

JAMES PERAZZO, ACTING DIRECTOR, MAYOR'S OFFICE OF DATA ANALYTICS
TESTIMONY BEFORE THE CITY COUNCIL COMMITTEE ON TECHNOLOGY
OVERSIGHT ON 2017 OPEN DATA EXAMINATION AND VERIFICATION REPORT
WEDNESDAY, DECEMBER 6, 2017

Good afternoon. My name is James Perazzo, Deputy Director for Strategic Management at the Mayor's Office of Operations, and Acting Director of the Mayor's Office of Data Analytics, also known as MODA. I am joined today by Craig Campbell, Special Advisor at MODA.

On behalf of the Administration, I would like to thank the Committee on Technology and Chair Vacca for their unwavering support of the NYC Open Data initiative. The Committee and the Administration are united in a commitment to transparency in government. The Open Data Law brings this commitment into community organizations, classrooms, and digital laboratories all across the five boroughs. New Yorkers mine and manipulate our data to develop solutions to a wide variety of issues while sharing information on how City programs operate and what they achieve. The Open Data policy that was borne from the Law is unparalleled among American municipalities, and the amendments this Committee has passed over the last three years will ensure that the program thrives in perpetuity.

One of those amendments is Local Law 8 of 2016, the Open Data Examination and Verification (E&V). This law requires MODA to work with three agencies each year to verify their compliance with the Open Data Law. Following the results of MODA's first Examination and Verification report, which we discussed before this Committee in January of this year, MODA developed a more robust examination process for the 2017 cycle. On December 1st, we submitted our findings on the Department of Buildings (DOB), Department of Environmental Protection (DEP), and the Fire Department (FDNY) in a report that is publicly available on the [Open Data website](#).

The process was a success. Each agency identified new datasets for publication and new ways to improve datasets already on the Open Data Portal. Further, the process affirmed each agency's commitment to Open Data. While the full Examination and Verification process is described in detail in the report, I will briefly highlight its major elements here.

The first part of the examination was a dataset questionnaire, which required agency Open Data Coordinators (ODCs) to conduct a thorough inventory of datasets at their agency. This consisted of two main components.

The first was an inventory of the information technology and data systems that support the agency's core business processes. This required ODCs to describe each system, its users, and its

supported operations to ensure that the agency's mission-critical datasets are accounted for in the examination.

The second was an inventory of instances where data is already formatted and transferred at the agency. This required ODCs to identify the existing channels where public and internal stakeholders consume the agency's data, such as operational inter-agency data exchanges, public reports supported by agency datasets, data provided for the Mayor's Management Report (MMR), and data used to fulfill freedom of information law (FOIL) requests.

Finally, the Agency Commissioner, or their designee, was required to certify the completeness and accuracy of the inventory to ensure the full participation of relevant staff in the process. In addition, the Agency General Counsel, or their designee, was required to sign-off on the data inventory, certifying that the process was conducted in accordance with the statutory parameters of Open Data Law.

In completing this process, each agency identified at least one new dataset to publish on the Open Data Portal. DEP identified datasets on the quality of water in City reservoirs and the geographic location of catch basins. DOB identified the Certificate of Occupancy dataset, which the agency published online in June. And next month, the Fire Department will publish the location of fire alarm call boxes and an inspections dataset that were identified during the E&V process.

The examination also included a survey of data quality for datasets already on the Portal. Publishing data online is only the beginning of making data "open": truly opening data requires updating it regularly, identifying and fixing errors, and engaging users and incorporating their feedback. All three agencies identified ways to improve their existing datasets on the Open Data catalog. DOB made data quality improvements to an existing dataset, while DEP and FDNY identified datasets that should be archived from the Portal.

Following the examination process, MODA met with Open Data Coordinators at the surveyed agencies to discuss the challenges and opportunities they face in their role. Insights from these discussions are reflected in the "Recommendations for Better Citywide Compliance" section of the report, which list ten specific actions for MODA, DoITT, and City agencies to take to improve the Open Data program. For your convenience, these recommendations are appended to today's testimony. These recommendations fall into two categories.

