CITY COUNCIL
CITY OF NEW YORK

----------X

TRANSCRIPT OF THE MINUTES

of the

COMMITTEE ON TECHNOLOGY IN GOVERNMENT

----------X

June 29, 2009
Start: 01:20 pm
Recess: 03:44 pm

HELD AT: Hearing Room
250 Broadway, 14th Floor

B E F O R E:
GALE A. BREWER
Chairperson

COUNCIL MEMBERS:
Bill de Blasio
G. Oliver Koppell
Letitia James

A P P E A R A N C E S

Sam Naim
Assistant Counselor
To the Mayor

Ariel Dvorkin
Special Assistant
For Policy and Governance

Andrew Hoppin
CIO
New York State Senate

Ian Jacobs
Head of Communications
W3C

Benjamin Kallos
Co-founder
Open Government Foundation, Inc.

Philip Ashlock
Open Planning Project

James Vasile
Legal Counsel
Software Freedom Law Center

Morgan Reed
Executive Director
Association for Competitive Technology

Chris Keeley
Associate Director
Common Cause New York

DeNora Getachew
Director of Public Policy/
Legislative Counsel
Citizens Union of the City of New York

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

Rachael Fauss
Research and Policy Analyst
Citizens Union of the City of New York

Joshua Breitbart
Policy Director
People's Production House

Frank Hebbert
Manager of Geographic Information Systems
Regional Plan Association

Thomas Lowenhaupt
Founder
Connecting.nyc Inc.

Silona Bonewald
Director
League of Technical Voters

Kayza Kleinman
NonProfit HelpDesk
Jewish Community Council of
Greater Coney Island

Andrew Brust
Chief of New Technology
Twentysix New York

Jay Sulzberger
Consultant

CHAIRPERSON BREWER: Good morning, I'm Gale Brewer, a City Council Member and head of the Committee on Technology in Government. This hearing is starting and we're here to talk about open government data. I'm joined by Council Member Bill de Blasio and I know other members are on their way. This is a very hot topic. At what we call Andrew Rasiej's Conference, but it is called Personal Democracy Conference where some of us were this morning, the Mayor in a video mentioned that he too was interested in this topic. So, needless to say, it's something that I think New York City can be cutting edge. It follows on the interest that I've always had on making sure that government data is public. Many thanks here today to Jeff Baker, who is counsel to the committee; Colleen Pagter, who is the policy analyst; certainly Junal Malhotra and Sam Wong from our office. So, just to open, talking about open government data, it's the concept that the publicly accessible data generated by the public sector should be available to the public electronically via the internet in open raw formats. I think "raw" is the key word there.

Adopting open government data standards would make access to information easy and affordable while also promoting transparency and accountability. Open data permits deeper and more varied analysis of government data by enabling two or more datasets to be integrated together. In December 2007, a group of 30 open government advocates, and I think that actually translates to a lot of companies, developed eight principles that define open government data. They include complete, that all public data should be made available. Public data is data that is not subject to valid privacy such as personnel records or health records that are personal or security issues or privileged limitations. Data should be timely. Data should be made available as quickly as necessary to preserve the value of the data. Data should be accessible. Data should be available to the widest range of users for the widest range of purposes. Data should be machine-processable. Data should be reasonably structured to allow automated processing. We introduced number 991, known in City Council terms as Intro 991 of 2009. It requires all public records to be made

available on the net through a single web portal formatted to enable viewing by web browsers and mobile devices and also in raw or unprocessed form. This bill would mandate that records be made available without any registration requirement, license requirement or restrictions on their use and be presented and structured in a format that permits automated processing. It requires DoITT, which is the City's Department of Information Technology and Telecommunications, to promulgate rules establishing an internet record policy and a technical standards manual for the publishing of public records on the net by city agencies by January 2010. In other words, there would have to be an analysis and a plan for how each agency would conform to Intro 991. Basically the legislation requires each agency in consultation with DoITT to review the public records under its control and to develop and submit an agency compliance plan to the Mayor and the Council no later than January 2010. In addition, agencies must classify the public records under their control as immediate, priority and legacy. Immediate is any public record that

can be made available on the net within 30 days of the agency acquiring or creating such record. Legacy is any public record that due to its size, complexity or technology constraints cannot be made available on the net by July 4th, 2011. All other public records shall be classified as priority. Now, the good news is that this topic is being discussed I think in New York today. It is also my understanding that this particular bill goes into legislative format whereas other states and cities have certainly, such as Washington, discussed this but it's been a policy regulatory and not legislative. I also want to say that at a Participation Camp workshop at New York University this weekend with Matt Coprider [phonetic], Sam Wong and I were there and talked about this bill. I know some of you in the audience were there and I appreciate your participation. I think there were lots of great ideas as to how this information can be used. So we will hear from the administration, but we were fortunate that in that workshop was Craig Newmark who is always trying to deal with Craigslist. Obviously one of the issues that we all would like to get more information on

is housing, where is the mold, where is the affordable housing, where are ways in which we can use this data that would help people who are trying to deal with the housing constraints in the City of New York. Of course, there is no end, and we'll talk about them later on, as to the type and quality of the data that could be available to New Yorkers and in fact everyone who is on the net. Without further ado, I'd like to call up the administration and ask them to participate in this hearing. Is somebody here from the administration that is going to testify? Thank you very much. Why don't you introduce yourself? We know that the Commissioner of DoITT was sitting next to me early this morning, so all he has to do is sit and listen all day and you guys actually have to talk. We appreciate your being here. Thank you.

SAMI NAIM: Good afternoon Chair Brewer and members of the Committee. I am Sami Naim, Assistant Counselor to Mayor Michael R. Bloomberg, and am here to testify on behalf of the Administration on Intro 991. Joining me today is Ariel Dvorkin, Special Assistant for Policy and Governance. Transparency and accountability have

been the cornerstones of the Bloomberg Administration since day one. Indeed, over the last seven years, we have worked hard to provide New Yorkers with the information they need to hold their government accountable. A few recent examples include the NYC Stimulus Tracker. It's a new tool that allows citizens to track how the City is using funds that stem from the federal American Recovery and Reinvestment Act. This website enables the public to see what projects are being paid for with stimulus dollars, when the contracts for these projects are approved, and which contractors are involved. NYCityMap is an interactive map feature that allows users to search by any address, intersection, community district or City Council district, as well as by hundreds of place names, and to add desired layers atop these chosen locations, such as aerial photos of the City, building and property information, poll site locations, census data, neighborhood health profiles and statistics, restaurant inspection information, locations of educational facilities, and locations of transportation hubs, and much more. Citywide Performance Reporting is

a revolutionary online tool that gives New Yorkers access to critical performance measures from more than 40 agencies in an easy-to-understand snapshot format. CPR was recently nominated for a prestigious "Innovation in Government Award" from the John F. Kennedy School of Government at Harvard University. ESubmit is a new program offered by the Department of Buildings that allows developers to submit required items and supporting documents electronically, facilitating public review of construction projects in neighborhoods across the five boroughs. ARISParent Link is a Department of Education service that we are developing which will allow every parent or guardian of a student attending public school to track their child's current and past academic performance. And just this morning Mayor Bloomberg announced five technology initiatives to improve accessibility, transparency, and accountability in city government. One of the initiatives that we are particularly excited about is NYC Big Apps, the first annual software application, which we believe will enhance the transparency of city data. Modeled on Washington,

D.C.'s successful Applications for Democracy initiative, NYC Big Apps will encourage the use of publicly available, operational city data to develop open source applications, commonly referred to as "mash-ups", that will benefit New Yorkers, improve business efficiency, and save taxpayer dollars. With NYC Big Apps, developers will be encouraged to create innovative applications with nearly 80 raw datasets from across 32 different city agencies and commissions to make living, working, and playing in New York City easier and more enjoyable than ever. The bill before you today is intended to build upon the progress that the city has made. We commend the Council for its desire to make government more open, accountable, and transparent. However, Intro 991, as currently drafted, raises a number of complex operational and fiscal issues that are of concern. First, the bill's definition of "public record" would require a full review of the city's record holdings to determine which records would be required to be released to the public pursuant to the Freedom of Information Law or FOIL. Currently, the city stores at least 2.5

million cubic feet of documents in offsite facilities, in addition to the records that exist in agency offices and facilities. This roughly translates into 6,250,000,000 pages of documents. Second, the task of reviewing these documents would fall on agencies with limited resources. Currently, agencies dedicate at least two people to review documents pursuant to FOIL, a Records Access Officer and Records Access Appeals Officer. These officers are generally city attorneys who serve their agency in a number of ways. They advise policy makers, draft and review contracts, respond to legal claims against the city, and, of course, respond to FOIL requests and appeals. In order to comply with Intro 991, these officers would have to review all of their agency's record holdings to determine which records are proper for disclosure. For example, whether the release of a record could, by reason of its disclosure, violate a legally recognizable privileged communication, such as an attorney-client communication, undermine a law enforcement investigation, publicize a commercial trade secret, or invade the personal privacy of an individual. This is a

Herculean task to say the least. Third, once this review is complete, the bill requires the records slated for disclosure to be converted to electronic format, so that they can be posted on the internet. According to the Office of Management and Budget, the cost of scanning all these documents is estimated to be more than half a billion dollars. Finally, once the documents are slated for disclosure by agency FOIL officers, and converted into electronic format by administrative staff, IT personnel must then construct and maintain an infrastructure capable of accommodating these records. The Department of Information Technology and Telecommunications has informed me that they would not only need to hire substantial additional support staff and information architects to implement this bill, but they would also have to construct an entirely new records management system with disk storage and data processing power capable of meeting the demands of Intro 991. While we clearly support developments to make government more transparent, accessible, and accountable, the placement of a seemingly limitless amount of data online is not

necessarily the most effective approach, especially at a time when public resources are stretched to the limit. Indeed, the vast majority of the records currently warehoused by the city are rarely the subject of FOIL requests and it is unclear whether placing them online would make them of any more interest to the public. Therefore, it would seem not to justify the extreme expense of doing so. Of course, moving forward, as we develop new record retention strategies, opportunities to make new data available can be harnessed in a more cost-effective manner. Indeed, NYC Big Apps is the beginning of this kind of forward thinking. Accordingly, while we agree with the spirit and intent of Intro 991, we cannot support it for the operational and fiscal impacts it would impose. Thank you for allowing me to testify, and I would be happy to answer any questions you may have.

CHAIRPERSON BREWER: Thank you. At the workshop at NYU, I said that's what you were going to say. I appreciate your coming here to say it. The question I have is do you have any knowledge of what the federal government is doing

along these lines? Are you aware that President Obama is with a much larger budget trying to accomplish what Intro 991 is laying out?

SAMI NAIM: Just this morning, in respect to the five initiatives that were announced this morning by the Mayor.

CHAIRPERSON BREWER: I was there when he announced them.

SAMI NAIM: The CTO for the White House, Aneesh Chopra, said that these five announcements align well with President Obama's open government initiatives and that these practices that we are undergoing are best practices which should be replicated by other governments across the country. I think we're in step with what the federal government is doing. I think we're providing that meets user preferences and user needs. I think it's a successful strategy and recognized by the White House as being so.

CHAIRPERSON BREWER: The other question I have is when you talk about FOIL; having worked in an administration myself I know that there are really a lot of FOIL requests.

Sometimes people FOIL all agencies. They file individual agencies. It depends on what they're looking for. How much do you think FOIL costs on an annual basis versus if you put that information up? Would that not eliminate the need to do as many FOIL requests as are requested? You didn't mention the cost of FOIL in your testimony.

SAMI NAIM: One thing I would say is that this sort of amounts to a FOIL request of all records, which is a tremendous task. Again, more than six billion pieces of paper have to be reviewed. Our approach to transparency initiatives generally is to build them along a customer service model and to tailor them to user preferences and needs in order to provide New Yorkers with information that is meaningful, easy to understand, easily accessible and which keeps them, informed and engaged with government.

CHAIRPERSON BREWER: Do you know that other cities, without legislation, are doing what Intro 991 lays out? Washington, D.C. is an example. It's obviously a smaller municipality. Have you looked at any of their policies?

SAMI NAIM: We definitely have

looked at the Applications for Democracy, which has proved to be a great model for NYC Big Apps. Again, that was part of the inspiration to create this open data source competition that we're pursuing in the next year.

