

Testimony of
John Kaehny, Co-Chair NYC Transparency Working
Group *and* Executive Director, Reinvent Albany
and
Gene Russianoff, Co-Chair NYC Transparency Work-
ing Group *and* Senior Staff Attorney, NYPIRG
before the
New York City Council Committee on Technology
Hearing on Open Data on October 1, 2015

Good afternoon Chairman Vacca and Members of the Technology Committee, I am John Kaehny, co-chair of the New York City Transparency Working Group and Executive Director of Reinvent Albany. I am also presenting this testimony on behalf of my Transparency Working Group Co-Chair and NYPIRG Senior Attorney, Gene Russianoff, who was unable to attend today.

Thank you Chair Vacca for holding this important oversight hearing – your questions today were right on target! Thank you also to you and your colleagues, and the central staff, for developing new legislation to kick-start improvements to the Open Data Law. We strongly support the intent of the seven new bills and have prepared extensive written comments.

This said, we are here today to urge the Council and the Mayor to aim high and to use this momentum for change to make major fixes in the Open Data Law. I was fortunate to be one of the public stakeholders who helped negotiate the landmark Open Data Law in 2012. When the Open Data Law was enacted, it was widely considered the best in the world. Yes, the world. And indeed, the law has broadly succeeded in many ways and the City's open data efforts have made tremendous progress. But three years on, we are seeing serious problems with how data is

made available, the usefulness and usability of data, and the quality of data. Despite sincere efforts by MODA and DOITT, many agencies are not complying fully or usefully with the law, and there is widespread frustration among data users, civic technologists and watchdog groups.

Data Availability and Private Right of Action

Most importantly, the public data being published by city agencies is not necessarily the public data that the public wants. Unfortunately, the Open Data law has not made it easier for the public to get agencies to release data. When agencies refuse to publish public data, neighborhood and civic groups have two options: they can try to get council to pass a new reporting law or file a Freedom of Information Law request.

The City Council has passed ad hoc data reporting bills on taxi crashes, bike crashes, youth and foster care, hepatitis, student demographics and so on. But Council cannot pass thousands of bills requiring agencies to post specific data sets.

We believe the Open Data Law should be like the Freedom of Information Law, and the public should be able petition a judge to make agencies comply with the Law. Consider this: according to Legistar, since 1998 the Council has passed 398 laws requiring agencies to report on a specific activity. Only five of those 398 laws deny the public the right to go to court to see the law enforced – one of those is the Open Data Law.

The Open Data Law and the Freedom of Information Law are really two sides of the same open government coin. Everyone agrees that if the public could not go to court to get records disclosed—what’s called a “private right of action”—the Freedom of Information Law would not work. Sometimes agencies have to be compelled to disclose records requested via FOIL. Similarly, given our experience of the last three years, we think some agencies will not release certain public data unless they are compelled to. Since October 2011, the Open Data Portal has received 185 requests to publish data, DOITT has approved six, rejected seven and not resolved 170. This is not encouraging.

Data Quality

The city's open data program has two other major problems. Most vexing is that there is no useful way for the public to report errors, receive responses, or track the fixes made to published data. Open data experts fully expect government data to contain errors --- sometimes big errors. One of the basic ideas behind open data is that the public will act as free "data detectives" who will find data problems so the city can fix them. This is why the City's open data platform has a public comment tool. Unfortunately, the City has responded to only a handful of comments on the platform, and often goes months without answering any. The City needs to staff the Open Data Platform and fix reported data problems.

Data Usefulness

Our groups have devoted much time and testimony to improving the overall usefulness of the Open Data Platform and data sets. Most broadly, the city needs to start using the same data the public does. It appears to us that the City has three different versions of some datasets: one for the agency, one for the Mayor's "Data Bridge" system, and one for the public Open Data Portal. There should be one high-quality data set for everyone.

Conclusion

We are very proud of the pioneering work New York City has done on open data. The city, the council and public stakeholders have all learned an enormous amount in the last three years. To us, the main lesson is that the Open Data Law will never truly succeed unless the public is empowered by a private right of action. Opening government is always hard, but it becomes close to impossible when the public cannot get the law enforced. We need a private right of action, and we hope the Council and the mayor will act strongly to once again make NYC the global leader in open data.

