

To: NYC Council - Committee on Technology
From: Noel Hidalgo, Executive Director of BetaNYC



Re: Written testimony for NYC's open data laws oversight hearing

Wednesday, 21 September 2016

To the NYC Committee on Technology & Chairperson Vacca,

Introduction.

BetaNYC is a civic organization dedicated to improving all lives in New York through civic design, technology, and data. We envision an informed and empowered public that can leverage civic design, technology, and data to hold government accountable, and improve their economic opportunity.

We were founded in 2008 as a “meetup” to discuss open government in NYC. Our work empowers individuals and local communities to build a civically-engaged technology ecosystem and provide for an honest and inclusive government. We want New York's governments to work for the people, by the people, for the 21st century.

BetaNYC demystifies design, technology, and data to the point where anyone can use it, create it, and participate in the decision making process. We host a number of online platforms that provide the general public a mechanism to share ideas and data.

In the last twelve months, our community has grown 700 new members. We are now over 3,700 civic hackers who are ready to use our talents to help our neighbors.

How open data has grown our community.

In the last twelve months, we have hosted four significant events—NYC School of Data, NYC TreesCount Data Jam, a BetaTalk conversation on affordable housing data, and another BetaTalk on the release of the City's second largest data set, NYC 311 call inquiry data. We feel that these events have shaped where we are today, and point to a clear future.

We would like to thank Council Speaker Melissa Mark-Viverito; Manhattan Borough President Gale A. Brewer; Council Member Ben Kallos, NYC Department of Information Technology and Telecommunication; NYC Parks; NYC 311; Mayor's Office of Data Analytics; and the Mayor's Office of Technology and Innovation for joining us as partners.

We would also like to thank all of the government employees who joined us for all of our events. Without their expertise, we would never properly demystify how government technology and data works.

How we use the City's open data.

Re: Seven new pieces of open data legislation.

We fundamentally believe that these seven pieces of legislation were the right additions to make. We can see that they have strengthened the City's open data practice.

We were honored to be a partner with NYC Parks and prototype examples of good data dictionaries and geospatial / address standards.

We commend the Mayor's Office of Data analytics (MODA) for engaging with our community to gather public feedback and help ensure that the City's data users have a voice.

We absolutely agree with our colleagues at Reinvent Albany and NYPIRG on the following:

- *automating 100 datasets in 2016. That makes a total of 200 datasets out of 1,600 automatically updating on the Open Data Portal whenever the agencies internal data changes.*
- *publishing important new datasets including the City Budget, City Record Online, Seven Major Felony crime data, a huge TLC trip dataset, and NYC 311's call inquiry data.*
- *thirty of eighty agencies reported on FOIL responses that included public data that is or should be on the Open Data portal.*
- *DoITT and MODA staff are reading comments and requests on the Open Data Portal and responding.*
- *The administration's Open Data Team published their annual update on-time.*

Re: NYC's Civic Innovation Fellows with Manhattan Borough President Gale A. Brewer

Reference URL < <https://beta.nyc/programs/nyc-civic-innovation-fellows/> >

Since July 2015, we have partnered with the Manhattan Borough President Gale A. Brewer to develop an innovative civic engagement program we call the Civic Innovation Fellows. With financial support from the Fund for the City of New York and Data and Society Research Institution, we have explored how Community Board district offices can better use data and technology.

We have reviewed Manhattan Community Board offices, their data & analytics capacity, and how they share information across digital streams. We discovered that Community Boards have a desire to use open data but don't always have the bandwidth, education, nor tools to process the City's data. Frustratingly, we discovered that there are zero best practices bringing Community Boards to use 21st century tools nor teaching them how to use NYC's open data.

Using insights developed with Pratt Institute's SAVI program, we have outlined a framework to teach Community Boards, Council offices, and community based organizations how to use NYC 311's service request data and NYC Park's TreesCount data. Frustratingly, we have not found any financial support to teach the City's data to its residents.

