



Testimony of Amen Ra Mashariki  
Chief Analytics Officer and Director of the Mayor's Office of Data Analytics  
Before the New York City Council Committee on Technology concerning

**Oversight – Open Data Plan 2015 and**  
**7 Pieces of Legislation**

**1:00pm, Thursday, October 1, 2015**  
**14<sup>th</sup> Floor Committee Room, 250 Broadway New York, NY**

Good afternoon. My name is Dr. Amen Ra Mashariki. I am the Chief Analytics Officer and Director of the Mayor's Office of Data Analytics (MODA), as well as the Chief Open Platform Officer for the City of New York. Here with me is Albert Webber of the Department of Information Technology and Telecommunications' (DoITT) Office of Advanced Technologies. The Mayor's Office of Data Analytics is housed within the Mayor's Office of Operations, and works in partnership with Anne Roest, the City's Chief Information Officer and DoITT Commissioner, to fulfill the requirements of the City's open data law, which was enacted as Local Law 11 in 2012. The Administration fully supports that law and works tirelessly to ensure it is fully implemented.

The promise of Open Data is transparent, free, and accessible data. Data is more than just numbers. It's information that can create new opportunities and level the playing field for New Yorkers. It illuminates issues to invite changes in frameworks, and leads to insights that turn impenetrable issues into solvable problems. It's the invitation for more voices to join critical conversations.

I am here today to testify on the progress of the implementation of the open data law via our reinvigorated interactions with agencies and our open data vision and strategy, as well as on the proposed pieces of legislation by the City Council designed to enhance the open data law. Over the past year, MODA and DoITT have been working energetically to engage agencies in continuing to open data sets to the public. To date we have released 1,386 data sets this year, up from 1,268 last year. This includes a number of notable new sets like trip data from the Taxi and Limousine Commission, data from City Record Online, and "Pre-K for All" data. Along these lines, we have and will continue to work diligently with agencies to assist them in releasing their data.

New York City remains the gold standard in producing open data, both nationally and internationally. Our ability to work with agencies to open their data to the public puts the City far and away ahead of all cities with respects to the number of public data sets available on a portal. Our open data efforts have also earned New York City a top spot on the Sunlight Foundation's national rankings of Open Data cities. While we believe this, in and of itself, is a strong accomplishment, the open data law is not just about releasing open data sets. As stated in the

Open Data law's declaration of legislative findings and intent, open data serves to "streamline intra-governmental and inter-governmental communication and interoperability, permit the public to assist in identifying efficient solutions for government, promote innovative strategies for social progress, and create economic opportunities." MODA and DoITT are partnering to realize a vision and strategy that will allow us to achieve those stated outcomes of the law.

### **Strategy and Vision: Open Data for All**

"Open Data for All." With this initiative we are saying that open data is an invitation for anyone, anywhere, at any time to engage with New York City government. The Open Data Plan, which was released on July 16, 2015, establishes the City's vision to deliver on the promise of the Open Data for All initiative to the benefit of all New Yorkers. This is our vision of what we can do and what we aim to achieve with open data. The ultimate success of the Open Data for All initiative will not only be measured in the number of data sets that are published on the Open Data Portal, but in the number of New Yorkers who use open data in their daily lives. And that's not just the tech-savvy New Yorkers – it's all New Yorkers, in all five boroughs.

To deliver on this promise, we will:

- Continue to release new data sets and identify opportunities to provide data for all types of users;
- Focus on delivering quality data, enhancing the usability of the data portal, and expanding access to open data through use; and
- Continue to improve and enhance the technical infrastructure that provides the foundation for the entire initiative, and increase responsiveness to and flexibility for user needs.

With these goals in mind we are driving aggressively towards building an open data ecosystem that supports expanded access to open data and better quality of data, as well as enhanced portal usability.

*Expand Access:* Among the primary goals of the Open Data for All initiative is to meet users where they are. This year MODA and DoITT have begun a citywide open data engagement tour. This has included visits to CUNY campuses, visits to Borough Board meetings organized by the various Borough Presidents, online webinars, Open Data tutorials, and other community-specific speaking engagements. We'll be gathering feedback from the citywide engagement tour to determine how to best meet the needs of all New Yorkers.