CHAIRPERSON BREWER: But when agencies provide data, don't they scan it in sometimes so it wouldn't be so difficult to put it on a computer and get it out to the public? How is data accumulated now, say for instance on the homeless? I sit on General Welfare with Council Member de Blasio and there is a great deal of data that those who are homeless and anything that comes through HRA accumulates. How is that collected now? It's my impression that internally that data is available.

SAMI NAIM: What we do now is we try to engage New Yorkers and speak into their native tongue, which is plain English. So we provide data according to the public's preferences and needs and to respond to their specific needs.

CHAIRPERSON BREWER: What you're saying is it is sort of complaint-driven or request-driven?

SAMI NAIM: A lot of times we preempt those requests. For example, 311 Online, we're providing a lot of information online so that New Yorkers don't have to call 311 to find the information they need. Another example is a recent partnership with Google where we will be able to track what New Yorkers are looking for, what information they're searching for on the internet to be able to provide that information to them before they have to come to us. We think this represents also cost savings as far as the call center volume. I mean if we reduce the call center volume by 10%; there are annual savings of over $4 million.

CHAIRPERSON BREWER: I feel very strongly of the way which is that the public doesn't know what data they need because they don't know what the raw data is. I think that it shouldn't be up to all of us in government to decide what the public should have access to. Obviously you have to take certain security and other personnel and health issues into consideration. But I really do think that the raw data should be available to the public. They're

actually paying for it.

SAMI NAIM: We design our applications around the public, not around what we want to see or what we want out there. An example is Citywide Performance Measures. A lot of times they may show negative trends in street cleaning for example. We are able to identify areas that need improvement and adjust our operations and resources accordingly. So we rely on this information as much as the public and we believe that good information makes for good management and good public policy and good government.

CHAIRPERSON BREWER: Let's go back to the FOIL issue. Do you have any sense of how much it costs to provide two or so FOIL officers per agency per year?

SAMI NAIM: The salaries of those particular officers?

CHAIRPERSON BREWER: Or the time involved. Do you have some sense of that? In many cases that would not have to exist in such a degree if in fact the information was up on the net.

SAMI NAIM: As a former records

access officer, I could tell you that I was working on various matters, land use, constitutional law, administrative law and FOIL. It does encompass nights and weekends.

CHAIRPERSON BREWER: Yes, it does.

SAMI NAIM: But again, FOIL requests aren't this size in magnitude. FOIL requests exist in a point in time. It's hard to have a hard and fast rule when it comes to public records because government does so much and so much information is memorialized in different ways. So I hesitate to say whether or not like a particular public is without a doubt no subject to any exceptions of FOIL.

CHAIRPERSON BREWER: I think that's another thing to be discussed. The public can develop insights into government information and use that information. There are things that you might not think of that the public might be interested in. Is that something that you have thought about as you developed the mayor's five policies today?

SAMI NAIM: Definitely. NYC Big Apps is a great example of taking public data and

incorporating it into your own daily life. For example, if I'm a new parent and I want to take Parks data to find out where playgrounds are. Maybe I want to also take my child to a restaurant afterwards. I can get private data of child-friendly restaurants and really plan my day accordingly. It's a catalyst for innovation and a catalyst to also see what the public is interested in and how they would use public data and incorporate it into their daily lives. It's an exciting time and it's very empowering for a daily citizen.

CHAIRPERSON BREWER: The public wants to know how clean the restaurant is, what have been the past signs in terms of their inspections. That's what they want to know.

SAMI NAIM: The DOH restaurant inspection data will be part of Big Apps.

CHAIRPERSON BREWER: I know. But you want to have the background, not just for that but how clean the parks are. It will be something that you want the raw data.

SAMI NAIM: All that data is online in a quickly accessible, convenient, and easy to

understand formula.

CHAIRPERSON BREWER: How many agencies are on? EDC, Health and Hospital Corporation, NYCHA and some of the agencies that are not particularly city agencies, are all of those agencies on? Is DOE going to be part of your Apps in terms of what you're considering?

SAMI NAIM: I know DOE definitely will be part of our Apps in the early materials that I've seen. Again, we announced this data. We're going to roll it out later in the year. We want this to be as helpful as possible. Again, we're driven by the customer. So if it's helpful to the customer then that’s the direction we'll go.

CHAIRPERSON BREWER: Will EDC be a part of this also?

SAMI NAIM: I'm not sure as of now. But, again, we're driven by customer needs.

CHAIRPERSON BREWER: But people don’t know what EDC is so how can they ask? In other words, what I'm saying is the reason that I want the raw data there is because the customers don’t know what is available from city agencies.

They may know about Parks and they may know about schools, but EDC would be an example where people don't really know what EDC does. You're never going to get a customer stating in the same way.

SAMI NAIM: The goal of NYC Big Apps I should say is to provide as much information to the public that they can embed in their daily lives and use for their own purposes. So, to the extent that any data that's out there that we have, we'll certainly try to incorporate into Big Apps.

CHAIRPERSON BREWER: What about environmental impact studies from City Planning Commission?

SAMI NAIM: I think, again, we want to be driven by the customer.

CHAIRPERSON BREWER: The customers want the EIS.

SAMI NAIM: We want to give information in an easy to read, easy to understand and convenient format. I mean I would have to take a look and see to what extent we could provide EISs or what have you. But again, it's driven by customer needs and we want to address

those specific needs as quickly as possible. That's part of what customer service is all about.

CHAIRPERSON BREWER: This is a question that we've asked DoITT a couple of times, but is each city agency responsible for the publication of its own city government data policy, or does that come under the Mayor's Office or under DoITT? How does each agency determine right now publication of its data?

SAMI NAIM: I think we've done a lot of interagency coordination as of late. NYC City Map is a great example. Each agency used to have their own map features and we combined them in a one-stop shop so you can layer 311 complaints and capital construction--

CHAIRPERSON BREWER: [interposing] But you don’t have the raw data. You just don’t know how true it is.

SAMI NAIM: But you do have data that's easily accessible, easy to understand and that really engages and directly communicates with New Yorkers, and that's our goal.

CHAIRPERSON BREWER: How do you get ideas for what data should be available from the

public? Is it just hits on NYC.gov? Do you have focus groups? How do get that so-called we'll respond to what the public wants?

SAMI NAIM: We solicit feedback from customers in many different ways. For example, the customer service survey earlier this year on 311 to get an example of where we're doing well and where we need to improve. Another example was 311 online, again, taking the top complaints, the top inquiries and trying to address them and provide more information in these areas.

CHAIRPERSON BREWER: Despite Commissioner Cosgrave's efforts, the community boards are still quite upset with how they get the 311 data. Just so you understand what the real customer wishes, not necessarily the person who's perusing nyc.gov, but people who really want to see where the mold is, where the parks are, where the hot spots are, where are things that are deeper than one might get just by doing so layering, which is important, but it's not deep enough. Is that the only way you get information from the public is through that one survey? Are

there any other ways?

SAMI NAIM: Again, we announced collaboration with Google as far as tracking online the traffic where it goes and what people are interested in, and what categories of information that maybe we should be providing more information on. We are constantly working to solicit feedback. All avenues are open.

CHAIRPERSON BREWER: It was interesting this morning at Public Democracy, there was a woman from Microsoft who is a demographer and she was pointing out that as we use the web, there are more digital divides dividing us even using the web. In other words, the issue of who uses one kind of social network versus another. I just throw that out again because I feel really strongly that the data in its total form, as long as we agree to what that is, needs to be up there much more than what you are describing. I don't think the public really knows the extent of government data. Unless they have it available to them, they cannot know how deep it is. I do believe that there will be a cost savings as opposed to as much as you say it

would cost to put it up.

SAMI NAIM: I think that there is a shared goal here in providing information to the public and have an engaged, informed citizenry. That's one of the initiatives with Big Apps, with 311 online, with 311 Twitter and with 311 itself, in addition to all the other initiatives I mentioned. I think what we're doing is trying to find the most efficient and effective way to engage and information New Yorkers directly to as many New Yorkers as possible.

CHAIRPERSON BREWER: What do you consider raw data? What do you think is a percentage of actual raw data up on the net now versus what is actually available if we were to do it in the more extensive form of Intro 991?

SAMI NAIM: That granular of a discussion of what is raw data versus some other form of data, I really couldn’t speak to it. I think that's more of a DoITT inquiry.

CHAIRPERSON BREWER: They're sitting at the conference right now. I understand that.

SAMI NAIM: They certainly wanted

to be here.

CHAIRPERSON BREWER: I know and I appreciate that. When you say easy to understand, it sounds more like processed than what I consider raw data. What do you mean by easy to understand? I'm delighted to have Council Member Oliver Koppell here also.

SAMI NAIM: The goal of all of these transparency initiatives it to engage in a meaningful conversation with the public. That means more than just hitting them with a bunch of bureaucratic language. We want to speak their language which is plain English. A great example is CPR. It's easy to understand, easy to see snapshot of our performance across many different agencies. Again, it recently won an award from the John F. Kennedy School of Government. These are really innovate projects.

CHAIRPERSON BREWER: I used to read those proposals. Go ahead.

SAMI NAIM: That's why we were singled out by the White House for being a leader in this field and for engaging in best practices that should be replicated elsewhere in the

country.

CHAIRPERSON BREWER: When you have a document internally, isn't most of that information already online and scanned? In other words, when you're doing the homeless figures, when you're doing the mold figures, when you're doing the inspection for the Department of Buildings, et cetera, most of that information is already online and able to be given to the public in a raw data format. Why do you think it would cost that huge amount that you stated in order to make it available online? Again, not in what you call easy to use and I call processed, I'm trying to find a way to have something that is both available to the public and absolutely clear that it comes in an unprocessed format.

SAMI NAIM: Again, agency records encompass a lot of different categories. They contain a lot of different information. You might have 1,000 records that say something that you could say in two lines. Our goal is to be as user friendly with the public as possible. We feel that we further the goals of transparency the more user friendly we can be, the more we can directly

engage New Yorkers and the easier we can make it for them to find the information that they need.

CHAIRPERSON BREWER: So how do you work with each agency? How does DoITT work with each agency to determine that agency's published internet record policy? How do you move data from raw as I call it to user friendly or processed? How does each agency determine what data is going to go upline and what is not? Is it this consumer friendly consumer request? If you ask the general public what government does, they have no clue.

SAMI NAIM: I would say the difference between a successful transparency initiative and one that lacks success is whether or not it's customer-driven or bureaucratic-driven. We don’t have bureaucratic-driven policies. We respond to the specific needs of New Yorkers.

CHAIRPERSON BREWER: I would differ with the Department of Education, sir. Please. I'm just saying that is not true. That's just not true. You can make an attempt to be customer-driven and not bureaucratic, but to say that you're not bureaucratic-driven is a stretch.

SAMI NAIM: I take exception to that. I mean I mentioned earlier Parents Link, which is a great tool that's revolutionary that allows a parent or guardian of any child attending public school to monitor their child's performance current and past. I think that's a great tool and I think they should be commended for it. They worked very hard to meet the public need and I think they do a pretty good job. I'm very happy to be a part of it in a small, small way.

CHAIRPERSON BREWER: That is still a really challenging IBM system. I could go on and on about that system. Do not get me started about that system. I would say that just leaving DOE aside for a moment because I think parents do understand more of what goes on at DOE than the general public understands about Consumer Affairs or EDC or some of these other agencies. Go back to my question, how do you decide what agency's records go up on the net? In other words, if you talk to the Department of Consumer Affairs, which to their credit does lots of different inspections, how do you decide what raw data or what data goes up? Who makes that decision?

SAMI NAIM: Again, it starts with the customer. We build our entire initiatives, all our applications around the customer to satisfy their needs, to adjust to their preferences, to speak their language and to make it easily accessible to them. That is the basis of all our initiatives. It's been a practice that's been well recognized throughout the country. I feel like we are a leader in this field and it's a best practice that we embrace moving forward.

CHAIRPERSON BREWER: Do you have a published technical standards manual for the publication of city government data on the net, or do you just say if somebody is interested in this topic we'll put it up?

SAMI NAIM: I think the general premise is that we want to serve our customers well. We want to serve the public. We have certain goals in mind when we do a transparency initiative. It's not how much paper can you put up on the internet, it's more how much can you engage New Yorkers, provide them with up to date accurate information that's quick, convenient and

easy to understand.

CHAIRPERSON BREWER: So there's no determination per agency based on some policy regarding what data will go up. Is that what you're saying to me?