NYC 311's service request and NYC Park's TreesCount data set are two of the 1,500 datasets that should have educational rubrics attached to their data dictionaries. **While we love the data dictionaries law, this should be considered the floor, not the ceiling. Every data dictionary should contain mini-tutorials explaining how to best explore the data and how to embrace the data portal's functions.**

While some of my colleagues might criticize the current data portal's user experience, a skilled user can quickly navigate around a dataset and easily produce reports. Granted, you will need a lightning fast connection, a large monitor, fast computer, and a bit of luck, but for now, we have the best portal tool for the widest audience. **Moving forward, NYC needs to develop tools to better suit bulk data users. We encourage the city to explore open source data sharing tools that will give agencies the flexibility to host and share their own data. Fundamentally, we want Agencies accountable for producing high quality datasets.**

This year, will be working with the Manhattan Borough President and the Fund for the City of New York to improve our Civic Innovation Fellows program and create a simplified online curriculum for all. **We should note that this is something that this Administration and the Council would benefit from and we hope to have your investment as we build the next level of open data education.**

Re: NYC Parks Data Jam

Reference URL < <https://beta.nyc/2016/08/05/treescount-data-jam-2016-report-back/> >

The most important insight from the NYC Park Data Jam came from the evolution of a “user centered data release workflow.” Building off of user centered ideas, we believe that data sets should go through some user testing and like all technology products fit into a continuous improvement loop.

When it comes to continuous improvement of NYC's most valuable, most used data sets, we believe that every agency and every dataset should go through this release workflow.

- **Phase 1: Research & Discovery**
 - Establish target audiences
 - Draft data standards that appeal to broadest possible audience
 - Draft data dictionary
- **Phase 2: User Testing**
 - Release sample dataset for feedback
 - Perform user testing and get feedback on data dictionary & dataset
 - Develop a framework and/or guide as how to explore the dataset's important values
 - Allow time for revisions
- **Phase 3: Initial Deployment**
 - Upload data set and data dictionary to Portal with an event or video explaining the key features of the data.
 - Allow time for user testing on Portal and gather open feedback.
 - Share insights and communicate them out to the public. Continue to gather public feedback.

NYC Parks Data Jam Event Metrics

- Total Participants: 196
- Data Jammer Participants: 144
- Workshop Participants: 27
- Participant Diversity: 45% male / 39% female / 16% unspecified / 1% other
- Community Group / Public Stakeholder Groups: 5
- Children in on-site daylong childcare: 7
- Projects built: 23
- Winners: 5
- Organizers: NYC Parks and BetaNYC
- Partners: NYC Mayor's Office of Technology and Innovation, NYC Open Data, Civic Hall, Microsoft, Carto (was CartoDB).

Re: NYC School of Data

Referance URL < <https://schoolofdata.nyc/a-brief-recap-of-nyc-school-of-data-2016/> >

NYC School of Data proved there is a massive community in NYC that wants data and technology to be demystified. For one day, we featured 18 sessions, 40 presenters, 16 came from NYC Government, & three were elected officials.

A quarter of the event contained a data jam looking at how to address economic justice.

A critical component to our event was child care. This ensured diverse attendance. Several of our speakers couldn't attend if we didn't offer childcare. For older children, make it explicit they could attend with their parents.

Should the city host an open data summit, please offer child care. It makes a fundamental difference.

NYC School of Data event metrics

- We featured 18 sessions, 40 presenters, 22 are women, 16 came from NYC Government, & three are elected officials.
- 372 tickets; checked in 260+ people; 11 children under 10 y.o. attended the event.
- 49 people offered to volunteer. 30 labored.
- Eight NYC High Schoolers learned to navigate NYC's open data portal, manipulate NYC's 311 data, & map this data into CartoDB.
- One family attended and represented three generations: Grandma—organizer and data visualizer, Father—data advocate, & Daughter—aspiring civic hacker. □
- Our event leadership team consisted of seven, three are women & five are people of color.
- For a third of the day, our hashtag was a trending topic on Twitter — #nycSoData.
- Organizer and Host: BetaNYC
- Partners: Manhattan Borough President Gale A. Brewer, Fund for the City of New York, Data and Society Research Institution, Microsoft, Carto (was CartoDB), Accela, Code for America, Internet Society of New York, NYC Mayor's Office of Technology and Innovation, NYC Mayor's Office of Data Analytics, Council Member Ben Kallos, Reinvent Albany, and Civic Hall.

Insights from the last year

Three types of data users.

While there are many types of data users, this year's research has led us to see three general types of data users.

- The general public who wants to see what the numbers mean.
- The data hacker who wants to see the data and play with the data.
- The scientist, business, or government entity who wants to enter the matrix and download data bulk.

For NYC's open data practice to be the best in the world, it must consider that these are the three types of users and each one has their own unique needs.

Concern about the seven new open data laws.