The first pertain to MODA and DoITT, which should continue to provide strong leadership and proactive support to agencies in order to affirm the Administration's commitment to the success of Open Data in every area of government. The second set pertains to Agencies, which should leverage the assistance of MODA and DoITT in embedding open data into existing processes. To

be successful, Open Data must not be a one-off reporting requirement, but a foundational element of data analytics and data sharing within City agencies.

Given the pace of technological and administrative innovation, as well as personnel turnover and organizational shifts, identifying all public datasets at a given agency is a constantly moving target. The discovery of new datasets through this year's E&V process does not mark delinquency on the part of any Open Data Coordinator. Rather, it highlights the need for routine assessment to ensure that all business areas, data systems, and communications channels are synchronized with open data publishing requirements.

These insights are helping MODA build additional mechanisms to improve citywide compliance with the Open Data Law. For the annual report published last summer, all agencies were required to submit an abridged E&V Dataset Questionnaire. This year, we will look to make additional improvements to the reporting process for all City agencies.

The strength of the New York City Open Data program is that it touches every City agency and engages hundreds of thousands of New Yorkers each year. Our ongoing partnership with the Open Data Coordinators and other Open Data stewards at participating agencies is helping MODA build a more transparent City government. We wish to thank them for their full cooperation in this process and for continuing the good work to identify, prepare, and publish valuable datasets to the Open Data Portal.

Finally, before concluding, I'd like to state for the record that the City of New York is very lucky to have such a talented, hardworking, and dedicated Open Data team. As the Committee is aware, I have been serving in an interim capacity at MODA for just a few months. Within that time, I have been deeply impressed by the expertise and ethic of this group, and I know New Yorkers are better off for having them on this team.

At this point, we welcome the Committee's questions on the Examination and Verification process and results.

###

Recommendations for Better Citywide Compliance

The Mayor's Office should regularly communicate the Administration's ongoing commitment to ensuring the success of Open Data – not just as a transparency goal, but also as a data governance policy with well-documented return on investment.

MODA should develop internal awareness campaigns to assist Open Data Coordinators in making more agency staff aware of the benefits of Open Data.

MODA should continue to demonstrate the value of open data for data governance and analytics use cases through its Open Source Analytics Project Library and additional channels.

MODA should provide proactive guidance to Agency commissioners on ODC selection.

The Open Data Team should help Open Data Coordinators craft Open Data strategies unique to their agencies and provide additional documentation on open data organizational structures, return on investment, and compliance best practices.

MODA and **DoITT** should align Open Data reporting with other reporting requirements, including performance reporting and public records requests, to build on existing resource allocations for process efficiencies.

When an ODC leaves their position, **MODA** and **DOITT** should provide proactive support to assist in handoff to ensure smooth transition and knowledge transfer.

MODA and **DOITT** should investigate tools for metadata documentation and management.

Agencies should assess demand for public information across multiple public-facing touch points – including dataset requests, public records requests, and 311 service requests – and work with **MODA** to identify ways in which they can proactively engage public stakeholders through Open Data.

Agencies should have internal, cross-functional data governance teams that regularly convene and have Open Data on the agenda.

Testimony

**John Kaehny, Executive Director, Reinvent Albany
Member, New York Transparency Working Group**

NYC Council Committee on Technology Hearing on Open Data Audit Per Local Law 8 of 2016

January 24, 2017

Good Morning, Chairman Vacca and Members of the Technology Committee, I am John Kaehny, Executive Director of Reinvent Albany and Co-Chair of the NYC Transparency Working Group. Thank you for holding this oversight hearing and continuing to push the administration and its agencies to fulfill the promise and requirements of the Open Data Law.

This hearing is specifically about the Mayor's Office of Data Analytics, MODA, audit of the Department of Buildings, Department of Environment and Fire Department. The audit's larger purpose is to keep pushing open data forward by learning what is working and not working at some of the City's largest agencies.