SAMI NAIM: Well, I would say the starting point is meeting the customer's concerns and moving from that. I'm not sure if there is further analysis, but again it starts and ends with the customer.

CHAIRPERSON BREWER: All right. I don’t agree with that, but I'll let you stay with that if you wish. That is something that I absolutely do not agree. I think you need to have a technical policy. You need to have standards and you need to say that--

SAMI NAIM: [interposing] We're already taking ourselves away from the customer when we're doing--

CHAIRPERSON BREWER: [interposing] I think I know more about city government than most people, as does the other colleague here. Believe me; the customer on the street has no idea what city government does. They don’t know what

state government does. They don't know what federal government does. That's why I believe that we need to produce this data in a form that they can understand but it has to be much more comprehensive than what you're suggesting. The public does need their data. It's their tax money and they do need to have it now that we have the tools and the portals to be able to do that. It's a different world. Twenty years ago we couldn't do it.

SAMI NAIM: I think we agree on the common point that we want to provide the public with up to date accurate information to lead to an informed citizenry. But we want to do it in as efficient and effective manner as possible.

CHAIRPERSON BREWER: I think you sort of answered this, but has there been any effort to consolidate government data into one searchable web portal? Or is that what you consider nyc.gov? You're saying that you already do that?

SAMI NAIM: Well, nyc.gov is one of the largest municipal online portals in the country.

CHAIRPERSON BREWER: Well we're the biggest city.

SAMI NAIM: Over 70,000 visitors, over 2 million a year visit. It recently won a municipal government award as of last year.

CHAIRPERSON BREWER: It gets lots of prizes. I don't want to hear about all those prizes. Go ahead.

SAMI NAIM: I think it does a pretty good job. I mean as far as online portals go, it provides information, it is easy to use, it is user friendly and I think we're well served by nyc.gov and other initiatives such as City Map, such as NYC Stimulus Tracker, such as the Parents Link I mentioned earlier.

CHAIRPERSON BREWER: I know them all. Is any of that data currently published in a nonproprietary format that is usable by the public? An example of course would be XML. However, I also want to add one of the aspects is the City Planning Commission records. Those are definitely processed. That is not something that is in a raw format that the public would like to see to be able to manipulate on their own. Again,

you're not going to be changing the data, you're going to be manipulating in an open fashion so that the public can see what the original document is but maybe they want to look at it in a different format.

SAMI NAIM: I think we are heading in the same direction. Again, like the application NYC Big Apps would allow the public to take public data and incorporate it into their daily life for their own purposes. It would result in new ways to see old data. I think it's interesting because it shows not what we're interested, not what someone at an agency is interested in with the data but what the public is interested in with the data and their interests.

CHAIRPERSON BREWER: I think the issue there though is that it's still data that you have determined the public wants as opposed to the raw data. So maybe you're halfway there, but you just need to go a little bit further and talk about the raw data. The agencies have access to it. When agencies come and testify, they have access to it. All we're saying is that same data should be available to the public, again, with

certain restrictions.

SAMI NAIM: I think we're heading in the same direction. Again, with NYC Big Apps, it replicates best practices around the country, such as the Applications for Democracy. I believe you mentioned it in the opening statement. Again, this is a great project and we're really excited about it. It leverages our local creative community and sort of gives us a new interpretation of old data that we've had for a while.

CHAIRPERSON BREWER: Now, have you looked careful at how the data from Washington, D.C., San Francisco and Boston is being presented? It is in a format that people feel is raw data and is something that is going further than what you're proposing.

SAMI NAIM: Again, I'll talk about Big Apps. Application for Democracy sounds like what you're describing. It's giving data out there so the public can manipulate it and provide a fresh perspective on data that we've had for a while and really create new and innovative ways to look at information. That's been the model for

NYC Big Apps and the model for other transparency initiatives.

CHAIRPERSON BREWER: We're looking at something that's more sustainable and that is rawer than what you're suggesting. I understand that you think that what you're doing is sufficient. We're obviously stating that we need to do more.

SAMI NAIM: I mean I think we are always looking for new ways to engage the public and provide them with information and serve their needs. We're trying to do so, again, with the taxpayer in mind as well. We're trying to do it as effectively and efficiently as possible.

CHAIRPERSON BREWER: Either through RSS or some other format, do you tell New Yorkers when city government data has been updated? I know that we're certainly working to tell New Yorkers when there's a hurricane or a tornado or an oil spill or whatever, but how about data?

SAMI NAIM: There's a program called NYC Notify to let people know of the tornado disaster.

CHAIRPERSON BREWER: That I'm

familiar with. I'm talking about data, which is not as sexy as a tornado.

SAMI NAIM: Well I think DoITT would disagree.

CHAIRPERSON BREWER: You think that data is as sexy as a tornado? I don’t think so.

SAMI NAIM: Without comparing what's more sexy, again, I'm not very attuned to RSS and http and what have you.

CHAIRPERSON BREWER: Well just letting people know what's going on. In other words, how do you notify people where there is new information about the number of homeless, et cetera? People have to go now to what I call processed data in order to get that information. How do people know that there's a new census out for individuals who are homeless, just as an example, or new mold numbers or whatever?

SAMI NAIM: One example is 311 Twitter where through a mobile device you can get information from the city as it happens on maybe a health warning that's an outbreak or maybe like a heat warning and things like that. I mean we are constantly looking at new ways to engage the

public and provide them with information that they need. To the extent we can do that we will. We've proven that over the last seven years.

CHAIRPERSON BREWER: What I'm saying is there needs to be more notification when there is data that's available. I think that what Skyler did regarding the notification in terms of emergencies is terrific, but this is a different group. It's a different interest sector.

SAMI NAIM: Again, we roll out these initiatives so the public will take advantage of them. We don’t roll them out for the sake of rolling them out.

CHAIRPERSON BREWER: I understand that but I'm just saying data is something that as people get used to it they will actually learn more about government, feel more empowered, which is what we're trying to do to the citizenry and make sure that they are aware. It's hard to empower people. It's not something that any administration wants to do. I understand that. That's the whole purpose of data.

SAMI NAIM: Again, seven years ago we didn’t have many of these initiatives in place.

CHAIRPERSON BREWER: I know.

SAMI NAIM: We've done a lot to date and we'll do a lot more in the future.

CHAIRPERSON BREWER: We've been joined by Council Member Tish James from Brooklyn. Thank you very much. Council Member James, do you have any questions?

COUNCIL MEMBER JAMES: No.

CHAIRPERSON BREWER: I want to thank you very much. I think we're still going to be discussing this because we do not agree on the depth to which this data needs to be produced. I appreciate what you have done so far, but there needs to be much more depth in terms of what we are considering for the City of New York. I believe that the federal government does go further than what you suggested, despite what you heard today. I also believe that these other cities, be they progressive cities, are doing more than what we are and we need to do more. Thank you very much. First is Andrew Hoppin who is here from Albany, Ian Jacobs and Ben Kallos from Open Government. We're going to do panels of three and that's the first panel. Whomever would like to

start, go right ahead.

ANDREW HOPPIN: Thank you very much, Council Members and guests. I'm Andrew Hoppin. I'm the Chief Information Officer of the New York State Senate. I'm honored to have been invited to testify here today. I'm going to speak contemporaneously and I'll have prepared remarks that'll be turned in later today. I'm not really going to rehash the rationale for open data. I think Mayor Bloomberg this morning and Councilman Brewer have eloquently stated why it's so important. What I will say and what I will speak to is our experience in the Senate with why it's been important and how we think that may translate across different government entities. Everything we do out of my office relates to transparency, efficiency and participation in government. We've found that while people often of open data as being fundamental primarily for the purposes of government transparency, we've found it's as fundamental for government efficiency, even within our own enterprise and also for citizen participation in government. It's really the DNA for all of these things. So in that regard, the

first point I'd like to make is I would encourage the Council to think of other government entities as a customer here as well as the public. We often in the Senate need to make use of data as part of our policy analysis and policy making process that is not our data. And to the extent that other government entities at the city level, at the federal level, other state level entities publish data openly in an easy to access and easy to use format that assists us with our work. Please consider other government entities as a customer. That can enhance our efficiency as a body of government entities working together. Also, when you think about openness in data, I would encourage you to think also about openness in terms of the software code that may be written to make use of this data. Open data can be the tip of the iceberg really in terms of the work required to actually make good use of that data. So to the extent that tax dollars are invested in doing work to make data useful, I would encourage the Council to consider how to make the results of that work on software code as open as the data that goes into those applications. Then finally

openness in terms of communication; I think the awareness of the availability of data, the user stories about how data is successfully used and the ability to learn from one another across different levels of government really all comes still through human beings communicating with each other, whether online or offline. So all of the roadmaps, the plans, the deliberations, the successes and the failures that relate to the opening up of government data I hope will be thought of as being fundamental and open public records as well so that we can all learn from each other. So openness of code, communication and data is my second point. The third is that in the context of the Senate, I'd like to speak a little bit to the debate that we just heard between the offering of applications that are of immediate and evident utility to citizens versus publishing a lot of raw data online. At least in our context we haven’t seen there to be any conflict between those two. One example is our effort to publish legislation that we're working on. Online we have both a search interface which is quite refined and easy for citizens to use we think to search for

bill information. Also, on open.nysenate.gov, we publish the raw feeds of that data in an API so that you can actually write your own applications to build the sorts of search interfaces that any individual stakeholder might want for their own particular use case. Because we don't believe that we can do a sufficiently good job of meeting every use case for every subset of our constituency sufficiently, so we want to push the power to do that out to as many people as possible while still trying to do the baseline good job that we need to do for the most widespread use cases that we encounter. So that's my third point that I don't think there is a conflict between raw publishing of data and delivering highly refined value added services to citizens. The fourth point I would make is that I think citizen participation is a key part of this. It can actually help to bridge the gap between viability and the desire to publish everything, which is quite expensive. We have a lot of human beings in this state who have time and talent on their hands and there have been some good examples I think at the federal level of how you can use crowd

sourcing applications to actually empower citizens to add value to raw data and make it more refined and therefore a more useful data product. In the U.K. we saw an example recently of "The Guardian" newspaper taking literally scanned records of expenses incurred my members of parliament and going through and looking at those expenses and actually processing them into structured data. Taking unstructured data and turning it into structured digital data. While that particular case may or may not be of interest at the city level, I think we have seen time and again that if you've got something that's too labor intensive for government to do, there are now ways that we can employ citizens being of public service themselves in order to actually work with data and add value to it and republish it through government. So citizen participation and crowd sourcing is my fourth point. My fifth point would be that obviously there are legitimate points of debate here. I find myself conflicted on some of the issues that were debated here over the last half hour. I think it's key in terms of implementing this is coming up with good

legislation, if indeed it's implemented and however open data in the City of New York goes forward is that the process around doing it remain open. It's fantastic to have this hearing. I hope that the views of the people, the views of the city administration and agencies and the views of other government entities are represented consistently throughout this process, not just in the consider of it but through the course of opening up data. There's a timeline all the way through, I think it was 2012 or 2013, which agencies would be potentially opening up their data. Throughout that process I think it's going to be important to have all those voices represented as agencies figure out how to open up their data, publish their roadmaps about how to do it, test what works, learn from each other about what works and doesn’t and we in the Senate would very much hope to be invited to be a part of that ongoing debate and in turn to invite you to do the same as we look to open up the subset of state data which we work with directly. To that end, I do hope that there is a role for not only DoITT, but an entity within the city that represents the

voice of the people as you go forward into this. Thank you very much for inviting me here today.

CHAIRPERSON BREWER: I want to thank you in particular because there is a lot going on in Albany and I appreciate your being here because I think you're the voice of reason for the State Senate. Thank you.

ANDREW HOPPIN: No comment.

BENJAMIN KALLOS: Good morning, good government, transparency and community advocates, Council Members Lappin, Gonzalez, James, Liu, and de Blasio, thank you for recognizing the importance of this issue and for sponsoring this introduction. Council Member Brewer, thank you for chairing this committee, for being the first to propose this kind of ground breaking legislation for today's advocacy, and most of all for being an amazing legislator and role model for so many in the city. My name is Benjamin Kallos, I am here before you today as co-founder of the Open Government Foundation, a New York State not-for-profit which aims to bring greater transparency, accountability and openness to government by making information available

online for all to see. Government is like anyone of us, because it is comprised of so many of us, and it is subject to the same sicknesses and disease. Unlike us, we know the cure for most of the government's worst ailments and maladies. If we may be the first of many to quote him today, Justice Louis Brandeis famously wrote, "Publicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants, the electric light the most efficient policeman. Transparency and openness in government should be likened to medicine or a vaccine." Like Buckley's, a medicine taken to cure a recurring ailment like a cough, "It tastes awful and it works." Like a vaccine, dreading the shot is often worse than the shot itself. We have all acknowledged the value of vaccines, requiring measles, mumps, and rubella, amongst others, for all children who attend public schools. We must make sure our elected and appointed officials get a similar vaccine against corruption. While transparency might cause momentary pain, as when member item information was recently released and scandal uncovered, new reforms were created to

avoid future corruption, leaving both City Council Members and New York City healthier in the long run. I'm going to skip the next piece of the testimony.

