Dr. Amen Ra Mashariki, Chief Analytics Officer

TESTIMONY BEFORE THE CITY COUNCIL COMMITTEE ON TECHNOLOGY

OVERSIGHT ON MODA'S 2016 OPEN DATA EXAMINATION AND VERIFICATION REPORT

TUESDAY, JANUARY 24, 2017

I. BACKGROUND

Good morning. My name is Dr. Amen Ra Mashariki, New York City Chief Analytics Officer, Chief Open Platform Officer, and Director of the Mayor's Office of Data Analytics (MODA). I'm here to testify on MODA's implementation Local Law 8 of 2016, the Examination and Verifications Law.

Local Law 8 of 2016

Local Law 11 of 2012 (the Open Data Law) requires City agencies to self-submit annual compliance plans that lay out a schedule for publishing all of their "public datasets" by 2019. In the first three years of the Open Data program, this method resulted in the publication of more than 1,500 datasets from 60+ City agencies.

Advocates, the de Blasio Administration, and City Council – steered by the leadership of the Council Technology Committee and Committee Chair Vacca – felt we could do even better, and in late 2015 and early 2016 collaborated on several new laws designed to provide additional mechanisms to ensure the publication of every eligible dataset.

One of these laws – Local Law 8 of 2016 – was a new "examination and verification" process. Its purpose is to improve citywide compliance by creating a more systematic way to locate datasets that may have inadvertently or purposefully been excluded in agencies' self-reported Open Data plans.

In January 2016, Mayor de Blasio designated MODA as the office responsible for carrying out the examinations and verifications. As required by the law, MODA then created a plan for conducting the process, which was approved Department of Investigations Commissioner Mark Peters in March 2016.

I would like to express, on behalf of MODA and the Open Data program, our gratitude for the opportunity today to speak to the quality of the 2016 Examination and Verification Report, our plans for improving the process, and our roadmap for better citywide compliance with the Open Data Law.

MODA Plan

Our examination and verification plan consisted of three parts.

The first was a dataset questionnaire intended to familiarize MODA with each agency's current Open Data footprint, routine information reporting requirements, data management systems, and organizational structure. These questions served as prompts for the Open Data Coordinators at the surveyed agencies to verify that they have identified and reported all eligible datasets in the agency's highest priority information streams – including FOIL disclosures, data supporting key performance indicators, and data included in reports required by legislation.

The second part was a certification letter modeled after the letter all Agency commissioners signed when they submitted their first Open Data compliance plans in 2013. In this statement, an agency commissioner or their designee warrantied the accuracy and completeness of the information provided in the dataset questionnaire.

Finally, while users always have the option to nominate datasets for publication on the Open Data Portal, a specific window for public feedback specifically invited the public to participate in the Examination and Verification process. During the first two weeks of November, we invited feedback via social media and alerts on the Open Data Portal.

II. E&V REPORT

As specified by the Local Law 8, the Department of Correction (DOC), Department of Sanitation (DSNY), and Department of Housing Preservation & Development (HPD) were included in the 2016 Examination and Verification cycle. MODA submitted its findings on these three agencies to the Mayor, City Council, the surveyed agencies, and the public on December 1, 2016.

In summary, we found that:
- All three agencies are in good standing with the Open Data Law.
- Several data systems *may* have contained eligible datasets, but warranted further review.
- Three dataset nominations were received during the public feedback window, all of which were referred to HPD. One was further referred to the Department of Finance, and another was determined to not be an existing dataset. The third was a request for daily updates to HPD's emergency repair charges dataset, which was reported as "under further review" but since been added to HPD's Open Data compliance plan.

All three agencies cooperated fully and provided valuable feedback during our inaugural round of the Examination and Verification. We would in particular like to thank the three agencies' Open Data Coordinators, along with Commissioners Been, Garcia, and Ponte, for being forthcoming with our requests and ensuring that this process ran smoothly and on-time.