Additional Written Testimony of Reinvent Albany, co-chair of the NYC Transparency Working Group

Issues With NYC's Open Data Law & Implementation

Based on the experience of the last three years, we see three fundamental problems with the implementation of the NYC Open Data Law.

There is no legal mechanism to compel agencies to publish public data or to meet their own data publishing schedules. Agency compliance is essentially voluntary, and DoITT has failed to coordinate the release of high-value data sets in a timely fashion.

Many open data sets contain errors – sometimes serious errors – but the City has responded to only a handful of comments about these errors on the portal. The public needs a way to report errors, receive responses from agencies, and track the fixes made to the data sets. Currently, there isn't one.

There are glaring data quality issues, but only on the public open data portal; Data Bridge has highly-formatted and cleaned data. The tale of two data cities is unnecessary: MODA should provide Data Bridge data for everyone, and each data set should include community district, council district, police precinct, and so on. This would make the data more usable by ordinary members of the public as well as technical experts.

The council has introduced seven open data bills before this committee, Appendix A contains specific comments on each bill, but we have a series of recommendations which directly address the three problems we see with open data.

Recommended Solutions for Open Data Problems

I. Restore the Private Right of Action

Our first and most obvious recommendation is to allow members of the public to file Article 78 petitions against agencies which fail to comply with the open data law. This is the same process used in Freedom of Information Law requests when

agencies deny access to public records. Without it, FOIL would be completely stymied, instead of being the cornerstone of government transparency for the last four decades. The right to petition agencies to obey the law should be restored to the open data law.

Article 78 of the New York Civil Practice Law and Rules allows members of the public to petition a judge for an injunction against an agency's action or inaction. By default, every agency required to act (or forbidden to act) in a certain way can be the subject of an Article 78 petition for failing to uphold the law. However, laws may take away that right on a selective basis: for example, NYC's Local Law 46 of 2010 requires the Law Department and Mayor's Office to review each rule before it is promulgated in the City Record, but members of the public can't file Article 78 petitions for either office's failure to review rules.

It's extremely rare for New York City laws to remove this right from the public when passing new laws. In fact, of the 1,414 laws enacted by the New York City Council since 1998, only a handful have removed the private right of action. Local Law 11 of 2012, the Open Data Law, is one of those laws; members of the public cannot enjoin agencies for failing to publish data sets on the open data portal.

Since 1998, the New York City Council has enacted 398 laws which require an agency, office, board, or department to publicly file reports with the council. Like the Open Data Law, these laws are meant to bring transparency to government activities. Virtually all of these laws allow for Article 78 petitions to enforce their provisions, but the open data law is one of five such "reporting laws" passed since 1998 to block them. The other laws require agencies to report on highly politicized topics:

1. Local Law 62 of 2011: Requires the Department of Correction to report specific information about immigrants detained each year by request of the federal government.
2. Local Law 21 of 2013: As LL 62 of 2011, but requires the NYPD to report the same statistics.
3. Local Law 85 of 2013: requires the Chief Medical Examiner's office to report to city council on systemic errors in the course of performing its ordinary du-

ties. A companion bill, LL 86 of 2013, requires the CME's office to report on the number of employees who pass an occupational proficiency test each year.

4. Local Law 140 of 2013: Requires the city to create a website to track the expenditure of federal disaster relief funds in the wake of Hurricane Sandy.

New York City's Open Data Law requires agencies to put public records containing tabular data on the open data portal. Every single record on the portal is already subject to the Freedom of Information Law. The Open Data Law is not any more politically sensitive than FOIL. The other laws were passed after scandals and a historic tropical storm. This Council passed the Chief Medical Examiner's reporting law after the CME's office mishandled DNA evidence required the review of over 800 rape cases. Likewise, the Department of Correction and NYPD reporting laws were passed in response to fears of unconstitutional overreach by federal immigration authorities.

The open data law does not provide any more politically sensitive information than the Freedom of Information Law, but FOIL provides a private right of action for ignored requests while the open data law does not. FOIL simply wouldn't work if agencies knew they could ignore requests with immunity to Article 78 petitions. This is our experience with the Open Data Law to date, and this should be the highest priority for this council.

2. Shift Ownership of Open Data Law from DoITT to the Mayor's Office

We also highly recommend that the responsibility for implementing the Open Data Law be shifted to the Mayor's Office. In the absence of leadership from DoITT, the Mayor's Office of Data Analytics has assumed the role of creating a working group of agencies' Open Data Coordinators. MODA has done an excellent job stepping into this role, and we applaud the work that Director of Open Data Lindsay Mollineaux has done for open data.

