

To: NYC Council - Committee on Technology
From: Noel Hidalgo, Executive Director of BetaNYC
Re: **Int 1528-2017 & Int 1707-2017**



Wednesday, 20 September 2017

Summary:

- Support 1707-2017
 - We have concerns about extending the deadline to 2021.
 - We have concerns around the broad user/usage reporting language. The language needs to provide transparency around use while protecting people's privacy.
 - We encourage the City to adopt permissive copyright around data sets and data products. We highly recommend the adoption of Creative Commons Zero and the GNU General Public License v. 3.0.
 - We call for a public right of action to ensure data accessibility across Administrations.
 - We call for clarity around dataset compliance and the creation of a simpler interface to access which datasets are in compliance with existing open data reporting laws and which data sets have geocoded elements.
- Support 1528-2017
 - We are very interested in knowing which datasets are derived from the City's unified FOIL process.

Additional statement added as oral testimony:

For the last few years, we've discovered a number of data issues related to geocoding. Geocoding is the process of translating a point into a addresses or attributes, or vice-versa. Geocoding is how one turns an address into a specific location or finds out the community board / council district of a point.

New York City is unique in providing a free municipal geocoder accessible to municipal agencies and the general public. We both use NYC Geosupport produced by NYC Planning. Frustratingly, we continue to come across a number of issues that are not easily addressed. We call for the Administration and Council to provide adequate resources to turn this main-frame tool into a modern, open source tool that will enable to public to point out problems and collaborate to fix them. We call for transparency around the data, code, and process. For New Yorkers to know where they are going, they must know where they are.

To the NYC Committee on Technology & Chairperson Vacca,

Introduction to Testimony:

Thank you for this opportunity to throw our support for Intro 1707-2017 and 1528-2017. As a community representative, I want to briefly agnolish the great work of those who have come before us and paved a path for us to follow. BetaNYC would not be here if it wasn't for the pioneering work of GISMO, Fund for the City of New York, NYPIRG, Transportation Alternatives, and others who have helped shaped our community's understanding.

This committee's present and previous leadership has made NYC the most robust open data city in the world. From my travels, the world looks to NYC for its open data leadership. Today, I will share how we can stay ahead.

NYC School of Data

To celebrate the 5th anniversary of the city's open data law, inaugurate NYC's open data week, and celebrate international open data day, over 350 people attended NYC School of Data. In its second year, we've ironed out a platform that empowers community and government to collaborate. This year's conference helped community based organizations learn how to use the city's data, helped individuals level up their analytic skills, and build stronger collaborations across the city.

Data Jams

In the last 15 months, we've hosted two agency related data jams and on the eve of our third. As highlighted in the City's own open data report, we're helping agencies explore their data questions and derive new operational insights. We believe in a structure formula that combines civic engagement with government stakeholders to address operational questions. In the coming year, we look forward to working with other agencies to hone this model.

To continue our event programs, NYC School of Data and Data Jams, we would love to have municipal support and help cover operational costs. We make sure our events are open and accessible by providing classes, healthy food options, and free onsite

childcare. We could use city support ensuring that these programs are accessible to all who want to attend.

Civic Innovation Fellows

We are embarking on our third collaborative year with the Manhattan Borough President Gale A Brewer. In that time, we've exposed 25 CUNY undergraduates to the City's municipal infrastructure and educated them on the value of open data.

In the last two years, we've moved beyond open data and mapped out community board district office technical needs. This past year, we've build BoardStat, a tool that simplifies Community Board access to NYC 311 service request data. In collaboration with the Manhattan Borough President's office, we will launch this tool at this weekend's NYC 311 Data Jam.

Additionally, we have outlined how boards can digitize their attendance tracking and an ideal future of community board websites. Our hope is to get the city's 59 community boards to use and produce structured data. If we ever want to ask siri or alexa, "when is my next community board meeting?" we will need this data structure.

In the next year, we hope to have funds and document how community boards self organize their information flows. We will use this research to outline new open data trainings and build new dashboards.

To continue our civic innovation programs, we would love to have municipal support and help cover operational costs. Our goal is to affect community boards citywide, but can only afford to raise funds to support the Borough of Manhattan. We could use city support ensuring this program scales to all five Boroughs.

Future of NYC's open data program

Five years in we still seem to have significant open data inequality. Though our research we continue to find agencies confused about open data, data standards, and machine readability.

Primary example is NYC DOT and their East River Bicycle counts. Every day, I go over an East River bridge and activate a sensor recording my ride. When I go to the DOT's

website, I can't access this data. Instead, I get a PDF or a [link to the City's open data portal](#).

When I visit the portal, I get a button to download a compressed file with five formatted excel spreadsheets. These files are formatted for human legibility not machine readability. If this was truly a smart city, we would have long embraced a gamification, made this sensor data accessible to the public, and celebrated healthy public transport.

Where open data can be automated, we want it.

We are excited to see language around a two year review of the technical standards manual. We hope this bi-annual review helps educate agencies and gets them to comply with the law.

We are a bit concerned that the current language in intro 1707-2017. The current working seems to pushes the 2018 deadline to 2021. Ideally, we like to keep the existing deadline, and have language that continues the city's open data mandate.

Compliance with existing open data laws

In preparation of these remarks, I was going to point back to the DOT and highlight that they haven't made their street closures data machine readable. When you visit the DOT's website, you will not find a link to the [DOT's Street Closures due to construction activities by block](#) dataset. Yet, there it is on the open data portal! It appears that someone went through the process to automate this database. Frustratingly, this dataset doesn't seem subject to local law 108 of 2015, the law pertaining to geospatial data standards.

As a cyclist, I've been asking for this dataset for the last two years. Why? I prefer to ride in bike lanes. When lanes literally disappear from underneath my tires, I am subject to extreme danger with very little warning. (Sidenote, rarely does DOT put up signs to warn cyclists that streets will be milled.)

This summer, a one mile long bike lane was torn up for two months. Every day, I would have to go out of my way to see how far the DOT milled. Then, wait till they would start resurfacing the road, and wait some more for them to completely surface the road. Having this dataset incorporated into a transit app would be another step toward a smarter city.

Today, I'm happy to see we are one step closer, but this is one of 1550 data sets that isn't geocode in a way that complies with the Geospatial Open Data Standards.

A dataset of datasets

As part of the Local Law 108 of 2015, the city produces a dataset of datasets that reports out which datasets conform to the law and which one will conform to the law.

Nowhere do we get a list of datasets that give other granular data. For example, I want to know which datasets are coded for council districts, community district, or have some other mappable points. In the same way we would love to see greater sensor frequency, We would love to have greater data on the city's data. To that end, we would love to see a data set that shows which datasets are subject to which open data laws, are they in compliance, and, if not, how close are they to that compliance.

As we are discussing 1528-2017, it would be good to know which FOIL requests lead to open data sets and which data sets came from FOIL requests.

We are concerned about intro 1707-2017's vague language around "web portal analytics." We believe that the language should be phrased to protect individual and location privacy.

Private right of action

Also, it is clear that NYC's open data law needs some form a private protection. Sooner or later, we will get a Mayor who refuses to understand the benefits of transparency and accountability. As open data users, we need the ability to protect this access.

As your last mark on five great years, cement your open data legacy by giving us a private right of action.

Closing

Thank you for four great years of turning our open data advocacy into action. We strongly support intro 1707-2017 and 1528-2017.

Thank you,
Noel A. Hidalgo