The law also requires MODA to report any deviations to the plan we submitted to DOI. While we followed the DOI-approved plan precisely, we did encounter challenges carrying out of "verification" workflow that we had intended (**Fig. 1**). We explain these difficulties in detail in the report.
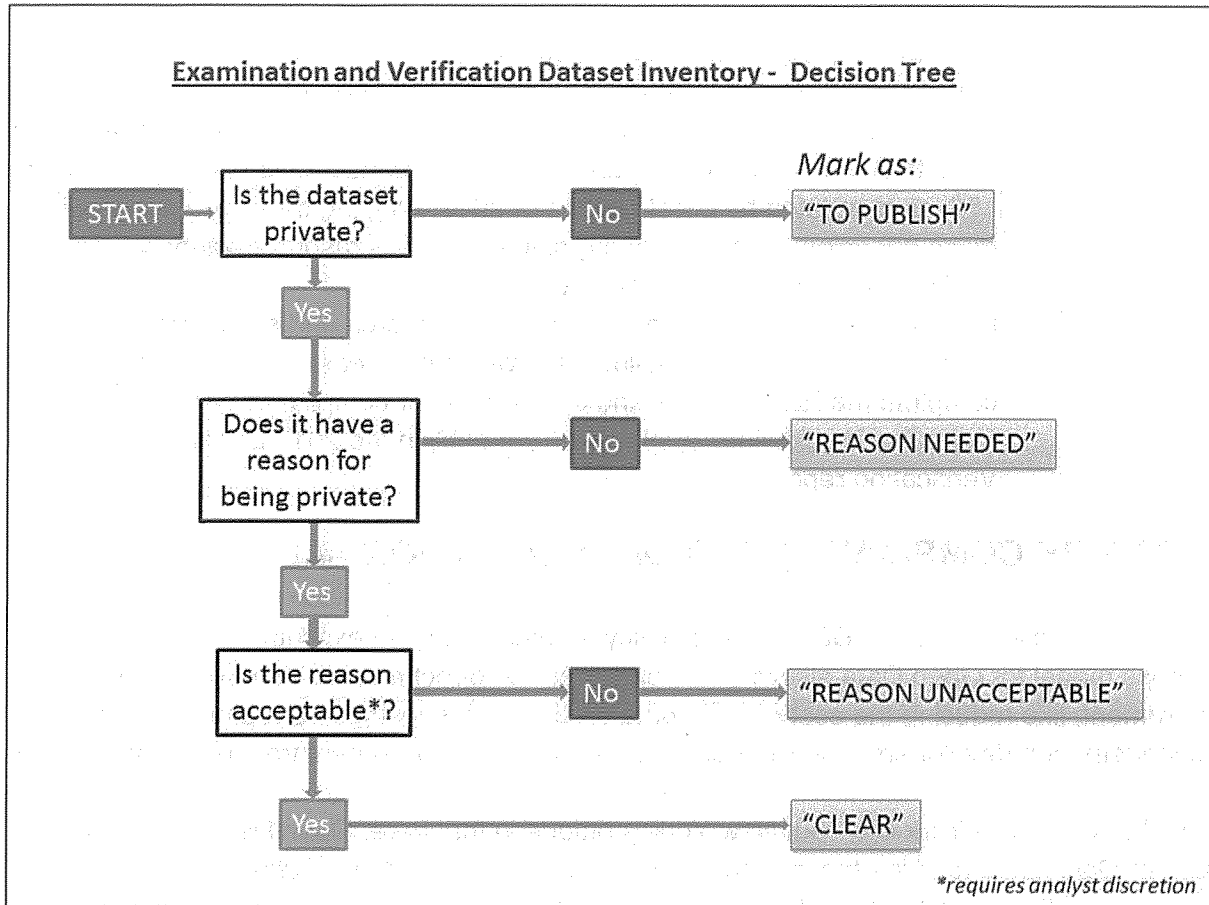
Examination and Verification Dataset Inventory - Decision Tree

START → Is the dataset private? — No → *Mark as:* "TO PUBLISH"

Is the dataset private? — Yes ↓

Does it have a reason for being private? — No → "REASON NEEDED"

Does it have a reason for being private? — Yes ↓

Is the reason acceptable*? — No → "REASON UNACCEPTABLE"

Is the reason acceptable*? — Yes → "CLEAR"

*requires analyst discretion

Figure 1

These challenges included issues determining the eligibility of raw data that may have never before been extracted as a 'dataset' from its transactional system, and conceptual questions on the scope of "public datasets" that were not immediately resolved by the definition provided in Local Law 11 of 2012.