This is the best practice in other major cities with open data laws; they recognize the need for strong executive oversight. Specifically, San Francisco, Chicago, Philadelphia, Boston, have all placed responsibility for implementing the open data law in the executive branch.

The Sunlight Foundation notes¹ that:

“Positioning an authority structure outside of a department or agency (i.e. outside of the IT department), and in the executive branch, provides that actor with more power to liberate data swiftly. Ideally, open data policies also include binding regulations with real consequences to not opening up data as scheduled.”

DoITT lacks the authority to impose real deadlines and real consequences on its fellow agencies. Other cities have recognized the need for strong executive oversight, and we recommend the council amend the open data law to provide her with the legal authority to continue her work.

3. Pass a Data Usability Amendment

We also ask that the council pass a “data usability” amendment to mandate that agencies take specific actions to make their data more useful to the broader public. Specifically, this amendment should require agency data sets with location information include not just uniform addresses and latitude/longitude information; data sets should have community board information for each entry, as well as city council district and police precincts. This would dramatically increase the usefulness of the portal’s information for the vast majority of users.

This should require little to no work on the part of agencies and DoITT. New York City already has sophisticated software to “geocode” data with addresses to determine latitude and longitude information: this is how the NYCityMap is created. Neither DoITT nor agencies actually use the geocoder before uploading data sets to the portal. Once data sets are geocoded, DoITT can automatically check latitudes and longitudes to place each entry in a data set in the appropriate community board district, council district, and so on.

In fact, this is already done for certain high-profile data sets, such as 311 Requests. This should be the rule, rather than the exception. Data on other Socrata-based open data portals, such as the New York State portal, is automatically geocoded. There’s no reason New York City can’t do the same.

¹ <http://sunlightfoundation.com/blog/2014/03/28/an-open-data-ombudsman-and-rethinking-oversight-authorities/>

4. Create a Formal Public Process for Requesting Data Sets

The open data portal contains a form that allows members of the public to request data sets. Since the portal launched in October 2011, just six data sets have been approved for publication based on a request from this portal, seven have been rejected, and there are another 170 which are still “open.” We assume most of these requests are no longer under consideration, but the process is completely opaque.

We ask that the council create a transparent way to request data sets for inclusion on the open data portal. The proposed legislation allows six months for a final response from DoITT, which is roughly six times slower than what’s allowed under the Freedom of Information Law for requesting the same data set.

5. Create a Way to Report Errors in Data Sets and Get Them Fixed

Along with the problem in getting data onto the portal, there are problems with the data sets once they’re opened. This is understandable: one of the purposes of open data is letting the public help agencies find data problems, so the agency can address them. The Socrata open data portal has a built-in platform for comments for exactly this purpose.

However, researchers from NYU’s GovLab have determined that there have been roughly 300 comments on the portal since it launched, and that most comments never receive replies. Even worse, commenters wait over six months on average to receive a reply. The result is that there is no meaningful way for the public to report errors, receive responses, or track the fixes made to the data set. Many users see no point to leaving a comment, and there are users of this portal who’ve resorted to hosting their own private versions of cleaned-up data.

We recommend the council amend the Open Data Law to require DoITT to monitor and respond to comments on the open data portal.

Appendix A

Reinvent Albany's Analysis and Comments on Proposed Open Data Legislation

Summary

We agree with City Council that it is time for the NYC Open Data Law to evolve. We are very excited that Council is pushing Open Data ahead and look forward to a serious discussion between council, the administration, and public stakeholders on legislation that will address the three major issues with NYC Open Data.

Major issues with NYC Open Data

Data Availability: Many civil society groups are frustrated with agency resistance to publishing public data sets that are of obvious interest. There is no mechanism to compel agencies to publish public data or to meet their own data publishing schedules. This remains a major flaw in the Open Data Law --- one that was identified when it was drafted. Open Data is not like FOIL, which has a legal appeal process.

Data Usefulness: Many city datasets could be made much more useful through simple additions like including an identifying field for community boards for location specific data, ensuring uniform geospatial data and consistent formats for dates, and reconciling differences in agency data sets, agency data sets modified by MODA and these same data sets on the Open Data Portal.

