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.