In the report, we propose ways to meet these challenges in our "recommendations on better citywide compliance." In particular, we recommended that:

1. **Agencies should look for ways to ensure that newly collected data can easily be published as Open Data.** For example, the contract procuring CitiBike's technical infrastructure included a provision mandating that the data collected be made public. Because this Open Data requirement was built into the technology itself, the ride data published by CitiBike is easy to access, easy to use, and in high demand.

2. **The Open Data team should take more steps to empower Open Data Coordinators.** Earlier this month, we surveyed all of the Open Data Coordinators to better understand their roles within their agencies, the challenges they face, and opportunities for improved coordination. The results are informing our ODC-

engagement and recruitment strategy, which begins with reconvening all the Open Data Coordinators in City Hall in mid-February.

3. **MODA should improve the Examination and Verification plan.**

    i.   For the 2017 cycle, the E&V cycle will coincide with the Open Data reporting cycle that culminates in the publication of the annual Open Data plan on July 15th, as mandated by Local Law 11 of 2012. Aligning these processes will make agencies' overall reporting obligations less resource-intensive, and the MODA verification process more efficient.

    ii.   In the meantime, we are conducting peer research and stakeholder interviews to produce guidelines that will make it easier for agencies to complete the E&V process and self-verify their compliance. We intend to make these guidelines public when we publish the 2017 Examination and Verification report.

III. FUTURE COMPLIANCE AND *OPEN DATA FOR ALL*

In many municipalities, Open Data is an IT policy. In others, it is an executive order. The business model for Open Data elsewhere is based on the direct returns it provides to City government, like reducing the cost of FOIL data disclosures, serving as a resource to internal and industry app developers, and representing a gesture of government transparency.

All hold true for New York City. But New York is unique in that here, Open Data is the *law*. Our vision of *Open Data for All* -- the idea that *Open Data belongs to New Yorkers* -- naturally followed from the fact that elected officials, via the constituents they represent, were so committed to this idea that they added it to the NYC administrative code.

We believe that the fundamental promise of Open Data is that it avails public information that gives people the knowledge to take action on their own behalf. As a policy, it is similar to the Freedom of Information Law, which guarantees that documents and information maintained by government be disclosed when requested by the public. But while FOIL is intended to deliver precise answers to specific questions, or disclose documents in a fixed form, data is more liquid.

It is, for this reason, that in addition to the report, we included itemized results of the agency dataset inventory surveys in the attached "Examination and Verification 2016 Results" excel workbook. This file is meant to provide users with a snapshot of the technical ecosystem of the agency and a better understanding of how data gets from an agency data source to a usable dataset on the Open Data Portal.