Data Quality: Many agency data sets contain errors. Unfortunately, there is no way for the public to report errors, receive responses, or track the fixes made to the data set. Put simply, there is no way to file a 311 request to improve a data set.

Recommended Solutions for Open Data Problems

1. A private right of action to make open data more like FOIL, and allow the courts to compel agencies to publish promised public data sets and compel agencies to comply with other aspects of Open Data Law.
2. Responsibility for implementation shifted to the Mayor's Office instead of DoITT.
3. A "data usability" amendment that mandates that agencies take specific actions to make their data more useful to the broader public, including simple things like ensuring city map data includes a field for community boards.
4. A formal public process for the public to request publishing of new agency datasets.
5. A formal public process for reporting problems with data sets and getting them fixed. I.e. a "311" for open data that allows problem to be easily reported and tracked.

Assessment of package of City Council Bills

Overall, we see this package of bills as small adjustments to the Open Data Law which will result in very modest improvements. We do not object to any, but do not feel these bills solve the fundamental problems with the open data law which were identified above.

By far, the most important bill is Intro-916 (Vacca) which mandates DOI audits of large agencies to determine how well they are complying with the Open Data Law. However, we believe that a private right of action – like FOIL has – is a far stronger solution to agency non-compliance. Of the bills in this package, Int. 916 will have the greatest practical impact.

Legislation

Intro 916-2015 (Vacca)

Intent of Bill

Use DOI audits to pressure nine major agencies to publish data defined as “public data” under the Open Data Law.

Mandate

Requires the Department of Investigation to audit agencies’ compliance with the open data law, including: Sanitation, Corrections, Buildings, Police, Transportation, and the Business Integrity Commission and three other agencies to be selected by the Commissioner of Investigation.

Pro

Most powerful of the bills in this package. Seems likely to get agencies to better comply with data publishing. Establishes concept of independent oversight of open data by a third-party agency.

Con

Unclear what the audits would look like. Legislation only specifically calls for reporting on public data sets which are missing from the open data plans. There are data quality issues, data usability issues, and other concerns.

Recommendations and Questions

Unclear why DOI is performing this audit. We’re not familiar with its operations enough to know if this an appropriate or realistic responsibility for this agency.

Intro 915-2015 (Vacca)

Intent of Bill

Ensure that agencies post the same public data on their websites and on the Open Data Portal.

Mandate

Agencies with data sets posted to the open data portal and their web site must update their data sets on the open data portal no more than three days after they update the data sets on their web site.

Pro

Prevent agencies from neglecting to update the open data portal when they publish new data.

Con

It could be difficult to automate this process in a satisfactory way. Much data on agency sites linked to the Open Data Portal is in an Excel spreadsheet format that may require a human to convert to a CSV file. The Open Data Law allows agencies to link data to the portal and does not require them to convert that data into a CSV.

Recommendations and Questions

We are not convinced this bill is needed. If it is, DoITT can be directed to put an updated copy of agency data on the Open Data Portal from links on agency web-sites. DoITT can automatically import files to the Portal from agency web sites. However, it can take a fair amount of work to convert highly structured spreadsheets into a CSV format that can be searched, mapped and transmitted by API via the portal.

Intro 914-2015 (Torres)**Intent of Bill**

Ensure agencies respond to public requests for the publication of data sets.

Mandate

Each request for data on the open data portal must receive a response within six weeks, after DoITT consults with the agency-owner of that data. DoITT must publish on the portal its determination as to priority inclusion of the data set.

Pro

Creating a robust process for the public to request data is an improvement. DoITT has two weeks to acknowledge the request.

Con

DoITT has six months to determine whether the data set should be “prioritized” for inclusion on the portal. That’s far too long. FOIL requires responses in one month. Also, it’s unclear what it means for DoITT to determine that a dataset is a “priority” for inclusion on the portal. How long does DoITT have to open priority data? What if DoITT declines to prioritize the data: will it be included in the open data plan anyway?

Recommendation and Questions

Great intent, wonder about mandate. Are there other best practices? Reduce time for a determination.

Intro 908-2015 (Palma)

Intent of Bill

Encourage agencies to publish data sets in the Open Data Portal when those data sets have been provided completely, or partially, to a member of the public in response to a FOIL request.

Mandate

When a FOIL officer responds to a FOIL request by releasing tabular data, he or she must provide that data to his or her agency’s open data coordinator. The open data coordinator may then prioritize the data for inclusion on the open data portal.