The passage of these seven new laws demonstrates that NYC's open data program has hit maturity. As a municipal practice, the city needs an investment ensuring this open data practice lives beyond the 2018 goal. Currently, the seven new pieces of legislation have created an unstable situation that will overwhelming the current open data team's ability to scale out NYC's open data practice.

Discrepancy in open data production.

We see several agencies who completely understand the benefit of open data and collaborating with the public. Yet we some who seemingly refuse to accept that law is an essential part of the 21st century.

In the past year, NYC 311, NYC Parks, Department of City Planning, NYC DoITT's GIS division, and the Taxi and Limousine Commission have produced data tools and data that are exemplary of the City's future. We hope their data teams continue to be supported and given the resources to lead by example.

More civic engagement events.

For the agencies that understand the value of open data, they have a unique opportunity to partner with the public and use hackathons and/or data jams to explore new insights and improve their data quality.

Quality data and data guides.

Community Boards and the general public are desperate for usable data. More importantly, they are desperate for content that will help them make sense of the data, aka an open data curriculum. San Francisco has a series of video guides that teach the general public on how to use their data portal < <https://data.sfgov.org/videos> >. We would love to see these videos appended to data dictionaries and see other mini-tutorials explaining how to best use the data portal's functions and how best explore the data.

Growing municipal data standards.

The future of municipal open data does not have one central clearing house processing all of the City's data and share it with the public. The internet does not work where there is one website servicing all of the world's information. While we agree that there should be one central catalogue of data resources, we fundamentally feel that Agencies need to own the responsibility in producing high quality datasets and understanding how the public uses its data.

For NYC to be the number one open data practice in the world, it must adopt a practice of establishing data standards, protocols, and coach Agencies to use those standards and protocols. NYC should be exploring

open source data sharing tools that give Agencies the flexibility to host their own data and interface directly with the public.

User center data release workflow.

When it comes to improving NYC's most valuable or most used data sets, we believe that datasets should go through a review process that bakes in public comment. This workflow is modeled after practices used in manufacturing and software development. We call this a user centered data release workflow. This workflow ensures that continuous improvement through public feedback will strengthen data quality and data products.

Enlisting public engagement through the data release process is a key part of every insight previously stated.

What we see'd like to see.

Fundamentally, you cannot cure NYC's open data future with a particular product but a dedicated practice.

We continue to encourage this Administration and the Council to place a significant investment in a dedicated open data team. This is the only way to ensure the City's data practice can scale across all City's agencies.

In an ideal world, we would like to see a Chief Data Officer resourced with a team and dedicated resources to shepherd the city's data practice, data standards, and outline education best practices. We do not see this office as counter to the current leadership. We see this team as a complementary—a policy, consulting, and product shop. From our research, this office would focus on the following.

Leadership & Standards

- With the City's technology and data leadership, draw together agency leaders to ensure data and technology standards produce data as a renewable resource.
- That technology procurement practices selects systems that allow for data version controls, API driven backends, with modular or open sourced capabilities.
- That these systems are always available, reliable, consistent, accessible, secure, and flexible to support an agency's mission.
- Help agencies perfect a feedback loop around data quality and public comments.
- Provide leadership to steer the production of citywide data policies and data standards.
- Ensure public feedback baked into the City's Technical Standards Manual (TSM).

Technology & Tools

- Prioritizing large scale data projects in conjunction with data owners.
- Standardize and automate future dataset dissemination.
- Sets policies for responsible data systems.
- Ensures that data reform and modernization—how agencies collect, uses, manages, and shares data—moves toward the stated goal of building “fact-based, data-driven decision-making” programs or policies.

- Oversees the development and stewardship of data sharing tools that enable agencies to share their own data and collect feedback directly from data users.

Evangelism

- The evangelism pillar would reach out to the public, industry, academics, and other branches of government to promote data, data services and tools.
- Inward, this pillar would connect the city's data managers to world's best and the brightest to ensure that Agencies are thinking about their best practices.
- Outward, this pillar would develop collaborations that further development of open data products and services.

Education & Trainings

- Inward, trainings would cut across agencies teaching and promoting data proficiency skills, geographic information systems, and data tools like MS excel, and R.
- Outward, this pillar would ensure that specific uses of data are taught and shared across the city. Community boards, institutions, and organizations would serve as a feedback loop to thoroughly define the public's need for technology, data, and data quality.

While some of these things can be done from the outside, we are at a make or break legacy point. For the City's open data practice to lead the world, they need to be done from within government.

Thank you for giving us this opportunity to give our story.

The BetaNYC Leadership Team