we rely heavily on access to publicly available data. I reviewed the pending legislation and I am going to make the assumption it will soon be locked in some shape or form.

As we sit here today seven nations and eight states and eight U.S. cities have already adopted open data legislation. There are currently over 270,000 federal data sets available from just a start of four one year ago. There are over 250 applications using these data sets and this is just the beginning.

Some examples include an application that shows the amount of aid given to

each country by the U.S. including detailed facts and news related to that country and the aid given to it. An app to see the adoption of broadband in the United States and things as innocuous as publicly available listings of who is visiting the White House and whom they're visiting. These apps I just mentioned were all built at RPI in Troy, NY. Just an example of what's popping up with these publicly available data sets.

The web has gone through an incredibly evolutionary process over the past 15 years and right now we seem to be in the open data stage. There's an immense appetite to take data especially in semantic form and turn it into valuable applications that range from consumer driven applications and also those that benefit the greater good of our society.

To me, data is beautiful. I liken it to a child that needs to be raised properly with love and good guidance with and without structure, to be socialized with context and to grow to provide back even more to the next generation.

Data can lead us to a cure for

cancer by way of the NIH's ontology. It can help us to find a lost child, Amber Alert data which is publicly available and can hold our leaders accountable for how our tax dollars are spent, Public Funds Research. We are all products of social construction and data is no different. It needs time and attention, it needs to play nice with others, and needs to explore relationships in order to grow so it can live on its own. Data like a child can bring joy, make you laugh but can also make you agonize as well. Simply put, to me, data is organic and we are just in the infant stages.

To most in this room, raw data is valuable data. And as a community of technologists we would be willing to raise the data for this city as we go forward.

On a more practical level the transparency of democratized data is an incredible leap forward for local governments and will bring New York City into the center of what will be a rapid growth movement over the course of this decade. An example of this is making available all government job openings and out placement

services for all laid off government workers. We can also make available public space that the city is no longer using because of the downsizing of the government. This could be used by start ups and entrepreneurs just if they knew about the availability.

From transportation to public safety New York City will see a reawakening from its release of data. As I mentioned we should not underestimate the societal importance of raising a child well and the same can be said for data.

CHAIRPERSON GARDONICK: Thank you very much.

RICHARD STANTON: Thank you.

CHAIRPERSON GARDONICK: Go ahead. Hit that button one more time and you'll be all set.

DAVID WEBER: My name is David Weber, I'm senior member of the ICM, member of the OASIS XML Public Standards Organization, and XML evangelist and a long term implementer of sharing solutions for government applications.

There's basically two types of data that we're considering here, the structured and

unstructured data. People are very used to unstructured data that they see on web pages constantly. The problem with that is that it's difficult for search engines to be able to harvest and link hence harder for people to find consistently.

The alternative is structured data which is used to publish data sets which is freely accessible via the data mine for example. But then not all of that data is created equally as the previous speaker was eluding to. Without a vocabulary, lexicon and approach you end up with a lot of noise in the system that actually over time inhibits access to the data as the mine grows.

So future proofing your data is very and as we know technology moves extremely rapidly so it's very difficult to pick particular flavors of standards only to find that then those are actually a problem later down the road.

So this is a paradox. How do we do data right and standardize how it is done on the one hand and what we don't want to be is prescriptive so that it blocks out new innovations and new technology as we move forward. And then

added to that is the risk of vendor lock in. We all know about that how selecting a limited set of providers that ultimately develop special software that you need to access data provides inhibitors to how you can get everything so using open public standards and open friendly source technology is therefore key so an earlier speaker talked about that.

So rather than adopting open vendor API which initially may be alluring you have to be very careful then that's these are not then the sole sources and that people have direct ways of getting the data if they need to. And notice that API's can also harvest data and another speaker mentioned this so you can have people register what information they're interested in but you can also see who is requesting your data.

So there's a lot of big challenges here. I want to mention the NIEM initiative, the National Information Exchange Model, should look into that and how's that's gone about providing a common platform for federal government to share information and similarly OASIS worked with the state of California on election record reporting.