Pro

Using FOIL to trigger the publication of open data is a great idea. It cuts down on agency workloads for processing the same requests over and over.

Con

The language in the bill is vague and confusing. For instance, the bill calls for FOIL disclosures to be used to “prioritize” data. What does this mean? This bill doesn’t require the agency to update its open data plan in response to FOILs. The public wouldn’t see how agencies are using FOIL to release data sets.

Recommendations and Questions

Is this bill intended to add to agency open data plans by using FOIL to “remind” agencies that they forgot to include specific data sets on their plans? This would be a good step forward, but that language about Prioritizing needs some clarification. Mandate that if an agency provides public data in response to a FOIL request, that agency must add the data set to its open data plan immediately, even if it’s unscheduled for release.

Intro 900-2015 (Kallos)

Intent of Bill

All public data should use the same geospatial data formats to enable multiple layers of map data to be used together. This bill codifies into law the standards already established in DoITT’s open data Technical Standards Manual.

Mandate

Data sets published on the Open Data Portal which include street addresses or other geographic information shall include a data field with standard geospatial reference data (latitude and longitude or X/Y).

Pro

Standardizing open data makes it easier for the public and government to map and to use multiple layers of maps.

Con

This language would not create any new responsibility for DoITT or other agencies.

Recommendations and Questions

The open data law already directs agencies to publish data in accordance with DoITT's Technical Standards Manual. That Technical Standards Manual, in turn, already directs agencies to publish geographic information in a uniform fashion and with geospatial reference data formatted in the WGS 84 or New York State Plane system in §4.2.1.4: "Geospatial data must be published in the Web Mercator coordinate system (WGS 84/EPWG:3857) to make the data easy to use with popular online mapping services."

DoITT has failed to enforce this provision as a regulation, and we are not sure what difference this makes as a law.

Intro 898-2015 (Gentile)

Intent

Improve search and archiving by improving and standardizing the description of datasets—the "metadata"—on the Open Data Portal. This codifies into law rules published in DoITT's open data Technical Standard's Manual.

Mandate

Applies primarily to DoITT. Requires each data set on the open data portal to have a plain language data dictionary, which will describe each column heading and define any acronyms or terms of art. In addition, the dictionary will describe units of measure, the range of possible values, and any relationship between the columns. The dictionary may include other information. The dictionary for data sets that are updated less than once per day must include the date the data set was generated and uploaded to the portal.

Pro

Standardized and improved metadata would help users find the data sets they're looking for.

Con

The Technical Standards Manual already contains lengthy regulations for the metadata included with each data set. DoITT and other agencies have failed to

provide these for three years. We are not certain passing a law would improve the state of metadata.

Recommendations and Questions

Allow the public to enjoin agency open data coordinators with Article 78 lawsuits; remove the sentence “this chapter shall not be construed to create a private right of action to enforce its provisions” from the existing Open Data Law so members of the public can ensure that the provisions are enforced.

Intro 890-2015 (Cabrera)

Intent

Ensure that changes in data can be viewed over time, allowing longitudinal analysis of data.

Mandate

Requires DoITT to archive copies of data sets which are replaced (e.g. a list of licensed taxi drivers) instead of continually added to (e.g. a list of all 311 requests since 2010), if more than half the entries have changed from the previous version. This archival version must be created at least once a year.

Pro

This solves an important data availability problem by archiving old versions of ever-changing data sets. Longitudinal studies of open data depend on having months’ or years’ worth of data points, something that’s not always possible currently. This is a great improvement.

Con

The bill as written only requires the archiving of datasets of which a “majority” of the data is removed or replaced. The bill notes “permit and licensing data” is a category of data set which would be covered by this provision, but it seems unlikely—for example—that 50% of the licensed taxi drivers in any given year were not licensed the previous year. Therefore, that data set would not be covered by the letter of the law, despite being within the spirit of the law.

Recommendations and Questions

Lower the threshold for archival versions of data sets from “majority.” Perhaps 25%, or 10% of the data.

With that one change, this becomes a very good bill. However, the Socrata-powered portal does not have the ability to archive old versions of data sets, and is unlikely to gain that ability in the future. DoITT would need to create a program to “crawl”, download, and archive open data sets on a third-party site. This isn’t a problem with the bill as written, but is a problem with the Socrata software that the Open Data Portal uses.