CHAIRPERSON BREWER: That would be great. Go to some of your suggestions. That would be great.

BENJAMIN KALLOS: If I may, I'd like to just go over our experience as a not-for-profit with FOIL.

CHAIRPERSON BREWER: Sure.

BENJAMIN KALLOS: Earlier this year, our Foundation sent a Freedom of Information of Law request to the New York State Assembly and Senate as well as the New York City Council. FOIL is the only means for good government groups and reporters, let alone citizens, to gain access to most government information. It is worth mentioning that most of this information must be made available within 5 to 20 days and paper records often have a statutory cost of 25 cents per page. In our experience, the New York State Assembly followed the FOIL to the letter. They quickly responded within 5 days statutory time

Formatted: Font: 13 pt

line. They quickly delivered an electronic copy of every single bill and vote since 1995, which we promptly posted online at NewYork.OpenLegislation.org. The New York State Senate has since delivered committee voting information for 2007 and 2008, which was also promptly posted on the same site. With a small addition; you made a point regarding RSS feeds and alerting people as to what type of information was happening. Our foundation has already created Data Gov Tweets, which is a Twitter account so that people who are interested in what's being posted at data.gov can get it on their Twitter feed. We're also created RSS feeds for various types of legislative bill actions so that you can get a feed in XML showing when a bill that you're following or certain bills have made it out of committee or are being voted on. Although both legislative bodies have been cooperative, many State Legislative records remain in paper form. For example, the New York State Assembly's committee votes for 2008 number 5,356 pages and at the previously mentioned statutory 25 cents per page will cost $1,339. While the State

Legislature has been compliant, the New York City Council has requested 90 just to respond to the same request and our appeal, detailing how the City Council could and should comply with the FOIL was recently denied. There are many instances where government bodies have demonstrated bad faith in compliance, requiring litigation to release public information, costing government and those exercising their rights hundreds of thousands. While some critics of open government or this legislation might argue that freedom of information and open meetings law provides for sufficient access, we would simply point to our current interaction with this very institution where transparency, accountability and openness remain blocked while we wait three months for a response that the law dictates, and other legislative bodies have demonstrated, should take no more than 5 to 20 days. In fact, in our foundation's efforts to set legislation free we've learned that government has incentivized itself to keep information secret. The government generates information that has value by virtue of the fact that it affects constituents. The government then

pays a vendor to help it internally manage that information. At the same time as that vendor is getting paid by the government, the vendor becomes the only source of that data in manageable form. The vendor then licenses access to the public who need it so badly that they are willing to pay for it. We have been advised that should the State Legislature begin using our website or should they implement our free source code, it would save millions a year in management fees and eliminate costs in the tens of millions currently paid by the public to a vendor to gain access to the same information that should be made free by the government rather than a not-for-profit. In an open, free democracy it shouldn't be up to the government to decide what information is important enough to be released to the citizens for free. Instead, all non-private data should be released in its entirety, and citizens should be empowered to decide what information they will consume, when they will consume it, how, where, and why. Our mission is simple: wherever government neglects to release critical data to the public, whether through apathy, indifference, or impracticability,

we will step in and use every means available to release the information for free ourselves. When Council Member Brewer sent out a call for testimony that was rightfully distributed widely throughout the open government internet communities, asking the Internet community to help "highlight the advantage to web developers if there was city data available in an open data format", given our recent experience, having developed a website for state data and wishing to develop a website for city data, our foundation heard this call loudly and clearly.

CHAIRPERSON BREWER: Ben, can you just summarize? You're doing great. Can you summarize the rest of this?

BENJAMIN KALLOS: Sure.

The suggestions are next and I'll wrap up quickly. Data must be available over the internet in accordance with the open principles of open data to facilitate development because in its absence three problems virtually preclude any developers from building websites or applications to benefit New York City. First, if the data is not available over the internet, developers can't

aggregate it, because it isn't there to collect. If the data is not freely available over the internet, it requires burdensome, technical, lengthy, and expensive FOIL requests, discussed earlier. Second, if the data is not in an open data format, developers won't be able to use it to create websites or applications. Again, developers are in a technical, specialized fields, and they like presenting technical information to a user in simple and useful way. Many developers don't like to parse through flat files that result from FOIL requests. This poorly formatted data only presents yet another obstacle to making New York City's data useful. Our foundation can testify that data obtained through a FOIL request, must be parsed into a format that can be easily manipulated by a website or application, which is time consuming, difficult, and requires an esoteric skill set that is uncommon even in the most dedicated developer. Third, without strong legislation, government agencies are extremely unlikely to do what this introduction would require of their own accord. Our experience has shown that government is slow to take initiative

and sometimes seeks to avoid compliance with the existing FOIL law. Once your introduction is passed into local law it will encounter similar resistance in compliance like FOIL. Our foundation recommends adding a self-enforcing provision aside from litigation to provide remedies to citizens, good government groups and journalists for failed compliance, which would state that city agencies may not charge FOIL statutory fee for any records produced in exclusively physical paper form, after July 4, 2010. Without this legislation and strong incentive for compliance, New York City will not be leading our nation with open data standards, let alone keeping up with national trends. Thank you again for considering the adoption of open data standards. We look forward to working with the New York City Council to make open data a reality in both the short and long term.

CHAIRPERSON BREWER: Thank you very much. We'll certainly talk to the general counsel about the lack of FOIL responses, or we'll check on that.

IAN JACOBS: Hello.

CHAIRPERSON BREWER: Hello.

IAN JACOBS: My name is Ian Jacobs. It's a great pleasure to be invited to speak today. It's a little bit by accident. I was here for other reasons. I was invited to the hearing and it turns out the organization I represent is mentioned by name in the proposal, which is the World Wide Web Consortium. I'm the head of communications for the W3C. I'm thrilled to hear about the proposal. Just a quick word on W3C, we were founded in 1994 by Tim Burners-Lee, who is the inventor of the World Wide Web. Data is a great passion of his, especially these days. Our mission is to create standards that are open and available for free to make the Web available to all, anywhere, on any device. So some of the standards you may recognize include XML and HTML. Right now we have a new develop stack of standards for data called the Semantic Web. Some of them you may not have heard of yet like RDF and SKOS and OWL. But we have been developing these standards for open data for ten years. I think 1998 was the first standard, RDF. We've been developing communities to help make the best use

of these open standards, including for example the health care and life sciences activity. More recently, the e-government activity has really caught on fire. So we have an active interest group where people can participate and learn and contribute to best practices for putting data on the Web. So it seems to speak exactly to the problem you mean to address in the proposal. I'm happy to provide other information about how we operate in our public accountability process ourselves and anything you'd like to know about the standards. I have just a couple of words about the proposal which I just read upon arriving. I think it's, of course, the perfect answer to use these open standards for data longevity and for cost reasons and to build one web. I heard a cry for raw data now, which Tim Burners-Lee had the audience chanting at the lecture earlier this year out in California. As far as the cost of doing this, it seems like a good way to start is to put up the data that's already available electronically using these open formats. Again, Tim Berners-Lee recently published a document in response to the request

from the Obama administration about how to put data on the web, so that's available as of last week. I can share that URL with some people if they're interested. One thing that raised an orange flag for me was the requirement for a central portal. Points of centralization tend to cause blockages and it would seem that it would both lower costs and make life easier for people if they could sort of publish their data independently and then it could be aggregated by any number of sites. I heard the chair speak in favor of RSS feeds and that sort of thing. So I think aggregation is a good way to help reduce the costs and to sort of free up each agency to publish what it has at the speed that it's capable of doing and to establish collaboratively good practices for doing so. Again, W3C is very, very interested in the participation of governments internationally, but we recognize that there are needs at different levels. I appreciate the comment about interagency and intra-agency needs. We have an open forum and welcome participation and help in building these good practices. I'm

happy to answer any other questions you might have. Thank you again.

CHAIRPERSON BREWER: Thank you all very much. We did mention in our legislation that there would be more difficult data and easier to put up data. So it's sort of along the same lines that you just described. We would ask that the data that's available go up. So I think we're thinking along the same lines. There are some legacy systems that might take longer. When you have talked to other cities or the federal government or maybe other countries, particularly in your role, have you found that some of the roadblocks that you heard discussed today have appeared? Or are they able to get this data up in the format that our bill outlines?

IAN JACOBS: I have not participated actively in those discussions. I'd be happy to have one of my colleagues in the interest group talk to you about that. The one comment I did hear was that some of the state governments were nervous about things like RSS feeds. They were not as up to date in technology as maybe the federal government was. I have no

way to back up that statement, but that was one comment that I heard. I also know that people have invested a lot in XML and may be wondering what the difference is and why they should go to RDF. Without belaboring the technical details, I think RDF it turns out is easier to merge with other data. For technical reasons it's just much easier to put it together and create mash-ups, which I heard was one of the big goals. Getting from XML to RDF is quite simple. That was one of the comments that I heard.

CHAIRPERSON BREWER: In Albany, do you find that there are other state legislatures that are trying to do what you're doing? Is there any kind of discussion going on around the country as to what different legislatures are doing?

ANDREW HOPPIN: We have asked and we have begun to form a de facto community of practice with our peers in other states. We have not found formal processes and specifications that we hoped would have existed that would make our jobs easier. We went to the National Council on State Legislature to talk about taxonomy for legislative data so that you could look up a bill

in one state and know that you're also able to the find the analogous bill from another state at the same time. That has not been done to our knowledge. We're hoping to set a standard, again, by doing it in an open way all the way along, publishing our roadmap, inviting other people to critique and add to it that we'll set a standard that won't be just good for us but good for other states. We're actually putting up a wiki that will help other states to publish their analogous experiences alongside ours so we can all learn from each other. I think at a city level there's a direct analog as well in terms of if you figure something out, help other cities to leverage off that good work and vice versa.

CHAIRPERSON BREWER: The National League of Cities would be the place to go. Back to the issue of the portal, I think we have 80 city agencies here. That’s a lot. I think that when you talked about centralized versus individual, there has been a big push due to the constraints of budget to try to do more under the umbrella of DoITT, which is our technology agency. I'm just saying that I understand the

centralization argument can be a challenge if there are barriers and slowness that result. But I do think that agencies need to be somewhat collaborative if there is a resource. How do you answer that question?

IAN JACOBS: I think W3C has a lot of experience in these collaborative processes and so I think giving them a forum for doing so should be low cost and everyone would benefit. Aggregating the information that they publish may be of service to the citizens that can go to one place and find what they need. But it should probably not be the only way for them to reach that. So it can be a convenience but without sort of hampering the agencies from putting their data online.

CHAIRPERSON BREWER: Are you working then with other municipalities through the National League of Cities or the U.S. Conference of Mayors or National Association of Legislatures or anything like that to try to work with best practices? Is that something that has come up as you talk about e-government?

IAN JACOBS: We're revising the charter of this interest group as we speak and the emphasis on creating a forum for this to take place is already built into the draft charter. I have not myself chatted with these agencies. I'm taking notes so that I can go back and find out if we have. It's certainly our goal to produce sort of best practices.

CHAIRPERSON BREWER: Thank you very much. This is fascinating and the conversation will continue. Thank you very much. The next panel is Philip Ashlock from the Open Planning Project, James Vasile from Software Freedom Law Center and Morgan Reed from Association for Competitive Technology. Whatever people send in, we will incorporate as part of the record. Needless to say, there will be much discussion back and forth. Whoever would like to start please proceed.