*Deliver Quality:* We also are ensuring data quality through the development of applications such as the Business Atlas. The Business Atlas is a portal that was created through a partnership between the Mayor's Office of Data Analytics and the Department of Small Business Services. It uses open data and also partners with a small startup here in the city called Placemeter to provide market research for local entrepreneurs.

*Enhance Usability:* Earlier this summer we partnered with our open data platform provider to launch a new feature called Data Lens. Data Lens utilizes artificial intelligence technology in order to display data to everyday New Yorkers in a way that makes sense to them and is visually intuitive.

## **The Proposed Bills**

The proposed bills amending the open data law are all focused on our shared goal of improving the quality, usability, and accessibility of open data.

MODA and DoITT have long believed that the Open Data Portal should be the primary location where agency data sets should be available, machine readable, and accessible. We have engaged agencies with this in mind. We understand that in order to do this, many costly mechanisms and processes must be in place. We firmly believe in working towards the goal set forth in Intro 915 to accomplish this. MODA and DoITT look forward to discussing with the Council how to best put a comprehensive plan in place to achieve this public data set updating capability.

We believe that in order for us to truly achieve a ubiquitous open data community, MODA and DoITT must be responsive to users as well as provide quality customer service to users. Timely responses to public requests on the Open Data Portal are paramount to us achieving this level of customer service. Intro 914 is very much in line with that goal and would enhance our responsiveness to such requests, and MODA fully supports the intent of this legislation.

MODA already instructs agencies to look at data that they have released in response to a FOIL request to decide whether the released data should be considered a public data set to be included on the Open Data Portal. Commissioner Toole of the Department of Records and Information Systems (DORIS) is currently building an Open FOIL system. We have discussed building a data exchange with this system that feeds directly into the Open Data Portal so that we can get automatic feeds of FOIled data. Such an approach would accomplish the goals set forth in Intro 908.

Enhancing the usability of the Open Data Portal is a key objective for MODA and DoITT. Ensuring that data dictionaries are available for each data set within the portal, as envisioned in Intro 898, will enhance the user's ability to understand and use the data. This is central to what we are trying to accomplish when we say Open Data for All. We support the intent of this bill and look forward to working with DoITT and the Council to identify the most appropriate way to implement this capability.

We believe that managing data standards throughout the portal is extremely important. We believe both high quality data dictionaries and the use of open data standards throughout the portal are achievable with agency and open data coordinator (ODC) engagement. Intro 900 is a great way to begin the conversation around citywide standardization of geospatial data, but would ultimately require processes that will require the expertise of the full tech leadership of the city, intimate agency engagement, and the investment of significant resources.

To ensure usability and access to all open data we are going to add language to the technology standards manual around archiving and maintaining old records instead of purging them after a

certain time period. Intro 890 similarly seeks to accomplish a strategy for the preservation of historical records. We must maintain a balance between keeping a clean, easy-to-use portal, preserving important historical data, and the cost of storing significant amounts of data that may have outlived their usefulness.

With respect to Intro 916, we have legal and practical concerns with the audit proposed by the bill. In particular, it appears that the bill would authorize the Commissioner of the Department of Investigation to inspect all of an agency's data, records, and information—including sensitive, confidential, and privileged information—in order to determine whether the information comprises a “public data set.” Although we support the principle of developing mechanisms to ensure compliance with the open data law, we do not think the proposed audit is the best way to achieve that goal. We are open to discussing alternative means for advancing agency compliance and accountability.

I would like to thank the Council Technology Committee for providing the opportunity to testify today and for your continued support of open data. At this point I'd like to turn it over to Albert Webber from the Department of Information Technology and Telecommunications to walk you through some of the finer details of implementing the open data law and the new enhancements we have released in response to feedback from the community. After Mr. Webber's testimony we will be happy to answer any questions you may have.

**DEPARTMENT OF INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS TESTIMONY  
BEFORE THE CITY COUNCIL COMMITTEE ON TECHNOLOGY  
OVERSIGHT ON OPEN DATA AND PROPOSED LEGISLATIVE PACKAGE  
THURSDAY, OCTOBER 1, 2015**

---

Thank you, Amen, and good afternoon Chairman Vacca and members of the City Council Committee on Technology. My name is Albert Webber and I am pleased to speak with you today about many of the improvements made to the NYC Open Data Portal since we last testified.