And this had a lot of interesting aspects that people just don't see when they start sharing information. Who's the authoritative source? Who has access to it? And when and how?

CHAIRPERSON GARDONICK: Thank you.

DIANA VITETTI: Hi, good afternoon. My name is Diana Vitetti, associate director for Common Cause New York. Common Cause New York is a non-partisan advocacy organization that fights for increased transparency and honest and accountable government. Thank you for the opportunity presented here today to allow us to speak about how government transparency can be expanded through the creative use of developing technologies.

Many local governments nationwide are figuring out how to use the internet to make government data more accessible. The goal is to utilize the technological power and usefulness of websites and mobile applications and even perhaps format change of how citizen's think about their city and its government.

Open data models lend itself to being a more inclusive and more transparent

government, cornerstones of our democracy.

Initiative number 29 before us today would further the slated goals of local law 11 which was first introduced by former Chair Brewer in 2003 to position New York City as the leader in the nation to using technologies to improve the efficiency and accessibility of municipal government. The provisions of this bill that would make data sets publicly available through linkage with a city web portal in a manner that is easily accessible promotes the public interest by allowing data sets to be meaningfully reviewed and utilized by the constituencies.

When I was walking into this room I noticed that most people here actually had out either and iphone or blackberry. I've used one as well so how great would that be if we could be able to harness all of that into looking right now at different data sets and actually be focused on interacting with our city governments.

The provision that all public records be made available in the raw and unprocessed form is the right step in making sure that the integrity of the data sets remain intact

and there is no perception that data has been aggregated or compiled in any subjective manner.

The intention of this bill is that all public records shall be updated as often as necessary to preserve the integrity and usefulness of the record. It also helps to maintain the continuous flow of open data to the public and creates a paradigm for best practices for city agency reporting. Too often data sets are outdated, they are not updated in real time, creating obstacles for those who are looking for information that will help better serve their communities.

Whether it is information regarding property sales, department and building issue permits, or as an organization such as ours that looks to analyze to compare data results and to look at performance metrics of different city agencies.

However some of the recommendations that we have for the committee today, I think one or two are already reference that we would like the committee to consider amending the proposal to require the record policy and technical standards

drafted to specifically address mechanisms for public input and oversight regarding any shortcomings of the data available. This would compliment the affirmation provisions already in the bill that seeks to maintain the data's integrity.

This could be done relatively easily such as simple comment features that you find on blogs or online submissions forms are simple ways to allow the public to provide their thoughts or concerns to the relevant agencies. Thank you for your time.

CHAIRPERSON GARODNICK: Thank you. Let me just follow up on that last point since you didn't quite get to finish it. You want to add some sort of a opportunity for the public to comment on the quality of the information that's being put out there. Is that correct?

MS. VITETTI: Correct. It would also, I mean that I think that if the public does see discrepancies in datas or if there is reason to believe that certain information is put out there should be some kind of public comment feature that allows a more thorough interaction of

the public with the city agencies. I mean if the reason behind this and we all know is to not only increase transparency and government accountability but also decrease requests on FOIA, a good way of doing this would just be having and interactive feature right there that allows constituencies to register their complaints or concerns and get feedback right away.

CHAIRPERSON GARODNICK: Have you seen the public comment element on data mine that exists and the Commissioner testified about earlier today.

MS. VITETTI: No, I have not.

CHAIRPERSON GARODNICK: Okay. I was just going to see whether you had any comment on whether that satisfied any of these concerns or whether you think that we should be going further but we'll leave that for another day.

MS. VITETTI: I'll look on that.

CHAIRPERSON GARODNICK: Okay. Council member Brewer.

COUNCIL MEMBER BREWER: Thank you. Thank you Mr. Stanton. You have obviously worked on some of the other data sets around the country.

MR. STANTON: Yes.

COUNCIL MEMBER BREWER: Which ones, or just generally what are some of the aspects that work that you would like to see replicated here in New York just generally.