MORGAN REED: Good afternoon. My name is Morgan Reed with the Association for Competitive Technology. Thank you, Chairwoman Brewer, for having this hearing on Intro 991. A quick explanation, ACT, the association is an

education and advocacy group focused on small technology-based businesses. We represent over 3,000 software developers, systems integrators, IT consulting firms, and e-businesses and we generally advocate for a healthy tech environment that promotes innovation, competition and investment. We are very active in supporting the goals presented in Intro 991 and we look forward to working with you and others to make it a reality. I'll vary a little from my written testimony to say that there seems to be violent agreement amongst all of us of the need for open data. We of course are big champions of the concept of raw data as developers. We've written a paper recently where we discussed what we think are the three principles of what we call we-government and that's like access, participation and accountability. I thought I'd move away from the agreement area and talk about something a little bit different, which is focusing on the developers. We've talked about this in platitudes and what the government should do, but I thought it was interesting that the gentleman from the Senate, he said that the State of New York has a

lot of people with time on their hands and talent. I'd like to make sure that we find ways to make sure that those people don't have time on their hands and that they have jobs. So what I'm interested in, in pursuing the open data element, is to give them an opportunity to build applications that actually result in revenue in their pockets that are ad-supported or they can sell subscriptions to. They can bundle with other proprietary information from the financial industry and build applications that actually move us in a positive economic direction. I thought it was interesting that one of the other areas that the gentleman from the Senate touched on was this concept, and he's exactly right, that there are more than just one set of customers out there. We've heard a lot of people speaking about citizens as customers. My membership actually has real customers, the folks who actually pay the bills. So when I look at opportunities from the open data element, I see opportunities for our membership to sell services into other government agencies. That we can actually go into a government agency and not have to have 10 meetings

with 14 roundtables that bring in 2 departments, but that we can go into a Department of Education and say I'll build you an application using data from another department and I will merge that independent of an interagency working group, an interagency funding process. I'll make it happen with your own revenue stream and give you the information you need for internal customer purposes. So I think there is a whole element of this, we want to serve customers of the broader citizenry, but I think there is an amazing opportunity to add efficiencies that gets us out of some of these elements of siloing. We all know there is this hoarding mentality that happens. We all know that getting two agencies to work together is difficult. I think open data presents a huge opportunity for my membership to sell services to build platforms independent of interagency working groups. I think that's a phenomenal opportunity for the state. I also think that one of the areas that we should focus on a little bit is industry best practices that we're seeing today. Where are we really seeing the rubber hit the road? This has been touched on

a little bit, but we all know that Chinese saying of teaching a man to fish and he eats forever. I think this debate about open government and your focus on raw data can really be broken down in that. When the government hands us a flat file of information and they throw it over the transom, it's just one fish. What we really need to do is start looking at systemically having government agencies conceive of their data in three key characteristics. It needs to have open data services built in from the ground up. It needs to be open for application services as well. In other words, I should be able to tie into the data stream, respecting security and paying attention to privacy, and immediately have access to data in real time. And then lastly, it needs to be technology neutral. One of the elements that have been phenomenal for us is the Kindle. I don’t know if the chair has a Kindle but you've heard of a Kindle I'm sure. The Kindle technology is actually something based on a format developed by a company called Mobipocket, which was an ACT member. We all love open standards and we love standards activity and we love the process so to

speak that takes place going through standards activities. But really the cutting edge of innovation often happens apart from standards. It's from the guys who don't have the money to fly to Geneva to sit down in a meeting with 14 lawyers. It's a company like Mobipocket that says we know how to deliver a format on a mobile device this size or this size or this size and we're going to put it in the marketplace and we're going to see if the real customers, the citizens find it acceptable. So I want to make sure that as government develops technology and puts it out into the marketplace that it does so in a technology neutral way that allows us to build on top of it. That may mean that we're going to take it in a direction no one's ever thought of. We can also look at what the big players are doing.

CHAIRPERSON BREWER: Can you just summarize a little bit?

MORGAN REED: Yeah. We should take a quick look at what the big players are doing and I think there's some guidance there. The large players, such as Google we know, Amazon has their web services and Microsoft has what's called their

Open Government Data Initiative or OGDI. I'm going to take a moment to touch on OGDI because it's been one that we've heard a lot of developers talk about. This is a product from Microsoft, of all things, and it is platform neutral and open source. It's actually an operating system that Microsoft has made available so that you can store data there and take it out in real time. Yet, my members who don't use Microsoft products can access the technology on it. It's available to be reached out to through Adobe flash, JavaScript, PHP, you name it and I can get access to it. So it's interesting to see that even the big players, even the big proprietary players are seeing opportunities to provide data in this format. So ultimately I think we know that from the developer's perspective we agree it's got to be raw. We want to see it tech neutral and we want to see it built from the ground up with the concept that I'm going to provide this in real time and look for ways to make it interagency operable. I would say that all of our recommendations fit very much in line with what the bill has in place. Thank you very much.

JAMES VASILE: Good afternoon, Chairperson Brewer. Thank you for this opportunity to speak today. My name is James Vasile and I am legal counsel for the Software Freedom Law Center. The center is based here in New York and we're a nonpartisan nonprofit legal services organization whose mission is to provide pro bono legal representation to those who make technologies that promote free communication and innovation. This bill makes government data available to New Yorkers in two different ways. The first is the web portal and we've all talked about that, but the most important thing is the raw data. Everyone has sort of been focused on that and I am going to be focused on it as well. The exciting thing about giving New Yorkers raw data is not just about the city giving up its data and the city being transparent and the city sharing things with people, it's about what we can do with that data once we get our hands on it. The amazing thing is that we have no idea what people are going to do with the raw data once we give it to them. We have no idea what their needs are and certainly as much of an idea as they have

themselves. There are so many different ways we can use this data that we haven't even thought of yet, that we will only discover once we actually hand it to people and see what they come up with. The only way to do that is by giving them the raw data. I'd like to talk a little bit about the kinds of things people might do just to show sort of the exciting possibility that exists there and then talk a little bit about who they are and why they're going to do these things. If we give people this data, they can combine data from Consumer Affairs, the Police Department, the Health Department, Sanitation and take all that government data and combine it with non-government data that's available from other sources on the internet, including proprietary sources of data. We're going to start getting services that allow us to answer questions like where is the nearest place I can park without having to wake up for alternate side tomorrow? That might sound like a trivial thing, but that's a huge deal in the lives of a lot of New Yorkers. I mean the number of people who have to plan their lives around Mondays and Thursday mornings so they can be near their

car is quite large. You can answer questions like, which restaurant in my neighborhood has both great reviews and a good health record. That's something you can't do just from government data. Ideally, you want to do it right in the place where you're already looking for your restaurant reviews. You don’t want to have to go to a government web portal to find out this information. You want that information already out there, already in the service you're already using at Yelp.com or Yahoo or Sidewalk or whatever place you're going to find a restaurant review. You want that information already there. You don’t want to have to take a second break from what you're doing to try to find the data somewhere else. The other thing this can do in addition to just sort of daily lifestyle stuff is help businesses plan their mode of operation in the city. If you could answer the question where should I put my electronics store so that it has the least competition and the most foot traffic?

CHAIRPERSON BREWER: That's the analogy I like the best. Go ahead.

JAMES VASILE: That question of where I should put my business and how I should structure my business to get the most out of city data and city life is really important. You're not going to get that just from the city. Someone at a city desk is not going to magically wake up one day and say I think the public needs to know where there is a lot of foot traffic and little competition for electronics stores. The fact is, that's a very specific kind of demand and it's the kind of demand that people have for themselves. They wake up one day and realize they need it and they need a way to answer that question. The city is never going to provide people with a way to answer that question. If the city creates a web portal, it is not going to go that deep. It is not going to get that niche that people will be able to answer that specific a question. But that specific question is exactly what people need answered and that's why we need the raw data. The last example I'd like to give is somebody either looking for an apartment or looking to sell an apartment. It would be really useful to find out average prices of apartments similar to the one

you're looking for within a certain block radius or something like that. That's an extremely specialized application. We can sort of imagine the city doing that at some point. But we could also imagine the real estate market doing that for the city and for consumers at no cost to the public, which I think would be a better win for the city in terms of how it spends its dollars. The reason to put all these things online is that we can't predict what people are going to do with it. And even if we could predict what people were going to do with it, we would never get down into the weeds, into the details to satisfy everybody's smaller needs. I'd like to give just three examples of how the public uses this information currently. These are three really good examples of how volunteers have stepped up to organize government data and present it to other people. There is a web service called Edgar. Edgar is a database of SEC filings. What some people don’t know is that Edgar, which is tremendously useful to the financial and legal services organizations, especially in New York, is that Edgar was started by a volunteer. There was a guy named Carl

Malamud who just started putting SEC filings online. He decided that this was information the public should have and he was going to share it with them. Nobody knew at the time how crucial this information was. Nobody knew how useful it was and how vital to the daily business lives of so many companies that this information would be. He ran it for two years just for free because this is what he wanted to do. He kept saying to the SEC that I'm doing your job for you. I'm giving people this information that you should be giving them. The SEC said thank you, keep on doing it. He said you should be doing this. At some point, he decided that he had had enough, that he couldn’t just keep on running this service. So he put a notice out to all his users and said he was turning it off. If you want this service to continue, call the SEC. So many people sprang up and called the SEC and said they needed the information and they used it every day that ten years later this website is still running. People are still using this information. It turns out that people need this information. Other people want to give it to them but all they need is a

little bit of help from the government to give them the data and enable them to take that first step. The second example is an example of something that failed. I have a colleague who had trouble finding schedules for bus service to his Windsor Terrace neighborhood. He tried to write a computer program to grab the data from a city website and send it to people's cell phones on demand. This is a service that I think a lot of people would use because I don't actually know the bus schedule in many neighborhoods and when I do arrive I have no idea whether I'm going to be able to get a bus or not. He was able to find bus service information online, but he was only able to find it in PDF format. He couldn't find it in a standard format that was capable of being processed automatically. So he couldn't just have a computer program grab the bus schedule every day and send it to whoever wanted it on demand. If he had been able to do that, every single one of us would be able to access that data right now from our cell phones whenever we wanted it, even if we don't know we need it in advance, we could have found it and used it. That's a really powerful

example to me of the difference between that service existing and that service not existing is the availability of the raw data in a format that let's it be automatically processed. The last thing that I'd like to talk about in terms of an example is a website called ridethecity.com. There were two people who met as students of urban planning at NYU. They formed this website called Ride the City. You enter your address. You enter a destination address. It tells you the best way to get from Point A to Point B that is safest for bicycles. The way it does this is by finding streets that have bicycle routes on them, which is really easy information to get. But what would make this website be even more useful and more accurate is if they could take into account bicycle accident statistics so they could actually tell you which streets really are safer, not just which streets might be safer because they have bicycle routes. These are three really good examples of how we take government data and use it in new and innovative ways that frankly I don’t see the city coming up with any of these services anytime soon even if we did get a web portal to

allow us to do it. I think the last thing I would like to address is the issue of cost. We had a gentleman from the city earlier. He testified that reducing call center volume by 10% would save the city some 4 million. I presume he was talking about 311 call center volume. Now, 50% of 311 calls are simple requests for information answered from a database of 6,000 entries, 6,000 little facts that most people call about. If this bill were to cause those 6,000 entries to be released, we can turn every computer in New York into a 311 service center and save the city the cost of a large percentage of 311 calls. The other thing I'd like to address is he kept mentioning NYC Big Apps, which sounds like a great program. I just wanted to say that in order to implement NYC Big Apps they must be aggregating quite a lot of data. What I'd like to see is that data, because they've already decided that is worth the cost of aggregating. So every piece of data they're putting into NYC Big Apps we should have. Frankly, if they gave us the data, we would build NYC Big Apps for them. We'll build what we need and we won't have to wait for the city to detect

and fill our needs. NYC Big Apps sounds like a great way to help New Yorkers, but if you give us the data we can help ourselves. I would suggest that we get the data and we get the software code behind NYC Big Apps and we will see exactly what people can do once you give them the ability to help themselves. Thank you, Councilwoman Brewer for your time.

CHAIRPERSON BREWER: Thank you very much.