To date we have unlocked nearly 1,400 data sets and maps via the NYC Open Data Portal, sourced from more than 80 New York City agencies and entities. As part of the annual July 15 update to the NYC Open Data Plan, City agencies identified an additional 282 data sets to be released through December of 2018. We also work closely with the Mayor's Office of Data Analytics (MoDA) and agencies to continue identifying data to be released or listed in the Open Data Plan beyond the yearly July update, and we make monthly updates to the Plan reflecting those updates.

**Data Integrity**

To date we have automated the refreshing of approximately 120 data sets, to ensure their timely and accurate delivery. Many of the 120 data sets automated were in high public demand, including 311 Service Requests, Police Department Motor Vehicle Collisions, and Restaurant Inspections from the Department of Health & Mental Hygiene.

This year alone 18 new data sets have been automated, including Derelict Vehicles from the Department of Sanitation (DSNY); For-Hire Vehicle and Active Drivers from the Taxi & Limousine Commission; and the City Record Online, Green Book, and City Store from the Department of Citywide Administrative Services. Automation – aside from the initial work done to implement the process – means that no human intervention is necessary for a data set to be refreshed.

Additionally – and particularly topical given one piece of introduced legislation – we recently completed a proof-of-concept that should allow us to begin the automated delivery of geospatial data sets to the NYC Open Data Portal.

With regard to the 1,200-plus data sets that are not automated, we work closely with agencies and their Open Data Coordinators to manually refresh them. This process is often time-consuming and labor intensive, but our team diligently works to ensure timely data.

We are also undergoing an effort to consolidate data sets where possible. There are a number of data sets broken down by year, borough and other attributes, making it difficult for users to export a holistic view of the data they desire. By consolidating data sets we can allow users to more easily locate and download data, improving the overall user experience. This consolidation effort may decrease our data set totals, but in keeping with the themes of *Open Data for All* that Dr. Mashariki just discussed, this approach will in the long run facilitate fuller access to data sets and maps for our many users.

**New additions**

As we continue our efforts to make data sets more usable, we also continue the march toward opening new data. Since our last update, the following new data sets have been made available:

- Universal Pre-K Data – Locations of facilities, number of seats available, website, and contact information.
- Workforce 1 Job Listings and Recruitment Events – Job openings and recruitment events.
- Civil List Data – Salary information for all New York City employees.
- TLC Trip Data – Information on the millions of trips taken by New York City's yellow taxis on an annual basis.

Data sets scheduled to be made available by the end of the year include:

- Bridge Ratings – Bridges rated good or very good by calendar year.
- Reportable Disease Surveillance Data – Information on cases of reportable diseases (not including STDs, Tuberculosis or HIV) in New York City.
- Events and Cleanup Costs – Names, locations, and dates of events permitted by the Street Activity Permit Office and the DSNY cost of clean-up efforts billed to event organizers.
- School Budget Overviews – Showing individual school budget allocation by year.

Additionally, earlier this summer we partnered with our open data platform provider to launch a new feature called "Data Lens." Data Lens is a machine-learning enabled tool that allows for the creation of auto generated charts, graphs, and search features so that users, be they data scientists or everyday New Yorkers, can easily view and interpret data sets. In short this tool helps to make the viewing of large data sets less intimidating, and more accessible to a broader audience.

These ongoing efforts encapsulate the spirit of the *Open Data for All* initiative Dr. Masharki discussed earlier, and reflect the de Blasio Administration's commitment to keeping New York City at the vanguard of the open data movement. And as we continue this work, the continued support of and feedback from the city's vibrant civic technology community has proven invaluable – as has the support, led by Chairman Vacca and the Technology Committee, of the City Council.

A number of the bills introduced last month align very closely with the aims of *Open Data for All* – making open data more usable, more accessible, to New Yorkers of all stripes and in every corner of the city. Some of these bills would mandate efforts already underway, while others fit squarely with the de Blasio Administration's five borough strategy to enhance the open data initiative.

Taking a brief look at each in turn:

**Intro. 914** would establish response timelines for public requests on the NYC Open Data Portal, and help set user expectations about when the data they have requested can be made available within the context of the City's larger Open Data Plan. We wholeheartedly support the intent of Intro. 914 and have no significant concerns about this proposal. As you know, as part of the annual update last July we committed to updating the plan on a monthly basis, and regular feedback on requests for new data sets fits into that increased engagement strategy.