Let me speak to that larger purpose for a moment. The City's open data program had a rough time making the transition from the Bloomberg to de Blasio administration and it took groups like ours making a ruckus with City Hall to get things back on track. We are again concerned about staffing and resources for open data.

Seven months ago, Rafael Carvalho, one of two senior open data managers at DOITT, left for another agency. He has not been replaced, and as far as we can tell there is no hiring process underway. Our take is that DOITT's inability or disinterest in replacing

this key open data position has slowed and undermined the Citywide open data effort. The person in this open data manager position is one of two people at DOITT who have the seniority to effectively work with, cajole, and massage agencies into getting their data published in a useful form. We ask that Council ask DOITT to fill this important, funded, position as soon as possible --- especially given the likelihood of citywide hiring freezes in the next budget.

The Audits

I have a number of general comments about the open data audits, but would note that we have had two business days to review the agency worksheets.

Broadly, we think the audits --- if done in the spirit they were intended --- can be extremely valuable tools for prodding specific agencies and for identifying systemic problems with open data implementation.

This said, MODA's audit is not nearly as useful as it could be. The audit report needs to include a concise summary of how well each audited agency is doing, what works well, what needs improvement and how each agency is going to meet its open data mandates. Without this, the audit is missing the forest for the trees and it is hard for even the informed public to know how healthy the open data forest is at the audited agencies. Put another way, without a lot of time and work, it's hard to tell which agencies are top open data performers and which are laggards.

This said, we think that this year's audit shows that the Department of Buildings is serious about open data and the Fire Department is not. We were impressed by DOB's detailed responses in the various audit worksheets, the number of datasets published and fact that vast majority of datasets have standard geocoding and data dictionaries.

In contrast, the Fire Department seems to have few datasets, lacks geocoding and data dictionaries and included sparse details about plans for improvement for each dataset, doesn't use the same data internally as it publishes and did not bother filling out a key worksheet which details which legally mandated reports are based on public data.

Like BetaNYC, we would also note that the Fire Department rejected a Department of Health request to publish more information about EMS dispatch data based on what we believe to be an incorrect assertion of privacy. (EMS dispatch data is available as open data in other cities, notably Seattle.) If our assessment of FDNY is unfair, we would like to know why.

Specific Questions from Audit

- What does MODA mean when it says that the term "dataset" is "loose." (page 4/17 Examination Report.)
- Related, which dataset does MODA believe agencies should be putting on the open data portal? Is this something that needs to be clarified in law?
- Did the audits find any examples of their own or other agencies using their agency data more effectively thanks to open data? If not, are they asserting that their data was easily accessible within their own agency and from other agencies?
- What did the audits tell MODA about how open data was making it easier for people within or at other agencies to access their data more easily?
- Do these agencies seem to think that open data helps them fulfill their mission?
- Do these agencies think open data has reduced their FOIL workload?
- Has MODA considered interviewing agency stakeholder groups when doing the audit to find out what they think?

To: NYC Council - Committee on Technology
From: Noel Hidalgo, Executive Director of BetaNYC
Re: **Open Data Examination and Verification Report 2017**



Wednesday, 6 December 2017

Good Afternoon, Chairman Vacca and Members of the Technology Committee, I am Noel Hidalgo, Executive Director of BetaNYC and a member of the NYC Transparency Working Group.

I want to start by thanking Chairman Vacca, the members of this Committee, and the Council for your continued commitment to oversight hearings for the Open Data Law. Your ongoing energetic support for Open Data has made New York City a global leader in open data and is hugely encouraging to open data advocates inside and outside of government.

BetaNYC wants to thank this Council and this Committee for being amazing partners. We hope for the next four years, we can find collaborators who are as willing this current committee.

Second, I want to thank the City's open data team for continuing to slog on with a notable loss of staff. It is a shame that the City's open data team isn't up to its full staff complement.

General Observations:

It is clear that this report is shaped by a lack of resources. Several of last year's recommendations are not mentioned in this report and don't seem to carry over. Specifically, 2016's 1i, 1ii, 1iii, and 3i, 3ii, and 3iii.