MR. STANTON: Well, I think someone previously said that replicating the format of RDF which is coming out of data.gov would be a wonderful thing to do. I think though that in terms of time and interest. Just having raw data will allow entrepreneurs to figure out what to do with it. You know I could sit here and come up with probably a thousand different applications for the data in my own mind, you do that across tens of thousands of developers and people and they're going to come up with things that are going to be wonderful.

The bottom line though is that there are a lot of public safety issues that would benefit from having multiple interactions of users, understanding the data that's out there, understanding the relationships of that data. Just to give you an example of that someone pointed earlier to on ontology in New York. We

like to say our company, it's very important for us to know the difference between Madonna the musician and Madonna the religious figure or the Bronx Bombers being the same as the New York Yankees. All this data, all these relationships, all this understanding comes from being able to take raw data and glean from it these relationships and understanding and context.

So I think that because there's not a lot of technologists maybe involved in the process of creating the law itself, it may not be understood that just getting it out there can get us started to a point where wonderful things will flourish and I think that was best illustrated when I said that over 275,000 data sets are now available and it's growing very quickly and I do spend a lot of time in Washington, recently talking about this both to State department and other organizations that are moving so quickly, it's incredible at a federal level.

COUNCIL MEMBER BREWER: Did you want comment, sir? What works?

MR. WEBER: Push that twice. First of all I did just email you my testimony for you I

apologize I didn't earlier print it out. I am too much XML. Yes, what I see is that this complexity challenge is a really difficult one. Working with the federal government in the NIEM initiative on standardizing definitions of things and getting consistent formats is really really important. I know there's a big push to get data out there in whatever format that you can. And I hear that but conversely the simple measures that you can take early on in the process so that you don't end up with a big mess upfront, and what we're talking about here is not rigid standards but flexible ones where things are named in a coherent consistent way so that you have predictability about the information that you're seeing. And the other aspects of this, obviously this XML, this many different flavors of as this gentlemen was mentioning, this RDF, you know don't bet the farm on one thing.

What I heard earlier we talked earlier about spreadsheets and that's one area that you can publish data with but a little known fact is and I included this in my written testimony that you can actually build spreadsheets

that are interoperable with XML so you can view that data either which way. You can download the XML, view it as a spreadsheet.

These are very empowering things that you don't want to limit by saying we only want to take these particular flavors so what NIEM has done is provide a broad range of guidelines to developers to follow to ensure consistency across the community and the use and development of open standards and what I've particularly been working on for the federal government is building open source software to facilitate that. I know you heard that earlier but I can't stress how important that is because then you have resources that anyone can access and use not only externally but you yourselves.

You talk about cost of developing this for New York and if you focus your efforts and monies on building tooling that will really accelerate what your own internal people can do to develop this and again I mention that California state election law, they were going down a hellish convoluted path that involved a year's worth of development and we were able to show them how to

do that using open source and simple approach in a few weeks so it’s very important.

COUNCIL MEMBER BREWER: Well thank you very much.

CAHIRPERSON GARDONICK: Thank you, one last question for you from Ms. Vitetti. We hear a lot today about ways that this data can be used by web entrepreneurs to make more accessible and user friendly. You noted that the purpose of Common Cause is to strengthen public participation and faith in our institutions of self government. Can you say a little bit about how you think this might empower communities or help the good government process?

MS. VITETTI: I think I would have to say that I could also see it from a good government point of view and also as someone coming from an elected official’s office. I’m understanding frustrations of constituencies when they would come regarding requests they had given to FOIA, waiting for those requests and there was a lot of resentment feeling that this data and information should be readily available to them. They shouldn’t have to go to request it they

shouldn’t have to come to an elected official’s office to follow up on it. I think also from a good government perspective having this data readily accessible even for us cuts down on the amount of time and the resources having to track down certain data, have to find it in the raw instead of looking at aggregates and looking at trends over time.

There is also a way for constituencies and I think it works for both the government itself and the constituencies to look at performance evaluations for agencies, to look at case loads, to look at number of people that were served, to look at over time and then even you know from a good government perspective we might compare that let’s say what budget for the organization or what money is being spent on and we’re talking about the public’s monies, we’re talking about the public’s trust. I think this changes a lot of the perspectives on the ground about what the government is or not making available and just that perception itself that we see so prevalent especially nowadays will definitely help by creating a government that

people see as open, people see as responsive to their needs especially in local communities.