Intros. 890, 898, and 915 embody proposals that speak directly to what *Open Data for All* is all about, and we look forward to working with the Council on each to explore ways that our shared goals may be realized.

**Intro. 890** relates to the retention of archival data on the NYC Open Data Portal, and rightly references the Technical Standards Manual as a place for the appropriate language to reside. To the extent that old data sets are removed from the Portal without explanation this proposal is in keeping with *Open Data for All*.

But mandating that old data sets, replaced by those from the same period that are more accurate and complete, have to remain on the Portal indefinitely, is problematic. This can be confusing to visitors and runs counter to making City data more understandable and usable for all. There will also be some costs for additional storage capacity as data sets – like some of those we mentioned earlier, containing millions of rows of data – become increasing large and complex, but we look forward to discussing with the Council a path forward on data set retention.

**Intro. 898** requires every data set on the NYC Open Data Portal be accompanied by a plain language data dictionary and is aimed at exactly the right place: Making what can be very bureaucratic data more usable and understandable to the layperson. Data dictionaries and supporting data set documentation were at one point common on the Open Data Portal, and can be a valuable standard to more formally implement.

Of course, the role of providing headings, explaining columns, and defining the technical terms and acronyms included in data sets needs to be completed by the agencies submitting their data to DoITT and MoDA for posting and/or automation (not by DoITT or MoDA directly), and we therefore agree with its inclusion as a requirement that can be made more specific as part of the upcoming Technical Standards Manual refresh. We look forward to discussing with the Council the most effective means of implementing data dictionaries.

**Intro. 915** requires that public data sets updated on other City agency websites be posted to the NYC Open Data Portal within three days. This would help ensure continuity and consistency of user experience when searching for official New York City data, which is directly in line with both the original intent of the City's Open Data Law, as well as with *Open Data for All*. While DoITT is constantly working with agencies to automate data sets, there still remains work to be done. There will be some cost and resource considerations to be addressed in pursuing this effort but we share with the Council the goal of keeping City data accurate, updated, and consistent across all channels – and we look forward to working with you on making that vision a reality.

The next proposal, **Intro. 908**, establishes a new process for making agency FOIL responses publicly accessible. The spirit of this proposal squarely aligns with the de Blasio Administration's commitment to the promise of transparent, free, and accessible open data. As part of that promise, we are committed to ensuring that agency responses to FOIL requests are widely accessible to the public.

To that end, the Department of Records and Information Services is currently developing an OpenFOIL Portal that will launch by the end of this year. This centralized website will empower the public to easily submit and track FOIL requests for all City agencies, as well as access the vast majority of records responsive to previously-filed FOIL requests, even where there is no accompanying "data set."

Such proactive disclosure through a centralized repository will dramatically expand public access to government records. Therefore, rather than duplicating work to be completed in the coming months, we are happy to explore crafting a technical solution whereby the records that will be posted to OpenFOIL are automatically transferred to the NYC Open Data Portal.

The final two proposals, Intros. 900 and 916, have elements with which we agree but also present significant staffing, cost, timing, or legal concerns.

**Intro. 900** requires every data set on the NYC Open Data Portal that contains address information to utilize a standard field layout and presentation of that address information, including corresponding geospatial reference data.

Having a single City standard for address and geospatial information data is ideal, but we do not believe it is a goal best achieved through open data legislation. While Open Data would benefit from such a standard, the standard itself should be developed via a larger collaborative process whereby agencies, advocates, and users can work toward a generally agreed-upon, adaptable standard for address and geospatial data. Once established, the systems utilized by agency databases feeding the NYC Open Data Portal would adhere to that standard when producing data for publication. Older legacy systems, once replaced, would then also be required to incorporate those standards. This is a worthwhile effort, but one that will require additional resources and time – substantially more so than the 90 days proposed – to implement effectively.

In the interim, we continue to make all the relevant data – tabular, shape files, etc. – available on the Portal itself for interested individuals to use in geocoding the data sets they are most interested in. This indeed is one of the earliest, truest promises of open data: Provide it to the public and let them do with it what they will, without further interference.