PHILIP ASHLOCK: Hi, my name is Philip Ashlock. I'm from a nonprofit organization here in the city called the Open Planning Project. Councilwoman Brewer, I'd like to thank you for inviting me here and for proposing this legislation. I'm going to keep my comments short because most of my points have already been made by others and it's running late in the afternoon and I would like to get back to the Personal Democracy Forum. The Open Planning Project is a nonprofit that builds open source tools to facilitate civic participation and improve civic life in general. We have also tried to make a point of finding models in other cities that have

been successful for different technology or livability issues and try and propose them in places like New York. One thing that we've been doing recently is looking at all the different instances of 311 services in other cities and by private organizations and companies and have been creating what this legislation would refer to as a volunteer consensus body around open 311, it's open311.org, which we just set up. One point I would like to make about the use of the word open since it's so vague. There's really sort of two parts to openness. There is transparency and there is participation. In developer's terms that's read/write. What's really interesting about the Big Apps project that's been proposed is that it's modeled on the Apps for Democracy contest that's been held in Washington, D.C. There have actually been two of these Apps for Democracy contests. The first one was held last year and it was solely a read-only model where it was using existing data to sort of prove the importance of transparency. They are currently holding another one which ends in two days, based around Washington, D.C.'s open 311 API. That's a

read/write API, so it's transparent and it's participatory which means that citizens not only can see the information that's been put into that system, but they can also add to that system on their own. That's one point I would like to make about the meaning of openness. Otherwise, I think I'm just going to run through some brief examples of how we've encountered open data in city agencies. I was actually asked to come here based on some experience at our organization and experiences some of my colleagues have had outside the organization and within the organization dealing with the MTA. Now, I understand that the MTA is not necessarily under the purview of the city but I think that the experience was representative of other agencies. Just to give you some background of how MTA's scheduling data is made available, as was just brought up the usefulness of that. Currently there are only two places that I know of where bus schedule information can be received in an electronic web-based application and that is MTA.info and Google Maps. The way Google Maps gets this information and the same way anyone else would get this

information is by contacting MTA and they would provide a compact disc of the data. The reason they provide a compact disc is it helps them track the information and which organization or private entity has received that data. I don’t completely understand the reason they need to track that data, but that is why they do that. Because it's on a compact disc and not in an automatic delivery system that request has to be remade any time each schedule change has been made. Consequently, whenever the MTA changes a schedule, Google has to submit another request for a compact disc. Then they receive that compact disc and they have to translate it into another form. If you look for schedule information on Google Maps it usually takes a week or two for it to come back up to date. That is one good argument for providing it an automatic structured format. Another is just a very simple example of accessing 311 data. It is actually possible in an automated way to get 311 data that's been submitted to the system, except for the fact that they put a firewall to accessing that data through aCaptcha which is a typical way of preventing spam from getting in a forum. Do

you know what aCaptcha is? Now, aCaptcha is typically only used to prevent spam from malicious data being written onto a server. They're using aCaptcha specifically to prevent you from reading the data in an automated way. There is nothing malicious you can do. They're allowing humans to read it, just not a machine. Another piece of information that I actually don't have very many detail about is the NYP used to provide crash stat data and I could be wrong about this but I've been told that they no longer provide that. So these are three cases where there is data that is available in an automated electronic way but are intentional means of preventing that from being available to the public. Just to keep it short, I will leave it at that.

CHAIRPERSON BREWER: That was three fantastic examples and three fantastic testimonies. I appreciate it because it will help us as we try to negotiate this bill. It will help us explain to the public because I think people don't understand. I think what bothered me the most with the administration's testimony is thinking that the public knows what they want.

The public can only be empowered if they have the data to use in a way that you all described. There is so little understanding of government unless you really need government and then you do find it helpful hopefully. But it would be great for us to empower people to use government in a different way as you described. I think that's what the glory of the internet is and that's what we should be using. Are there any specific comments you have on our bill that you would change in any way? That would be the only other question I have.

PHILIP ASHLOCK: I have one which is in response to the representatives of the administration who were commenting on the 3 billion pages of archived documents that would need to be scanned. It does take a lot of time to scan documents, but there are a plethora of institutions, not just in the New York area but all over the country that would love to help contribute to that effort if there was something from the city to make sure there was nothing sensitive that was being digitized. Archive.org is a perfect example of that and they have a very

sophisticated, refined way of scanning documents in a very quick way. I would suggest that if it would help alleviate those concerns and it take past the 2013 deadline to convert those 6 billion pages, that if each agency was continuously in the process of achieving that goal past the deadline that there would not be a penalty. If they were being provided with every means to help achieve that goal, then I think that that should fit under the compliance.

CHAIRPERSON BREWER: I mean certainly we would play with the deadlines. We were really clear that issues that are easy to deal with, low hanging fruit go now and then work on those that are more complicated. We were really clear in the legislation. Does anyone have other comments about the bill?

JAMES VASILE: The only other thing that I would liked to have seen in the bill would have been a clear statement as to the lack of assertion of any copyright over the data that comes out of this process. Right now in San Francisco there is a bit of an argument over who has the right to publish Muni arrival and

departure times. A company was hired to process some data for the city and provide the public with this information. They're quite upset that the public is now using that information and publishing it and chopping it up and managing it for themselves. What I would like to see is an avoidance of that argument here in New York, to make it really clear that the city is neither asserting any control over this data or going to permit anyone else to assert control over this data.

CHAIRPERSON BREWER: The issue is, of course, the public will see the original document because that's always there. So you can always see what that was and if it's changed, the public can see the original document. To me that would be pretty clear. Thank you very much for your testimony and enjoy the rest of Andrew Rasiej's Conference. The next panel is Chris Keeley from Common Cause, DeNora Getachew and Rachael Fauss from Citizen's Union and Joshua Breitbart from People's Production House. Whoever would like to go first can go right ahead.

CHRIS KEELEY: Good afternoon, Chair Brewer and members of the Committee on Technology in Government. Thank you for the opportunity to speak here today. My name is Chris Kelley. I'm an associate director with Common Cause New York. Common Cause is a government watchdog organization, nonprofit citizens' lobby and leading force in the battle for honest and accountable government. Again, thank you for this opportunity to speak here today. We see this in a lot of ways as a very, very strong improvement of the already groundbreaking law that the chair sponsored a few years ago, Local Law 11. This is really taking that up another notch in a very important way. That bill's stated intention was to position New York City as leading the nation in using information technologies to improve the efficiency and accessibility of municipal government, and using the internet as a powerful means of accomplishing these twin goals. Clearly, Intro 991 fits into that framework. Common Cause testified on Local Law 11 in February and we'd like to revisit a few of those recommendations we offered then. The first of which is that we

suggested that Local Law be amended to include data quality standards. At this point we see that the current Intro before us requiring DoITT to issue a record policy and technical standards no later than January 4, 2010. Centralizing these standards under DoITT as opposed to requiring greater flexibilities with different agencies to do it we think is a valuable step. Having standards across the board is a very important way for those sorts of information to be compatible with one another because there is going to be information from HPD and DCA that are both going to want to interrelate with one another and having those standards across the board are very valuable. Also in February we urged the Council to mandate that Local Law 11 establish administrative mechanisms allowing individuals to seek and obtain correction of information disseminated under that law. It doesn't seem that it's in this bill and that's something that we would encourage DoITT to include in the policy standards. That a way for individual New Yorkers as they're using this information, if they notice inconsistencies, if they notice problems with it,

having a defined mechanism for them to maintain that dialogue with the city so as to not lose track of that real relationship that is going to be developed through this sort of proposal. Also in February, we urged the Council to require all agencies to report annually on the number and nature of complaints received. Similar to the last recommendation, we think that that's something that should be the standards put forth by DoITT that in addition to having that back and forth dialogue we need to make sure that that is available for everyone to understand what sort of shortcomings people are experiencing so that we can have a full public discussion about how to address them. In addition to those three that we outlined in February, I'd like to touch on a few more recommendations. First and foremost, as we're talking about having that sort of dialogue and bringing New Yorkers into the process and understanding what the government is about, we think that a draft of the technical standards manual put forth by DoITT should be made available to the public before its final issuance in January of 2010 so that we can have that discussion to

make sure that those standards are going to be setting up a framework where people can really be using this information in a valuable way. In addition, the definition of record, while it does seem very encompassing, it seems very broad. Records includes information in any physical form whatsoever that is kept, held, filed, produce or reproduced with or for any agency. Then it goes on to say including but not limited to and it delineates quite a few different types of documents. Just one other one in there that if we are going to be delineating them, we'd encourage video, web casts. We'd encourage those sorts of records to be included also because largely under the leadership of this committee, it seems that that's a direction the city might be going in the long term and that's something we'd want to make sure doesn't get lost in the shuffle here before that's the case. So, again, Intro 991 is a great step forward for the City of New York. We do hope that it is adopted. I'd like to touch on a few quick little points from the prior testimony. We've talked about other agencies being included. The EDC was one of them and the Board of

Education. One other one that maybe some of my colleagues here at the table would also like to see in there is Board of Elections. In thinking about other entities that aren't often subsumed by these sorts of rules, Board of Elections would be one of them. As far as the framework for talking about it as a customer relations issue, we think that primarily we're New Yorkers, not customers, and that we should be thinking about this as our right to know this information, our right to have access to this information. It shouldn't be thought of as a consumer/supplier sort of situation, but it is the public's information so it's out responsibility we think collectively to think through now to provide that in full. One other thing that came up in the testimony from Local Law 11 in February, and it seems to have been touched on again with the administration today, is the email subscriptions and talking about that as one way to stay in touch. Chair Brewer, you raised the question about RSS feeds. It seems that this was the second hearing in a row that we've been talking about this sort of government data availability and the

administration has had some very clear, if not misunderstandings, they certainly could use a brush up on RSS technology. By no means is that my area of expertise, but I'm sure there are a few dozen people that have been through this room today that would be able to sit down and have a conversation with them about the value of that and help them think through creative ways to do it. That's disturbing that it was four months later and we're still sitting here without, it seems, a very clear understanding of an RSS feed could offer in this sort of scenario. Once again, thank you for your time.

CHAIRPERSON BREWER: I think Board of Elections is state, otherwise I would agree with you. I mean that's the same problem we have. I can't do anything about that. I think DoITT understands RSS. The gentleman was not from DoITT, but I appreciate that. Thank you.

DENORA GETACHEW: Good afternoon, Chair Brewer, and other members of the Committee in Government. My name is DeNora Getachew and I'm the Director of Public Policy and Legislative Counsel for Citizens Union of the City of New

York, an independent nonpartisan civic organization of New Yorkers that promotes good government and advances political reform in our city and state. I am joined by my colleague Rachael Fauss, our Research and Policy Associate who is going to talk to a little bit more of the technical components and I'll talk to more of transparency and accountability of government components of the bill.

CHAIRPERSON BREWER: You're welcome to summarize.

DENORA GETACHEW: I'm definitely going to summarize. As a matter of fact, I'll turn it over to Rachael and she can go through the technical stuff.

RACHAEL FAUSS: We definitely support Intro 991's singular web portal. Obviously we all know what would mean in the bill, so I'm going to skip ahead here. The bill does not specify where on nyc.gov or a successor site the web portal would be housed or what agency would be responsible for maintaining the website. Given that we've testified on Local Law 11 before and we know how the Department of Records has

responded with Local Law 11, we think that they have some practical experience with implementing a bill that requires them to work with other agencies and to collect data, so we think that they could be perhaps a potential site. Obviously, something improved and better would always be welcomed. Whatever agency or agencies are ultimately responsible for maintaining the portal, we recommend that sufficient resources and staffing be provided to ensure that records are posted in a timely manner and that staff have appropriate technical expertise. DoITT should play a major role in the development of the website to make it as user friendly as possible. We support a lot of the more technical aspects of the bill. Specifically, web syndication technology, requiring records to be presented and structured in formats that permit automatic processing, not having restrictions on access or use of documents. I'm just going to touch on why RSS is an important thing. It let's the public know immediately of what documents are posted that would be of interest to them. Similar to how they would subscribe to news feeds when they want to

know the latest breaking news. We really feel that timely access through this type of technology can let the public weigh in on important decisions when they need to at that moment so they can effectively hold government accountable. Formats that allow for processing, such as CSV, will let users more easily process and analyze government documents by allowing them to create programs which can analyze trends and look at changes in spending or conduct other types of analysis. Just as a little side note, Citizens Union Foundation, our sister organization, publishes Gotham Gazette, which is an online news source. They would love to have this kind of access to information. They would be one of those types of organizations that would want to create the programs to process the information and provide it to a wider group of New Yorkers.

CHAIRPERSON BREWER: Every reporter would love to have this information.