CHAIRPERSON GARODNICK: Great. Well thank you very much and thanks to all of you. Now we'll invite our last panel up which is Dylan Golds and Thomas Lowenhop [phonetic]. Come join us. If there are others who wish to testify and have not yet been heard or not yet filled out one of the forms please do so now is your opportunity. And let's go right ahead and get started. Welcome.

DYLAN GELTS: Thanks guys. My name is Dylan Gelts and I work with startup in Brooklyn called Roadify and using open data we've been able to improve public transportation and minimize traffic in Park Slope Brooklyn. We're expanding quickly—it's an ongoing process—so Brooklynites are currently using Roadify to access and update bus schedules through text messages as well as share information about open parking spots to get cars off the road faster. And our bus platform uses the MTA's released information which is a good case study and is simply based on riders reporting a bus location so that other riders

waiting down the line could get a better idea of when it will arrive at their stop.

As I said we aggregated that real time user update side of it with the MTA's schedule to create a more flexible, more dynamic bus schedule that makes riders better informed and more participatory in their commute. That's why we say we're bringing community to their commute as my bright neon shirt says.

At a time when pains over cuts are running high the accessibility of the MTA data has allowed Roadify to help alleviate some of that strife and by releasing some of that data to developers and entrepreneurs like ourselves, government bodies are not only helping their constituencies but themselves as well. Promoting access to innovative and popular technologies allows governments to run more efficiently and offer solutions that otherwise wouldn't exist. I'll close just to say inherent to the democratization of information is the idea of participation and governments benefit when people participate so give us the capabilities to do that and we'll put it work for you. Thank you.

CHAIRPERSON GARODNICK: Great, thank you very much. Hit that button one more time.

TOM LOWENHOFF: Good afternoon. I'm Tom Lowenhoff, Director of Connecting.nyc a New York state not for profit advocating for the development of the nyc. Domain as a public interest resource. My presentation is on the DNS query log a soon to arrive database. By way of background within the next few years the internet is going to change in fundamental ways. It is going to be more intuitive, this will happen as the eye candy that issues the more top level domain such as .com, .org and .gov finalizes an application process. There will initially be hundreds then thousands of top level domains with names such as .net, .sports, .news. So the future holds Chase and Citibank moving from Chase.com and Citibank.com to Chase.bank and citi.bank. Espn will move to Espn.sports and the Wall Street Journal will find advantage in moving to wallstreejournal.news. With this transition people will come to see that the internet is far more intuitive than today and will begin entering

their domain name request directly. For example if you're looking for a bank you are likely to enter an index of .bank and directory of .bank or if you're looking for the news sources you might go to categories.news. And information about baseball might best be found on baseball in sports. It's going to be a different internet one way or another. Our dependence on search engines will be diminished.

In addition to the aforementioned sport news bank there will be city such as .paris, .berlin and my favorite .nyc. Imagine the nyc top over the main is fully functional in five years if people have come tor recognize the benefit of entering domain names directly rather than relying on Google. So people learn that it's faster and more direct to enter Mayor.nyc or citycouncil.nyc, firedepartment.nyc and police.nyc. The operator of the nyc tlv will connect each of these queries to the appropriate website and create an entry in a query log database. This query log will contain valuable information for marketing governments and civic life perspective.

Let me give an example imagine in

19985 we had an intuitive internet as I described it today. Baseball.sports, police.nyc and imagine the residents of Greenpoint Brooklyn started in entering inquiries such as hole in tree.nyc, spotted beetles.nyc, dying trees in Greenpoint.nyc, what happens to these queries? If they are for an existing website they will go directly to the site and I'll skip for a moment to privacy issues associated with that database and imagine it's a time like 1985 when the American, Asian long horned beetle has just arrived on our shores and residents of Greenpoint are entering intuitive inquiries so you can give information about the strange developments going on with their trees. Let's assume that none of these initiatives have existing websites, what happens to these erroneous queries. We advocate that this information can go into an error query database and be made available for all to inspect. Some clever researcher—you have the rest of it there.