**Intro. 916** would require the Commissioner of Investigation to conduct audits of certain agencies' compliance with the Open Data Law. While we – and all public officials – can always do a better job of serving New Yorkers, the City's Open Data Law is among the most ambitious, comprehensive policies of its kind in the world, and has set a model for many governments to follow. We are happy to work on its continued implementation and committed to doing even better still. There are a number of steps between making the accountability improvements Dr. Mashariki mentioned (for example, more clearly stating on the Portal which data sets are either removed or delayed, with detailed explanations) and requiring the Commissioner of Investigation to conduct audits in an area where the City has demonstrated such noteworthy, transformational success in just over three years. Most significantly, we are concerned that the proposed audits would require the Commissioner of Investigation to inspect *all* of an agency's data, including even sensitive, confidential, or privileged information, in order to determine the existence of public data sets. We have legal concerns about the Commissioner's authority to access such information, as well as concerns about the utility of such audits generally in advancing Open Data goals.

Thank you for the opportunity to testify this afternoon. Open Data has been and remains a priority of the de Blasio Administration and we thank our partners in the City Council, and across the civic tech and advocacy communities, for their continued efforts in making New York City's successful implementation even better.

This concludes our prepared testimony and we look forward to answering your questions.

Thank you.



## Additional Written Testimony of Reinvent Albany, *co-chair 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 notes<sup>1</sup> 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.

---

<sup>1</sup> <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/EPSSG: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.

**Date: October 1, 2015**

**To: New York City Council Committee on Technology**

**From: Tara Das, Interim Head of Social Work Library, Columbia University**

Good afternoon everyone,

My name is Tara Das, and I am the interim head of the Social Work Library at Columbia University. I'm affiliated with the Digital Social Science Center in the Columbia University Libraries, which supports data-intensive research by acquiring data collections and managing data in its spatial and numeric data catalogs. We provide research consultations to students, faculty, and other university affiliates. My responsibilities also include overseeing the university collections in government information issued from the federal, state, and city government.

With the majority of government information now in electronic format, including open government data, concerns arise with preservation of this information so that it remains publicly accessible over time. For example, a government agency may decide to remove a publication from its website. If they did not transfer a copy to an archive or library, users will find it difficult to access that publication. With this in mind, my colleagues and I decided to select datasets from NYC Open Data, where the topics were of interest to our users (e.g. crime and city planning), and maintain them in our data catalogs.

I took a much closer look at the datasets on NYC Open Data under the assumption that, if Columbia University was going to store this data, it needs to be shareable and reusable by others for research. Users cannot analyze data if they do not know what it means, how it was collected and cleaned, and what was the purpose of data collection. They need to be equipped with the same information that the data creators have in order to conduct research with the data, and appropriately evaluate city services. Close examination of NYC Open Data revealed that it was largely not reusable due to poor data documentation. For example, datasets were missing definitions of column names and values, and descriptions of how the data was collected and cleaned.

### **NYC Open Data case studies**

I present 3 cases using NYC Open Data to further explain what I mean, and will discuss the first 2 at this hearing.

1. The Department of City Planning has datasets related to Hurricane Sandy: Damage by Sandy by Land Use<sup>1</sup> and Damage by Sandy By Age of Building.<sup>2</sup>

#### **Damage by Sandy by Land Use**

---

<sup>1</sup> <https://data.cityofnewyork.us/Housing-Development/Damage-By-Sandy-By-Land-Use/tqvi-w9ww>

<sup>2</sup> <https://data.cityofnewyork.us/Housing-Development/Damage-By-Sandy-By-Age-Of-Building/mgjt-zuui>

Land Use	% of Total Destroyed by Storm	% of Total Red Tags
Commercial & Office	2%	1%
Industrial & Manufacturing	< 1%	1%
Mixed Residential & Commercial	6%	1%
Multi-Family Elevator	< 1%	< 1%
Multi-Family Walk Up	1%	4%
One & Two Family	88%	90%
Open Space & Recreation	1%	1%
Parking	< 1%	< 1%
Public Facility & Institution	< 1%	1%
Transportation & Utility	< 1%	< 1%
Vacant	1%	1%

### Damage by Sandy by Age of Building

Year Built	Inundation area	Destroyed by Storm	Red tags	Yellow Tags
Before 1983	84%	98%	94%	67%
1983-2001	9%	< 1%	4%	23%
2002 or later	5%	< 1%	1%	9%
No Data	2%	2%	1%	1%

How are we to analyze this data when we don't know what the columns mean, how the data was collected, and how the percents in the dataset are calculated?