1. *Agencies should make their technical ecosystems more accommodating to Open Data by:*
 - i. *Using automations, rather than manual uploads, to update datasets currently on the Open Data Portal.*
 - ii. *Writing Open Data requirements into procurements of new data systems and analytics technologies.*
 - iii. *Allocating more resources to Open Data personnel , especially Open Data Coordinators .*
3. *MODA should improve the Examination and Verification plan for future years by:*
 - i. *Consulting with the Department of Investigation on potential improvements .*
 - ii. *Creating clear guidelines and definitions of "data" and "dataset. "*
 - iii. *Creating clear guidelines on determining whether a dataset is "public" or "private ."*

With these recommendations, what happened?

Last year's report introduced a results snapshot. At high level, we could see agencies and their released data sets, soon to be released datasets, the status of automation, their update frequency, data related to the MMR, and public requests against these agencies.

As these reports moves forward, we would love for MODA to standardize a summary data table. Also, we recommend that the Council ask for a detailed snapshot to be included in every annual report.

BetaNYC and some public members were fortunate to get copies of accompanying spreadsheets. Frustratingly, these spreadsheets burry information that should within the report. Both the DEP and FDNY will archive a significant amount of existing data sets yet the 2017 report says that the agency identified a number of datasets that don't constitute a "public dataset" and they will be removed to improve the data portal's hygiene.

It is important for the public summary have more detail. The public shouldn't have to pour through spreadsheets to understand this audit.

Specifically to this audit, several questions that have come up.

1. Why is the FDNY's D. Reports is blank of any legally mandated reports? How was that approved?
2. If these reports were put in-place during the summer, why will it take DEP over a year to release two data sets? Shouldn't these compliance reports outline a path to compliance with some sort of compliance along the way?
3. What is MODA's "improved" examination and verification process?
4. How has it improved open data coordinators work?
5. What is being done with in these three agencies to move data and open data from insight to action?

Lastly, neither 2016 nor 2017 are easily accessible via MODA's website nor the open data portal. All reports and accompanying supporting documents should be accessible via MODA's website and the open data portal.

Recommendations on observations:

In 2016, in partnership with the NYC Parks Department, we outlined a data release workflow that benefits users and producers. Through this process we were able to make the 2015 TreesCount! data set one of the most accessible datasets on the city's data portal. This process outlined three phases.

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.

This "user centered data release" helped the Parks Department navigate three things: (1) collectively, we were able to understand data users and ensure data was accessible to them; (2) we were able to identify data quality problems; (3) we demonstrated value in and out of the agency. It is great to that this examination and verification report validates these ideas. Sadly, the report doesn't speak to this research.¹

When agencies are going to release data, there must be some sort of public and private stakeholder convening. For this to happen, the open data team and agencies need resources for engagement.

For the past two years, we have worked with the Manhattan Borough President's Office and along side Manhattan Community Boards. The open data team and MODA have supported our efforts. We can attest to this report's recommendations on better citywide compliance.

- Each agency is unique.
- Executive buy-in is crucial.
- Open data is an opportunity to teach users how City operations work.
- Open data drives data governance and analysis.
- People who use open data must be well-networked and trust each other.
- Organizational knowledge retention is a challenge.

It is great that this report outlines that open data drives governance and analysis, but there needs to be considerable support in moving data from awareness to action.

This open data examination and verification report is a testimony to slow culture change. Five years into the open data law, we still have agencies that still don't know how to review their datasets nor are prioritizing data sharing.

It is crazy to see the DEP say they will release data sets in over a year with no intermediary datasets. Additionally, it is crazy to read that the FDNY received the following public comment request from the Department of Health.

"Hello! I work for the NYC Health Department (DOHMH) and we are interested in using the EMS Dispatch Data that is provided on NYC Open Data. However, we would really like more detail to the data, if possible. Who should we contact to discuss this? Thank you!"