CHAIRPERSON GARODNICK: Just say, if you could just sum it up really quick we'd really appreciate it.

TOM LOWENHOFF: Alright so the

question is. This database is an error query log database would be a twitteresque database that the city would own and could use to find out what's going on and we could feel the pulse of this city by finding out what people are entering into .nyc.

CHAIRPERSON GARDONICK: Thank you very much. Let me just pose one question to Roadify. First of all we appreciate your creativity and how you managed to keep things moving in Park Slope. Help us understand what this bill would mean for you all if passed.

MR. GELTS: Yeah, so first off I forgot to mention that we were in the Big Apps competition and that was a fun experience for us but the more data that we have in terms of for us transportation is key the better we can service our public. We have big plans with including the alternate side of the parking data that can partly be found in the data mine currently through the RSS feed and such. We have big plans of including that into our systems so we can tell the 2000 people in Park Slope that participate in Roadify to better inform them so that the Sanitation people can run more efficiently and street

sweeping and stuff like that.

CHAIRPERSON GARODNICK: Council member Brewer.

COUNCIL MEMBER BREWER: Alright now let me just be devil's advocate because I can also go online and this is, figure out alternate side of the street parking but I assume you want to layer it that would be what you would do, in other words you have not just the alternate side of the street parking but you have information about other opportunities in other words if the bus changes and all these other things. How, can you be more specific even than what the Chair asked about how this information was available with all databases you could actually use and in what format. Very specific.

MR. GELTS: Absolutely. I'll be as specific as can be. I work in the space, I went and waited on an F train for 2 1/2 hours on Saturday all because I didn't go to the MTA website. And it's got to be more accessible and that's why we're doing it. We started with the text messages because they're the most ubiquitous most common. We're layering on not only

information but iphone applications, better web accessibility but it’s got to be more transparent and we’re working to aggregate all these different places of information into a more easy to use and accessible manner.

CHAIRPERSON GARODNICK: Thank you both very much for your testimony and that is the last panel of the day so I will conclude with a quick thought and give Council member Brewer and opportunity just by saying that we agree with Mr. Stanton, data is beautiful and this is an opportunity for us to truly expand the way that both the constituencies out there as well as community groups as well as web entrepreneurs can harness what is existing out today and put it to much better use so we look forward to making sure that we get this bill passed and passed quickly and I give you council member Brewer for some final words.

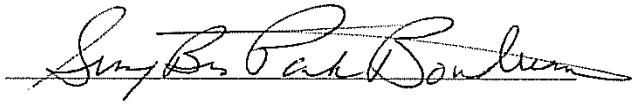
COUNCIL MEMBER BREWER: Thank you very much I think that New York will be a very exciting place to do this partly because we have data that is more diverse and I think we are a well managed city and it would show the aspect

that we are well managed and could take—we have so many entrepreneurs here and entrepreneurs I think who r understand the importance of serving their communities so this particular data would have so many important aspects to it and it will be very exciting so it's great to work with the Chair Garodnick and with the administration but finally I just want to reiterate one more time to thank Kanal Mahach and thank Sam Wong for all the work in putting this wonderful hearing and certainly the same over the years and Lou Klettner and Joely McPhee and certainly Jeff Baker from the committee and all those in the IT division who make it possible. We've had in this committee, we had the first tweet that went live and we've also had the first audio video in the other committee. We're kind of behind in the council but we're catching up but the data would be real and we look forward to working with you.

CHAIRPERSON GARODNICK: Thank you council member Brewer and with that this hearing on the Committee will be adjourned.

C E R T I F I C A T E

I, Sung Bin Park-Boudreau, certify that the foregoing transcript is a true and accurate record of the proceedings. I further certify that I am not related to any of the parties to this action by blood or marriage, and that I am in no way interested in the outcome of this matter.

Signature_____ 

Date _____July 15, 2010_____