2. The Department of Homeless Services has 2 datasets on NYC Open Data that overlap with their Daily Report, which is available in pdf format on their website.<sup>3</sup> But they're not all exactly the same. The time and subject coverage vary across the 3 sources. The datasets are the Daily

<sup>3</sup> <http://www1.nyc.gov/assets/dhs/downloads/pdf/dailyreport.pdf>



Report of Single Adult and Family Intake<sup>4</sup> and DHS Daily Report.<sup>5</sup> In addition to lack of data documentation, there are other issues with these reports. The Daily Report of Single Adult and Family Intake has only 1 day of data (1/28/13), and the other - DHS Daily Report - is routinely updated but does not contain columns that the first dataset and the agency website report do.

Questions that come to mind with these datasets include - How are the categories in the reports defined? How is the data collected and calculated? If all the reports come from the same data collection system, then why is the data split up as it is in NYC Open Data?

3. The Administration for Children's Services has one dataset on NYC Open Data -

**ACS Community Partners**.<sup>6</sup> It is described as a directory of preventative agency location centers, and has the following columns -

PROGRAM\_LST\_ID  
PROGRAM\_CODE  
PROGRAM\_NAME  
ACTIVE\_FLAG  
AGENCY\_NAME  
DESCRIPTION  
BOROUGH  
PROG\_CONTACT\_LAST\_NAME  
PROG\_CONTACT\_FIRST\_NAME  
PROG\_CONTACT\_PHONE\_NO  
PROG\_CONTACT\_FAX\_NO  
PROG\_CONTACT\_EXTENSION  
PROG\_CONTACT\_EMAIL  
PROGRAM\_ADDRESS\_1  
PROGRAM\_ADDRESS\_2  
PROGRAM\_CITY  
PROGRAM\_ZIP

---

<sup>4</sup> <https://data.cityofnewyork.us/Social-Services/Daily-Report-Of-Single-Adult-And-Family-Intake/sci4-yqgk>

<sup>5</sup> <https://data.cityofnewyork.us/Social-Services/DHS-Daily-Report/k46n-sa2m>

<sup>6</sup> <https://data.cityofnewyork.us/Social-Services/ACS-Community-Partners/9hyh-zkx9>

OFFICE\_MON\_FROM\_TIME  
OFFICE\_MON\_TO\_TIME  
OFFICE\_TUES\_FROM\_TIME  
OFFICE\_TUES\_TO\_TIME  
OFFICE\_WED\_FROM\_TIME  
OFFICE\_WED\_TO\_TIME  
OFFICE\_THURS\_FROM\_TIME  
OFFICE\_THURS\_TO\_TIME  
OFFICE\_FRI\_FROM\_TIME  
OFFICE\_FRI\_TO\_TIME  
OFFICE\_SAT\_FROM\_TIME  
OFFICE\_SAT\_TO\_TIME  
OFFICE\_SUN\_FROM\_TIME  
OFFICE\_SUN\_TO\_TIME

Some of these columns are intuitive enough to figure out, like the program contact information, address, and office hours. But there are still questions of how an agency comes to be defined as a preventive agency, what the criteria are for inclusion in this directory, and what is the different between a program and agency. In addition, these columns are difficult to figure out -

PROGRAM\_LST\_ID  
PROGRAM\_CODE  
ACTIVE\_FLAG  
DESCRIPTION (with categories for Family Treatment and Rehabilitation, General Preventive, Intensive Preventive, and Medically Fragile)

As I hope these cases illustrate, open data is not just a matter of making data available. Since most of this data comes from internal business operations, there are abbreviations and/or agency jargon in a lot of datasets. Truly open data - data that is usable by others - requires data about the data, or as librarians call it - metadata. People less familiar with the idea of metadata can understand it in terms of data documentation, data dictionary, or the data codebook that is essential to understanding a dataset that one did not create, and to interpreting any analysis that used that dataset. In research communities, it is expected practice to include metadata that describes the data collection background, instruments, data definitions, and methodology when sharing datasets.