We encourage users to take this information to request new datasets for publication with the dataset nomination feature -- which guarantees a formal review and timely response by Local Law 109 of 2015 (**Fig. 2**).
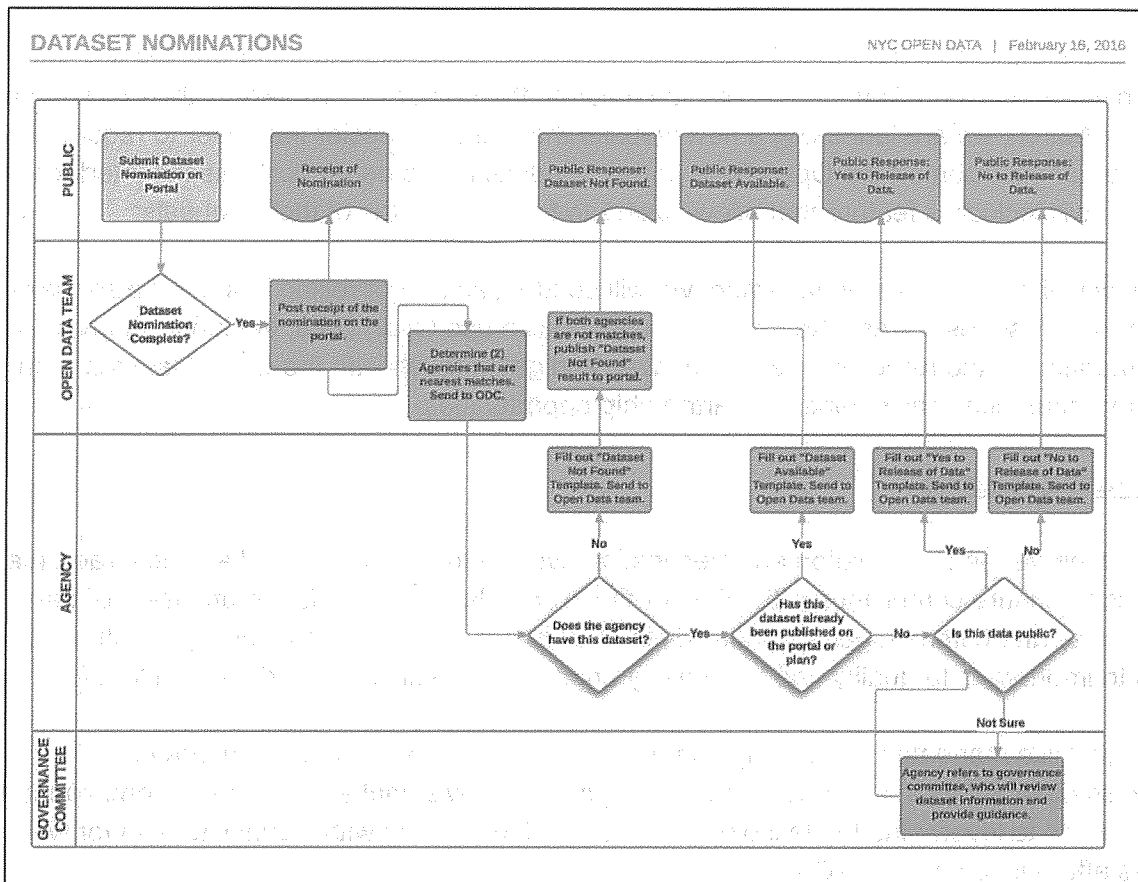
Figure 2

These requests are helpful in supplying parameters that guide decisions on how to represent data that may previously not have had any bound and fixed form. They also signal public interest in an agency's data, which according to the prioritization parameters named in Local Law 11 of 2012[1], instantly escalates the urgency of making a determination on how to publish an appropriate dataset.

A demonstration of public interest is crucial to the ongoing success of the Open Data initiative and its compliance with all the associated laws, as it makes agencies more aware of the data they have and maintain. This, in turn, spurs better upkeep and disclosure to the public – and the cycle continues.

---

[1] As stated in Local Law 11 of 2012: agencies should prioritize datasets according to whether the information:

    (1) Can be used to increase agency accountability and responsiveness;
    (2) Improves public knowledge of the agency and its operations;
    (3) Furthers the mission of the agency;
    (4) Creates economic opportunity; or
    (5) *Responds to a need or demand identified by public consultation.*

New website

It is also incumbent on the Open Data team to cultivate public demand. For this reason, when we join the NYC Open Data community to celebrate the 5-year anniversary of the Open Data Law this March, we are launching a new website that makes it easier than ever to request new datasets for inclusion on the Open Data portal. In addition to a better nominations interface, the new site will include a feature that allows users to track the status of their requests.

The new website, and the engagement we will conduct as part of rolling it out, will also make it easier to access resources to learn more about how to use Open Data, track our progress on implementing all the relevant Open Data laws, and get in touch with the Open Data team on anything from technical inquiries to partnership opportunities.

Concluding remarks

As we improve the Examination and Verification over the next year, we will also leverage the process to ensure compliance with all other Open Data laws, including the provision of data dictionaries and common geospatial fields in all datasets, as well as advancing our strategic goals to improve data quality and proactively engage the public on new dataset releases.

The legislative mandates, increasing user base, and growth of open data culture in City agencies have put us in a strong place moving into the law's final years of initial compliance. We look forward to leveraging the lessons we've learned in working with Council to plan for what comes after December 31, 2018.

Thank you.

Testimony of
Dominic Mauro,
Staff Attorney, Reinvent Albany
Member, New York Transparency Working Group

NYC Council Committee on Technology
Hearing on First Ever Open Data Audit Per Local 8 of 2016
January 24, 2017

Good Morning, Chairman Vacca and Members of the Technology Committee, I am Dominic Mauro, Staff Attorney of Reinvent Albany and a member of the NYC Transparency Working Group.

I want to start by thanking you 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.

