

**DEPARTMENT OF INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS TESTIMONY
BEFORE THE CITY COUNCIL COMMITTEE ON TECHNOLOGY
OVERSIGHT ON IMPLEMENTATION NEW YORK CITY'S OPEN DATA LAW
MONDAY, OCTOBER 27, 2014**

Good afternoon, my name is Albert Webber of the Department of Information Technology and Telecommunications (DoITT). On behalf of Commissioner Anne Roest, I'd like to thank Chairman Vacca and the members of the City Council Committee on Technology for convening this hearing on the implementation of Local Law 11 of 2012, the Open Data Law. Open Data is a major priority for DoITT, and I am here today to share our progress to date, highlight the recent enhancements we have made to the NYC OpenData portal and to discuss our strategic plan going forward.

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

Local Law 11 of 2012 requires that all public data sets maintained by New York City agencies or departments be made available on the Internet through a single web portal. DoITT maintains this portal, NYC Open Data, at nyc.gov/data. We currently have 1,268 unique datasets available via the portal. Each year, all agencies and departments are required to submit an inventory of public data sets that they have not yet made public along with the dates that they intend to publish them. This inventory is compiled by my office and published as the New York City Open Data Plan. The latest edition of the plan, published on July 15th of this year includes 345 data sets scheduled for future release. Since July 15th, 60 of the data sets on the plan have come due and I'm happy to report that we have successfully published 57 of them or 95%. We are on track to publish an additional 100 before the next edition of the plan, which is scheduled for submission to the Council on July 15, 2015. As per the law, all public data sets maintained by City agencies and departments must be published by December 31st, 2018 or a reason must be given as to why the data set cannot be released and when it is expected to be released.

To enable the Council and the public to track the progress we've made, we included listings of all the data sets which we published since the inaugural release of the plan, as well as any data sets that were removed from the plan and the specific reason why they were removed, for example, being merged with existing data into a combined data set. We also provide a dashboard of available data sets which is updated monthly and provides a concise listing of data sets currently available on the portal and those scheduled for future release.

I would now like to turn to some of the enhancements we have made to the portal. These improvements were made in response to feedback received from users of the portal and were developed in close collaboration with the Open Data community.

The portal has been improved to enhance search functionality. Initially, the portal relied on industry-standard technology, but it was unfortunately not sufficient to navigate the thousands of Official City Data Sets and Community Generated Views available on the portal. Working with the Open Data community and our technology vendor we were able to refine the search function to provide much more relevant results. For example, a search for "311," which is one of our most popular data sets, using the old search technology, returned the official 311 data set on the 9th page as the 214 data set in the list. A user would then have to click through pages of community generated views to find the actual data set. With the new technology released just last week, a search for "311" now yields the entire 311 Service Request data set as the first result.

While discussing the search enhancement I mentioned two types of data sets: “Official City Data Sets” and “Community Generated Views.” I’d like to take a moment to explain the difference. One of the NYC Open Data portal’s most useful features is that it allows users to create a specific view or filter of a data set and save it for future use. For example, a user may only be interested in the Department of Housing Preservation and Development’s complaints in his or her zip code and create a filtered view with only this information. We feel this is important functionality to allow users to access relevant information. However there was so much user activity on the portal that it became difficult to find the original data set among all the Community Generated Views. Therefore, our release last week created a feature that allows for searching or browsing of only the original Official City published data sets. While all the Community Generated Views are still available to users of the portal, users should now have a much easier experience locating the original City data sets.

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

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

Automation: When we first launched New York City Open Data portal in 2009, then known as the NYC DataMine, the Open Data law had not yet been passed, and all the data posted to it was done so on a voluntary basis. We will be focusing on automating these older datasets to ensure they remain updated and relevant.

Partnerships: There is a vibrant Open Data community in New York City. The improvements we made to the Open Data portal described earlier would not have been possible without the input and spirit of collaboration embodied by the New York City Open Data community. We value the feedback we receive both on what data sets should be prioritized and the technology used to deliver them and view our relationship as a true partnership. The publication of these data sets is useless unless we have users who use it to make positive change for New Yorkers. There are events nearly every weekend and evening, and we receive hundreds of comments from users about City data and strive to answer every one of them. To that end, we are in the process of adding a new staff member to ensure that all comments and data set nominations receive a prompt reply, and that the City is well represented at community events and that we remain engaged with this community. This representation will allow us to deepen the communication channels between the City and Open Data users, enabling us not only to be more responsive but also to communicate new data and features we have deployed or that are under development.

Innovation: Last week our Open Data portal vendor, Socrata, unveiled the next generation of its platform. This week we are working with the vendor and members of the New York City Open Data Community to conduct usability testing on the new site to ensure that it meets all the needs of the City and our users. This will give us the chance to give feedback to the vendor and to customize the product to fit our specific needs. The new interface will offer New Yorkers a host of new features:

- A new online experience focused on allowing non-technical New Yorkers, to easily access information in a way that's meaningful for them, including intelligent and automatically-created data visualizations based on the data being analyzed. For example, when New Yorkers search for 311 data, the tool will automatically create data visualizations showing Service Requests by Agency, a map of data by Community Districts, and a chart showing Service Requests created over time.
- Every dataset that includes location-based information can automatically be visualized by those attributes and by custom boundaries, for example by borough, community board, or zip code.
- Technical users can embed NYC data and visualizations in their own web sites and access them via tools of their choice.
- Significantly improved performance, allowing users to easily and quickly access and analyze millions of rows of data.
- Ability to query data based on location, for example how many pothole requests have been submitted within a specific Community Board.
- Pre-built tools to allow technologists to easily access Open Data based on the technology of their choice like iPhone or Android.
- Improved data automation tools, which will simplify agency Open Data automations going forward.

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

In closing, DoITT strives not only to comply with the Open Data Law but to excel in creating an open and transparent government. We are working diligently to ensure New York City remains a national leader in the Open Data space.

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