RACHAEL FAUSS: Absolutely. We also support that there not be proprietary requirements and registration requirements. On the technical standards and internet record

policy, we support the creation of a technical standards manual through the consensus approach that's lined out in the bill. But we believe that there should be greater specificity regarding the selection of voluntary consensus body standards and most importantly the public process for weighing in on standards. It's not clear in the bill whether DoITT would be using preexisting standards or developing new standards in consultation with voluntary consensus bodies. If the intent is to use preexisting standards, we believe that there should be specifically in the bill an opportunity for the public to comment on the use of preexisting standards. If new standards are to be developed, we believe that there should be more transparency about how voluntary consensus bodies are selected because the composition of such bodies is crucial to ensuring that there is a broad base of stakeholders represented. While we don’t think that there should be a mandate of who should be around the table in developing standards, the bill could require DoITT to report on the reasons for choosing particular consensus bodies.

Additionally, throughout the entire process, if new standards are developed, we believe that there should be the opportunity for general public comment throughout the process.

DENORA GETACHEW: I'll just summarize some of our concerns with respect to the agency compliance plans and how we can make this more transparent, along the lines of what Rachael talked about. We do support that there be rollout periods for different classifications of records because we understand that there might be some records that are difficult or complex that can't be put online immediately. We do recommend that there are ways to ensure that the agencies have appropriate resources and tools to meet the requirements of the bill and make sure the classifications are clear so that the agencies don't avoid posting things online because they classify them as things that are just too hard or too complex to do. We want to make sure, again, that the purpose of this is for New Yorkers to be able to access information and there are not impediments to them being able to do so. We would also recommend that the Council consider amending

the bill to create mechanisms to allow for greater evaluation of this implementation. Specifically, this could be in the form of an annual report compiled by agencies, including their statistics on the numbers and types of records available online, the number of hits they've received on the web portal and other issues related to implementation. As the Mayor's Office said, they want to know what users want and this is one way to make sure that they know what users want and what kind of information they should be posting online. We would also recommend that the bill require DoITT to hold a public hearing to assess the user friendliness of the web portal, solicit recommendations for how to improve the site and evaluate the timeliness of record posting. Again, we thank you for the opportunity to testify on this legislation. We hope that you will work to move forward with it.

CHAIRPERSON BREWER: Thank you. I think one of the ways that this could be evaluated, which is sort of what you're saying is through CoPIC, which is the Commission on Public Information and Communication. About six of us

know what that is. The point of the matter is that would be a good tool to get public evaluation and public input. Joshua, you wrote a lot. Are you going to summarize?

JOSHUA BREITBART: I'm going to summarize my summary.

CHAIRPERSON BREWER: Oh good.

JOSHUA BREITBART: My name is Joshua Breitbart. I'm the Policy Director for People's Production House, which provides young people, immigrants and low age workers with a comprehensive media education for the information age. This is a great bill. I want to thank the Chair, the committee staff and your staff for working on it. I discuss four areas in the written testimony where I think this would have positive impacts, including journalism, broadband adoption, government efficiency and quality of life. You just had the exchange about journalism and I think government efficiency and quality of life. I want to talk about broadband adoption just briefly. People's Production House spurs broadband adoption by showing people how they can use the internet to build community and improve

their lives, especially through digital media production. This bill supports that strategy and we believe it would stimulate demand for broadband in our city. Policy makers now widely understand that compelling content is key to increasing broadband adoption, but some still approach it as if they're adding more channels to a television. The internet is a two-way medium. The power comes not really from the ability to find what you're looking for, but from the ability to shape what you find. Having open data standards makes that possible. The bill paves the way for what DoITT referred to as value demonstration in its recent request for information on broadband adoption programs. It's identified broadband adoption as the key for getting universal broadband in the city. So I think that this bill really would contribute to that. I commend DoITT for the way it already presents data to the public in interesting and useful ways, such as NYC State and NYC City Map. They're cool websites, except that the DoITT way is only one way. As much customer research as they do, when they're coming out with this one application, it's just one application.

We saw an example of that before when the Assistant Counselor to the Mayor said that they put it into plain English, which I understand is a turn of phrase, but what is plain English to you might be Greek to me. Obviously talking about plain English makes that clear in terms of what New Yorkers need to understand something, it needs to be translated into a variety of languages and styles. I think current best practice in the field of broadband adoption programs acknowledges that different communities, even if they want the same information, want it in different forms. That's obvious if you think about it in public health terms. You wouldn’t use the same AIDS awareness advertisement if you want to reach teenagers, adult African American men, sex workers, different communities, even though the underlying piece of information that condom use prevents the transmission use of HIV might be the same. This bill takes that lesson to a grander scale, giving all New Yorkers greater ability to learn and share essential information about our city, expressing it on our language neighbor to neighbor and deputizing us in public education

campaigns. We heard many examples of what that could be. The bottom line is we can't really predict the applications that the mass imagination of the city would come up with compared to a much smaller number of smart people who work at DoITT. I have a couple of suggestions that I feel might strengthen the bill. The gentleman from Open Planning Project, as well as from the State Senate, talked about the need for application programming interface and having open APIs that makes it read/write. I don’t know if this is a current problem with New York City government websites, but federal government websites up until last year had a problem with not being indexable and searchable through commercial search engines. The Center for Democracy and Technology produced a report on that called, "Hiding in Plain Sight". Because essentially you could navigate to all this information, but if you Googled for it, you'd never get the federal government's information. I believe that was corrected in the E-government Reauthorization Act in 2008. Again, it's just a way of providing information to people where they are and not making them come to the government

website to find it. Also, I've spoken to this committee about the importance of mobile technology and I think it's important that the bill references that. I think that putting it into a mobile accessible format is one thing that volunteers or developers would do very quickly if the material was available in the raw data format and if there were APIs for it. But just to reiterate the issues, the Pew Internet and American Life Project has shown that disparate broadband adoption and personal computer use in the home, much less pronounced on mobile devices, African Americans and Latinos in particular are avid users of mobile devices for non-voice data applications. This bill would support delivery of city government information to mobile devices and as a result to African Americans, Latinos, seniors, young people, non-English speakers, people with physical disabilities and practically every group that suffers from a digital divide. Based on that, People's Production House strongly thanks this committee for taking up this issue.

CHAIRPERSON BREWER: Thank you all very much. How do you think that the RSS would

work? Is that something that other cities are doing? I know you mentioned Philadelphia and Washington, D.C. in your testimony but I feel really strongly that people need to get information when it's current. When you talk to Washington and to Philadelphia, not only in terms of information being current, but are there other ways that they're looking at this data and improving on it in terms of what they provide now? In other words, are they doing what we're suggesting even though not legislatively?

JOSHUA BREITBART: I'll just speak about Philadelphia. I just came last week and I'm going back tomorrow for meetings about the citywide collaborative application for the Broadband Technology Opportunities Program. This issue of open data standards came up there and it actually came out of the Economy Recovery and Jobs Working Group. It was exactly the electronics store example that you liked so much from before was exactly what they were thinking because there was a gentleman there from a neighborhood small business administration saying if we could have access to this kind of market information it would

help us in sustaining small businesses. It would help entrepreneurs. They were thinking about it very much as an economic stimulus engine in addition to promoting tech startups and transparency means and these broadband adoption programs. With respect to RSS, as Chris said, we heard from the Records Department on the Local Law 11 hearing. It's critical and it allows each agency to publish to its own website but have a single city portal, but also have that people can extract that information from whatever flood of city information that we're getting. RSS is absolutely a critical technology.

CHAIRPERSON BREWER: How about this issue of easy to understand? The administration used that endlessly. To me that has, as you suggested, different meanings to different people. Does the bill approach it correctly? What do you think about that easy? You could get caught in process is what I worry about.

RACHAEL GAUSS: I think there's a place for plain English and there's a place for raw data. I don’t think they're mutually exclusive.

CHAIRPERSON BREWER: That was a good answer to that question. Thank you very much for your testimony. The next panel is Frank Hebbert from RPA, Thomas Lowenhaupt and Silona Bonewald. This is the second to last panel. Thank you, go ahead. Tom, go ahead.

THOMAS LOWENHAUPT: First, I'd like to thank you for having this wonderful hearing. I'm going out a much more educated person about the issue. I'm Tom Lowenhaupt, the found of Connecting.nyc Inc., a New York State not-for-profit created to educate New Yorkers about the internet. The central organizing force behind our work is the .nyc Top Level Domain. Think of .nyc as being like .com, .org, or .gov, but just for New York City. Connecting.nyc traces its roots to the Internet Empowerment Resolution passed by Queens Community Board 3 on April 19, 2001. I was then a member of the community board and back then we advanced the concept of Open Data through the use of a Creative Commons license for our Community Board's website. And our Internet Empowerment Resolution envisioned using the .nyc Top Level Domain as the organizing force for New

York City's digital resources. My goal today is to alert the Council to the advantages of using .nyc as part of the Open 311 and Intro 991 initiatives. Basically stated, .nyc provides an unlimited number of names that might be used to identify digital resources. This includes data acquired as part of the city's 311 operation as well as all other databases. The naming of databases is just one example of the naming power of the .nyc TLD. It extends to assigning digital names to objects not normally seen as part of the digital world. For example, by naming every piece of street furniture, ever bench, light post, fire hydrant, tree, et cetera, the .nyc naming system can become a programmer's dream, leading to a more accessible and friendly city for both residents and visitors. It's part of a transition to what's called The Internet of Things. And beyond the local, the Geneva-based CORE is advancing a common naming standard amongst global cities that would allow for closer cooperation and coordination between the world's great cities. While the ICANN continues to ponder the process for activating .nyc, New York City can begin utilizing .nyc in

the public's interest as infrastructure for assigning intuitive human readable names to its digital resources. I thank you for your attention.

CHAIRPERSON BREWER: Thank you.

SILONA BONEWALD: First of all, thank you so much for allowing me to attend. My name is Silona Bonewald and I'm with the League of Technical Voters. I'm actually from Austin, Texas. I'm also here for the PDF conference as well. One of the things that I was asked to talk about is one of projects which is called Cit Ability, citability.org. Basically what we're asking is that all publicly available government documents be on the web and citable to a paragraph level. We have a pretty simple solution that's outlined on our website which is basically the domain plus the path plus the document name creates a unique identifier. You then data/time stamp it and allow people to walk through it on a paragraph basis. One of the interesting things that happens with that is earlier I heard people talking about figuring out the ontology of terms so that people can find things or figuring out how

to make all of those 6,000 different FAQs for 311 available to the public. If you were to go and give all of those this unique identifier that's easily readable, then people can go and use that link on their blog post, on other commentary and all of it would be searchable on the internet. You would have search engines like Google and Bing and all of those actually figuring out those ontologies for you. You wouldn't have to worry about any of this other technology just by using a very simple standard at the very beginning with the URL. So I would highly suggest looking at citability.org, going to the wiki and seeing the different papers. I've been talking with the New York Senate about this, recovery.gov and actually the French government. They're all interested in doing this piece. Thank you.

CHAIRPERSON BREWER: Thank you very much.

FRANK HEBBERT: My name is Frank Hebbert. I'm the Manager of Geographical Information Systems at Regional Plan Association, a nonprofit research planning organization serving the greater New York region. As a research

organization, we strongly welcome the proposals and the Introduction. Information created and released by City agencies is vital to planning and advocacy work. For non-profit and other community organizations, availability of data is often the limiting factor on what you can do. The central data portal in Intro 991 will make it easier to obtain information and to respond more accurately, speedily and efficiently on issues affecting the city. I'm going to skip over the benefits of open government and innovation in the testimony because I think other people have covered them very well. A couple of sort of suggestions and comments that we have is we think Intro 991 could go further on geographic data. It does not make specific provisions for greater sharing of currently unavailable geographic data. The definition of Record in Article 23-301 includes prepared maps. We suggest that the data could be more specific in opening up existing map services created and maintained by the city. Giving unrestricted, free access to all map data in common digital map formats could be transformative. We also think you could be more specific on data formats, so

that the definition of' raw and unprocessed' could perhaps be strengthened. I know it's been discussed a lot already today. But it's that data in machine readable and completely disaggregated that's absolutely essential for future flexibility and innovation in use that could be over the horizon and we can't necessarily predict today. We welcome the use of consensus standards for formats, and we urge that formats used be nonproprietary. We think that the D.C. Office of the Chief Technology Officer has a very good lead sort of case study I think that you could look at as a sort of baseline of minimum requirements for system designed, particular the variety and methods of data formats that they make available. Thank you.

CHAIRPERSON BREWER: Thank you. How is Texas doing? Are they providing data?