Also, our sincere thanks to Mindy Tarlow, the Director of the Mayor's Office of Operations, and DoITT Commissioner Anne Roest, who have helped staff up the City's Open Data Team, and have dedicated more time to open data issues. We also thank the open data audit team at MODA and DoITT for their earnest and professional work carrying out this first-ever open data audit. We are extremely pleased to see the administration comply with Local Law 8 of 2016 in a timely and serious way.

We have three comments on the agency open data audits. First, the administration's Open Data Team exceeded our expectations and gathered and shared with the public a great deal of useful insights.

Second—and we find this odd given the overall high quality of the audit—the Open Data Team declared all three agencies in compliance with

the Open Data Law. But the evidence they gathered raise questions about whether the agencies are complying with the Open Data Law.

The audits found twenty public data sets which the three agencies have not scheduled for publication on the Open Data Portal. The Open Data Law requires all public data sets be published by the end of 2018. The Department of Sanitation should be considered out of compliance with the Open Data Law until it puts the fourteen public data sets on a schedule for publishing before the end of 2018, and the same goes for the six public data sets identified at Corrections and HPD.

Third, and more positively, the Open Data Team lists a series of forward-looking recommendations on page five. We strongly endorse all eight of these specific recommendations, and hope that city council and public stakeholders are invited to engage in the process of implementing them.

We have additional written testimony which I will summarize.

In fulfillment of Local Law 8 of 2016, the Department of Investigation delegated to the combined MODA-DOITT Open Data Team, the task of auditing the Departments of Sanitation, Correction, and Housing Preservation and Development's compliance with the Open Data Law.

The Open Data team's audit was thoughtful and included a number of useful features:

- It describes the data sets which are used to calculate each MMR indicator for over a hundred indicators;

- It inventories each agency's "Technical Systems with more than 20 users," organized by agency program;

- It lists the agency personnel consulted for their expertise in their respective agencies' data assets; and

- It examined agency FOIL logs for repeated requests for public data sets, although agencies did not identify any data sets to be published.

However, we have serious concerns about two parts of the report. First, as mentioned above, we do not understand how the agencies can be considered in compliance with the Open Data Law when they have no plan to publish the twenty public data sets identified by the Open Data Team. According to the Open Data Law, a data set is either public or not public. The Open Data Team and the agency have to decide, and they have to explain why a public data set is not a part of an agency's compliance plan.

The Open Data Team explains that these twenty public data sets are clearly public or clearly private: they are "less definitive" and they "require further investigation." (footnote page 4.) But, the main purpose of the Local Law 8 audit is to tell the world how many public data sets an agency has published, has scheduled for publishing, and how many public data sets have not but should be scheduled for publishing. The audit raises concerns by failing to classify these twenty data sets.

Second, there should not be confusion about what a public data set is. The Open Data Law defines a public data set as a "comprehensive collection of interrelated data that is available for inspection by the public in accordance with any provision of law and is maintained on a computer system by, or on behalf of, an agency."

In other words, if a dataset is (wholly or partially) a public record subject to disclosure under the state Freedom of Information Law or is already shared on an agency website in another form, it is a "public data set" and should be on the Open Data Portal. The Open Data Team's apparent confusion about the definitions of the terms "public," "data," and "dataset" are alarming and the administration needs to work with Council and stakeholders to clarify and resolve these definitional questions or the Open Data Law cannot work. (Page 2, paragraph 8.)

Thank you the opportunity to testify.

THE COUNCIL
THE CITY OF NEW YORK

$$\boxed{\textit{Appearance Card}}$$

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

☐ in favor ☐ in opposition

Date: _____

(PLEASE PRINT)

Name: Craig Campbell

Address:

I represent: MODA

Address:

THE COUNCIL
THE CITY OF NEW YORK

$$\boxed{\textit{Appearance Card}}$$

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

☐ in favor ☐ in opposition

Date: _____

(PLEASE PRINT)

Name: Amen Ra Mashariki

Address:

I represent: MODA

Address:

THE COUNCIL
THE CITY OF NEW YORK

$$\boxed{\textit{Appearance Card}}$$

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

☐ in favor ☐ in opposition

Date: _____

(PLEASE PRINT)

Name: Dominic Mauro

Address:

I represent: Reinvent Albany

Address:

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