## **Outreach to NYC government agencies for data documentation**

Given the limitations in NYC Open Data, our immediate objective at Columbia University Libraries shifted from preserving it to acquiring data dictionaries for datasets. Since data dictionaries are not required for NYC Open Data, my first step was to contact the dataset owners through the portal (via the About tab in each dataset) to see if they can provide this information. No one replied. I reached out to the Mayor's Office of Data Analytics to see if they could assist with providing agency contacts or, in general, with acquiring data dictionaries. There was one conference call with them where these concerns were expressed, but no follow-up action. I then reached out to the NYC Transparency Working Group. Dominic Mauro of Reinvent Albany responded, and provided me with a list of open data coordinators for outreach. Since April of this year, I have reached out to open data coordinators with questions about specific datasets. I ask for definitions behind column names, and categories within columns, methods for data calculations, and processes of data collection.

This is a time-intensive activity. One data dictionary that I've created is included in Appendix B of this testimony, for the ACS Community Partners dataset. Since data dictionaries are not legally required, relying on government officials to voluntarily provide this information is not the best way of making open data usable for research. If I need to contact an open data coordinator about multiple datasets, putting all my questions re: data definitions and data collection methods in one long email is not the best way to get a response. This means I ask a few simple questions first, see if I get a response, and then follow up with phone calls or emails to ask more questions in order to create data dictionaries.

Completing just one data dictionary for one dataset requires sustained communication with an agency. While I received at least one response from some open data coordinators, I didn't receive any response from others, including Office of Management and Budget, Department of Correction, and Department of Transportation. Some agencies did not have an open data coordinator to contact, as of February 2015, but have provided datasets to NYC Open Data. These include Civilian Complaint Review Board, Fire Department of New York, New York City Housing Authority, and Health and Hospitals Corporation.

Moreover, government officials aren't always equipped with definitions and descriptions of how data is collected and what each category means. This data comes out of routine government services and programs, and wasn't originally collected for public release. Sometimes the people "on the ground" collecting this data for internal purposes don't have written documentation for what specific categories mean. I encountered this issue while speaking with Department of Homeless Services about their Daily Report. Other times, I was referred to laws to assist with understanding datasets. While helpful, I'm not a lawyer and couldn't be confident that I was mapping the appropriate value to the legal definition.

In sum, obtaining data documentation retroactively is not realistic.

## **Support for proposed bills**

In light of these experiences, and given the critical need for preservation and data documentation for NYC Open Data, I strongly support bills Int. No. 890, Int. No. 898, Int. No. 914, Int. No. 915, and Int. No. 916. Along with my support, I have the following comments for each bill.

Int. No. 890 - In relation to the retention of data on the open data portal

- Will the archival copies be stored in a publicly accessible location, e.g. City Hall Library or Municipal Archives?
- Archival copies are intended for datasets where “the majority of the data contained within may be subject to permanent removal or replacement.” How will this be determined? How can we avoid the situation where a dataset is removed without having first been archived?
- Re: the sentence, “The manual shall also include guidelines for the division of large data sets into groups of smaller data sets, disaggregated by discrete time units, when technical barriers, archival necessity or practical concerns require such division” - users will need documentation that details how datasets were divided in order to appropriately merge datasets for analysis. For instance, Civilian Complaint Review Board and Department of Education have each provided over a hundred datasets to NYC Open Data. Many of the datasets look like they can be merged, based on the columns, but it is difficult to know for sure without documentation as to how the data was divided.

Int. No. 898 - In relation to requiring a data dictionary for every data set on the open data portal

- This bill addresses many of the concerns outlined in this testimony.
- In addition to explaining “any relationship or connection between a column and other columns within the data set,” connections across datasets on NYC Open Data and the agency website should be explained. This would help clarify the confusion across the similar reports provided by Department of Homeless Services, as presented earlier.
- All datasets should have the most recent upload date and generation date posted, not just the datasets updated less than daily (as specified in the bill). Daily datasets may not include a date column with this information, and even if they do, there may be differences across date of event, date of upload, and date of generation. These dates should also be posted on the portal, perhaps in the About tab (where there are also dates of update and creation, related to the portal itself), and not posted in the data dictionary, as suggested in the bill.
- Re: the statement “The data dictionary may also include any additional information or description that can provide context to the data...” - the data dictionary SHOULD include this information. Description of the purposes and methods of data collection is essential for research and analysis, as discussed earlier.
- Data dictionaries should follow a standard template, and be located in an easy-to-find location, e.g. in the About tab under Description.
- External review of data dictionaries is also recommended. It is essential that acronyms and technical terms are explained, as the bill requires, but government officials close to

the programs that generate open data may not realize what an outside person would consider a technical term.