SILONA BONEWALD: Well, I'm very, very happy that you all are doing what you're doing and I hope that it serves as an example to my home state.

CHAIRPERSON BREWER: That's what I thought. I just wanted to ask. In terms of RPA,

because you certain have a great reputation and we all know that this data would be helpful to you because you take it and you do great things with it. Tell me just a little bit about what D.C. provides as opposed to what New York does not now. I mean I'm most familiar with the City Planning Commission which I find challenging here in New York. Second, I am dying to get the environmental impact statements up online. I don’t know if everybody knows what they are. I have to deal with developers all the time. The data sits on somebody's shelf for god knows how long and I don’t even know what happens to them. I guess they're somewhere at City Planning. Apparently one library per borough, if you're lucky, gets one. This would be a wealth of material. Those two as an example, is that something that D.C. provides? How is it a challenge here versus D.C., et cetera?

FRANK HEBBERT: I think the reason for bringing up the D.C. example was just the structuring of the data portal, some of the ways they make information like building permits and works permits available in a variety of both.

There is a feed and also as a map file. It's impressive. I don't think that it's a complete example necessarily for some of the things you're referring to and definitely not a complete example for map data. There is a lot of potential with Land Use data that the city currently only makes available in licensed forms. It could be open here.

CHAIRPERSON BREWER: I assume you use EIS data or even DEIS data. Is that something that you can get online now? Or is it only when you got to the developer or to the city or library?

FRANK HEBBERT: We can obtain specific data for projects, but it tends to be licensed for the use on that project. So that limits us in further research.

CHAIRPERSON BREWER: Having EISs online would be really helpful. Would you agree?

FRANK HEBBERT: Correct.

CHAIRPERSON BREWER: Thank you all very much. That's really helpful. Our last panel is Ben Woosley [phonetic], Andrew Brust and Jay

Sulzberger. Sergeant, he does have a prepared statement.

JAY SULZBERGER: Who's our third?

CHAIRPERSON BREWER: It doesn't matter; whoever is here, Jay.

ANDREW BRUST: Good afternoon. My name is Andrew Brust. I'm the Chief of New Technology at twentysix New York. We're a consultancy specializing in application development, business intelligence and other software technologies. I am also a native New Yorker and a former technology professional with the City of New York. In the mid and late 1980s, I was a programmer for the Department of Parks and Recreation and later I was the Computer Systems Director at the Department of Cultural Affairs. I'm sure you can understand that given my career history and my current position, I have great interest in this legislation. The language in Intro 991 seems to speak implicitly to a number of important features, advantages and a technology premise for the City's data sharing platform. But a number of these points deserve to be called out explicitly, so I hope it's okay that I do so

briefly here. Beyond those points, there are just a few extras that I wanted to touch upon. My first one was on raw data, so I think we've covered that in droves. I don't need to repeat it, but what I would say is that if agencies want to build a more end user application on top of the platform, that should be fine but that shouldn't really be the priority. The priority should be the platform. If the data is provided in the right format, then transformation of it from machine-readable to human-readable should be almost trivial. Today, the Atom Syndication Format, which is a particular schema within XML, is a common format for arbitrary, structured data and it can be rendered in human-readable form by most modern web browsers. The Representational State Transfer, or REST, standard is arguably the most popular service protocol for allowing such data to be queried. And so I would certainly recommend that Atom and REST be supported. But the reality is that there are other formats. I've listed some of them in my testimony. I don't think I need to read through all of them. But ultimately, there are going to be a number of

formats currently and a number of formats that have yet to emerge and it's going to be important to support all of them. The way to do that of course is to put something into a good kind of standard unified format and then have a platform where that format can be easily transformed into other formats through XML technologies or others. It might also be worth considering that the system should allow writing data, in addition to reading it. City residents should be able to submit a tennis permit request through this platform, pay a parking ticket, or even a water bill, or a City income tax bill. City natives should be able to request a copy of their birth certificates, and numerous other submissions should be accepted in addition to mere queries for information. Back on the reading side, users and systems should be able to retrieve non-structured data, including archival photographs of specific city lots, maps, titles and deeds, audio from major speeches made by the Mayor and video of Council meetings and hearings as well. Ultimately this could make up for the loss WNYC-TV. The fact is that Channel 31 was a video authority of record and the loss of it

has been significant. The data platform contemplated by this bill, if it supports rich media in addition to textual data, could bring about services that fill the gap from when WNYC was sold off, and it go well beyond what linear TV, broadcast television can do. Beyond the formats, protocols and contents that are produced, this system will require innovations in licensing as well. There was some mention of this previously. The availability of the data that this platform could produce will enable unprecedented analyses, products and services, useful for both commercial and social services pursuits. But to make possible a number of different query and data visualization services, applications will need to cache, aggregate, slice and dice the system's data. To do so, they will need to stage the data in local or hosted databases and the city should expressly permit this so as not to impede the innovation that would result. Beyond a permissive regime around the availability of the data, the city will also need to allow companies to make a market, and to charge for the value-added services they build on top of

the public platform. Certainly, companies should not be charging for the mere redistribution of the data, but they should be permitted, indeed encouraged, to build user-friendly front ends, interesting "mashups," innovative analyses, and inventive integrations of the platform's data. Google Maps should be able to show where the big potholes are. Zagat should be able to indicate which restaurants have a sterling Health Department inspection record. WebMD should be able to create heat maps showing which neighborhoods are hardest hit by an epidemic. And the New York Times ought to be able to indicate which boroughs and neighborhoods are getting the most, or least, arts funding. Retail consultancies should be able to show which precincts are best and least served by certain types of shops. Tourists should be able to see where the cheapest hotel rooms are and where the most availability exists. Members of this Committee should be able to see how well Verizon is living up to its commitment to deploy FiOS service to all areas in all five boroughs. Children's Aid Society should be able to

illustrate where concentrations of child homelessness and abuse exist. Food for Survival should be able to show which ethnic, geographic, economic and age groups are most susceptible to hunger in the city. And none of these organizations should have to stop and wonder whether they are using or republishing the data in some unauthorized form. I'll just kind of skip to a couple more points to keep it brief. The City and its agencies should be permitted to implement the back-end platform for this system as they see fit, whether they do so using Java, PHP, Ruby, C#, Visual Basic or even COBOL. I would imagine that agency implementations would need to be signed-off upon or certified by DoITT, but as long as they produce their output and solicit their input using the correct formats, standards, protocols and interfaces, that should meet the whatever litmus test may exist. I'd like to close on an issue of civic pride. The City of New York is a unique municipal government within the State. Most cities are contained within counties. The City of New York, as you well know, comprises five counties, and provides the services that in other

parts of the State are delivered by special districts, incorporated villages, towns, cities and counties. As such, our data standards system should serve as a model for each of these distinct types of government within New York State. So let's not just do this the right way. Let's do this in an unprecedented exemplary, creative and exciting way. Let's make this the time in history when the economy was down, but the great tradition of commerce and ingenuity in the City of New York was nonetheless invoked to bring about innovation, opportunity and a new standard in good government, adopted by other governments in New York, and other states. Thank you for your time.

CHAIRPERSON BREWER: Thank you very much. You're going to be the last speaker, so why don’t you come and sit up there. Go ahead Jay.

JAY SULZBERGER: I'm not sure what to say. So much has been covered and so much has been said that I wanted to say.

CHAIRPERSON BREWER: You can summarize sir.

JAY SULZBERGER: Obviously I'm on your side, Gale, on Intro 991. I'll just say a few

sentences. In truth, open standards can only be defended by free software and I won't include a discussion of the mechanism here. Finally, the last thing brought up about the issue of people doing add-ons and everything, if the data is published by the City of New York and it's freely licensed, then indeed if somebody can figure out a way to put something in the front of it and make money off of it, that's what free licensed means. It means that that can be done in addition to being reproduced for no money; somebody can make money off of it by packaging it up. They can't further restrict the actual data of course. Finally, I suggest a solution to the portal problem. The portal can be very simple. If there are 80 agencies, it can be a web page with 80 lines and links and that can be the portal. I was a little bit surprised to hear the fellow from the W3C worrying about the portal. If the portal is heavy, ugly and incomprehensible such as SourceForge, the default thing which is almost impossible to use, but if it's more like Google or a single, simple near text web page with a line for every single agency and a link then to the

agencies, that solves it. I believe in portals. There should be a central point. Thanks.

CHAIRPERSON BREWER: Thank you, Jay. Just introduce yourself.

JAY SULZBERGER: I'm sorry. I'm Jay Sulzberger. I used to do a lot of propaganda with various free software groups. I don’t do much nowadays. I've retired to a position as consultant. I guess I'm a hire to some consulting company that does some statistics for HRA.

KAYZA KLEINMAN: My name is Kayza Kleinman. I've been with NonProfit HelpDesk, which is a project through the Jewish Community Council of Greater Coney Island for a number of years. I'm not going to date myself.

CHAIRPERSON BREWER: The mike is hard to move so she's having trouble moving it. Go ahead.

KAYZA KLEINMAN: I think I got it this time. The first administrative testimony highlighted all the problems that we will have in trying to do this. Having heard this kind of thing before, I have to say we shouldn’t really let that get in our way.

CHAIRPERSON BREWER: It won't.

KAYZA KLEINMAN: Knowing you, I don’t believe it will. I think a lot of the advantages have been mentioned. I think it needs to be highlighted that the issue of open standards is important not just for citizens but also the data sharing between organizations which is something that I know has been a huge issue that’s been talked about more than once. I think it's a really, really important. I think even if we're talking about cost, I think that eventually the cost saving of being able to share data efficiently should help really cut the cost of creating open data standards. If citizens can share them, then certainly other agencies can manage to make use of that data as well. So it's killing two birds with one stone as far as I'm concerned. When talking about accessibility by people, organizations or whatever, I think we really need to keep in mind that whatever goes up there needs to be handicapped accessible. It's something I don’t think I heard anybody mention, but I think it's a huge issue. If there is a constituency that really could use the web, it's

people who have handicaps of various sorts, whether it's mobility, visual, et cetera. I think it would be a crying shame not to put that in from the ground where it's a lot easier to design it in than to add it on afterwards. It's a big issue. Another issue that was touched on but I think really wasn’t addressed were privacy concerns. I'm concerned in two directions. I keep up on this stuff and many municipalities have used privacy concerns as an excuse to not share important data. I think that needs to be addressed upfront. That's a good reason why there need to be policies and procedures in place. Private data should not be shared inappropriately, and that has to be built into the whole thing from the ground up rather than tacked on later. I do think that there is a place for easy-to-understand formulations of data, but as others pointed out, there are different ways of looking at it. What's English to me is literally not English to someone else. I think that both things need to be done. I think that's been covered and I don’t need to get into more detail about that. I would like to make the point about tech neutrality, which was

raised by somebody else. Besides the fact of all of the possible applications that can be built on a tech neutral platform, in terms of accessibility, the more technology neutral the data will be, the easier it will be for people and organizations with low resources to make use of that data. I don’t need to necessarily have the latest and greatest machines. If I can get it at a very basic level, then even I can put it turn on a new project on an old machine and do something useful with it. I think that tech neutrality serves a lot more than just the business, it also serves the less enfranchised. I would like to mention two examples that were mentioned in prior testimony of the unexpected things. There is a search engine that's in beta but it's publicly accessible. It's called Wolfram Alpha. Don’t ask me why they cal themselves that, but it is a fascinating site. You get on there and ask questions and it basically pulls together huge amounts of statistical data to answer questions. But this can only be done if the raw data is there. It couldn’t necessarily do it if it just had like the two summaries that somebody in the

city decides on. But if there is all that raw data, the calculations and computations can be done. This is great because nobody has to think in advance about what somebody is going to need. This is a really cool example. The Google Maps mashups are another very good example of the kind of stuff that can be done when data is freely accessible.

CHAIRPERSON BREWER: Thank you very much. Andrew, thank you, for all of your background you brought to the hearing. Thank you very much. I do want to say in conclusion that I think everyone agrees that the time is right for open government data. It should be done in the right way and making all public data available. What was interesting about today was that it can be person-friendly in one format and have the raw data at the same time. You have different users with different needs and the fact of the matter is they're compatible. They're not mutually exclusive. I hope that the discussion today leads the administration to understand that we're really serious about this. There is a lot of interest and I think that we can come to an agreement about

what should be online. Thank you very much. This hearing is concluded.

C E R T I F I C A T E

I, Donna Hintze 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 28, 2009