¹ <https://beta.nyc/2016/08/05/treescount-data-jam-2016-report-back/>

According to the accompanying FDNY spreadsheet, this response was denied. Did FDNY use this opportunity to engage with the DOHMH? What happened to such a promising opportunity for collaboration? Did MODA come in and help data sharing?

If we ever expect these practices to change within the next ten years, the Open Data team and its partners need to be taken seriously. Seriousness comes with resources and staff. From my understanding, MODA and the open data team need to be staffed up to its full complement and get a budget for internal and external engagement. We need this administration to hire a Chief Analytics Officer and to ensure that DOITT hires a replacement for Ralf Carvalho.

We love working with the open data team. We will donate as much time and energy as we can. With support from the Alfred P. Sloan Foundation, Fund for the City of New York, and Data & Society Research Institute, BetaNYC has been able to scrape up funds to support mutually beneficial programing. **We are always looking to combine MODA's objectives into our work and are very lucky to have them as a trusted partner. Our funding isn't permanent, yet we see a permanent need for our collective work.**

To be perfectly frank, the Mayor's office needs to dedicate financial resources to MODA and its partners. With all the things MODA is called on to do, it is amazing they were able to produce this report.

Now is the time for Council to work with MODA, DOITT, and their partners to properly resource the city's open data program. There needs to be money for internal education, internal/external engagement, and concrete roadmaps that link technology improvements to better data systems. MODA and the open data program needs a budget for programs that support internal change.

Thank you,

Noel Hidalgo

Observations from the three reporting spreadsheets:

DEP

- Number of existing DEP data sets - 48
- Number of datasets that say they have a data dictionary - 8
- Note 1 - As of 5 Dec, 22 listed datasets were not accessible. The reporting spreadsheet should indicate if these are to be archived. There is no listing of where this data will live.
- Note 2 - Two datasets are to be released on 31 Dec 2018. If this report was published in the summer of 2017, why isn't a preliminary release now?

FDNY

- Number of existing FDNY data sets - 18
- Number of FDNY data sets that will be archived - 11
- Number remaining - 7
- Number of remaining data sets that have data dictionaries - 3
- Time to get data dictionaries - unknown
- Note 1 - FDNY indicates which datasets will be archived.
- Note 2 - FDNY's D. Reports is blank of any legally mandated reports. This is odd.

DOB

- Number of existing DOB data sets on the data portal - 20
- Number of datasets that say they have a data dictionary - 20
- Note 1 - Number of times "We do not have the resources to publish this to the Open Data portal at this time." - 5
- Note 2 - This report uses terms that aren't explained. For example, what is an ibot?

**THE COUNCIL
THE CITY OF NEW YORK**

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☐ in favor ☐ in opposition

Date: 12/06/17

(PLEASE PRINT)

Name: John Kaehny

Address: _____

I represent: Reinvent Albany

Address: 148 Lafayette NY 10013

**THE COUNCIL
THE CITY OF NEW YORK**

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☒ in favor ☐ in opposition

Date: _____

(PLEASE PRINT)

Name: Noel Hidalgo

Address: _____

I represent: BETA NYC

Address: _____

**THE COUNCIL
THE CITY OF NEW YORK**

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☐ in favor ☐ in opposition

Date: 12/6/17

(PLEASE PRINT)

Name: James Perazzo

Address: 253 Broadway

I represent: MODA

Address: _____

**THE COUNCIL
THE CITY OF NEW YORK**

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☐ in favor ☐ in opposition

Date: 12/6/17

(PLEASE PRINT)

Name: Craig Campbell

Address: _____

I represent: MOCA

Address: _____

Please complete this card and return to the Sergeant-at-Arms

**THE COUNCIL
THE CITY OF NEW YORK**

Appearance Card

I intend to appear and speak on Int. No. _____ Res. No. _____

☐ in favor ☐ in opposition

Date: 12/6/17

(PLEASE PRINT)

Name: Alfred Webber

Address: 2 MetroTech Ctr

I represent: D-ITT

Address: _____

Please complete this card and return to the Sergeant-at-Arms