- If the bill is passed, section 4.2.2 on Metadata in The Open Data Policy and Technical Standards Manual will need to be updated. The manual states that metadata for columns within a data set is not required.

Int. No. 914 - In relation to establishing response timelines for public requests on the open data portal

- My comments here are related to what is the “on-line forum” referenced in the bill. Where is it? Is it the Discuss tab or the “Contact Dataset Owner” action in the About tab? Both of these options for public feedback are located within existing datasets, and so they are not intuitive places for suggesting new datasets.
- Moreover, agency officials very rarely respond to questions asked through either avenue. In general, there should a response timeline for inquiries, similar to FOIL requests.

Int. No. 915 - In relation to the timely updating of certain public data sets on the open data portal

- Yes, NYC Open Data must contain data that is as synchronized as possible with data published on agency websites. Lack of updated data on NYC Open Data lowers its utility, and in fact, we encourage researchers not to depend on NYC Open Data and consult agency websites for the most recent data.
- Looking forward, the operations underlying NYC Open Data need to be less manual for updates to happen seamlessly, and should follow Data.gov in harvesting data automatically from agency websites for new and updated datasets.

Int. No. 916 - In relation to an open data law agency compliance audit

- A compliance audit is an excellent idea, and should include data dictionaries and responses to public inquiries as part of the audit.

## **Conclusion**

If a primary intent of NYC Open Data is to make government more transparent and accountable to New Yorkers, that goal cannot be easily achieved if the public is not also provided with sufficient information that makes the data understandable and analysis meaningful. Simply transforming government information into open data format does not add this value. The proposed bills, particularly the bill requiring data dictionaries, are essential to making NYC Open Data usable for research and analysis.

## **Appendix A: The importance of metadata & data documentation**

Metadata for data can be categorized into four types. First, study-level metadata describes the study or data collection program, including the objectives, sample or population characteristics, and variables or indicators that are being measured. Second, file-level metadata summarizes the properties of individual data files within the collection. Third, variable-level metadata describes the variables measured in the dataset, and how they are defined and coded (including column headings and values). Fourth, and last, administrative and structural metadata describes the structure and technical aspects of the data so that the data can be properly maintained and preserved over time.

It is the lack of study and variable-level metadata in order to conduct research and analysis with open data that is of concern here. The lack of this metadata is a common limitation of open data – particularly for open data that was derived from agency internal operations and was not created for the purpose of public research or evaluation.

It's important to keep in mind that government data is not a unitary whole. It derives from multiple agencies with different organizational mandates, programs, and policies. Government scientific and statistical agencies (e.g. United States Census Bureau, National Center for Health Statistics at the federal level) have long provided open data, before the phrase came into vogue. Such scientific data is produced by government to support its programs and policies, but it has also traditionally been publicly disseminated for research and other purposes. Agencies provide these datasets for public reuse, data which are derived from formal data collection efforts such as surveys and censuses. These open data have been accompanied by detailed documentation and dictionaries (i.e. metadata).

Yet under current open government data initiatives, it is not only scientific and statistical data being provided. Government agencies create information when performing its mandated activities, e.g. legislative hearings, budgets, crime monitoring, traffic violations, number and type of 311 calls made. In sum, open data as issued from different agencies cannot be treated as a homogeneous whole. Individual datasets will have their own strengths, challenges, and deficiencies. Quality metadata can aid tremendously in minimizing inaccurate understanding of open data and inaccurate interpretation of open data analysis across government agency datasets.

## Appendix B: Data Dictionary for ACS Community Partners

Dataset Name	ACS Community Partners	
Dataset URL	<a href="https://data.cityofnewyork.us/Social-Services/ACS-Community-Partners/9hyh-zkx9">https://data.cityofnewyork.us/Social-Services/ACS-Community-Partners/9hyh-zkx9</a>	
Agency	Administration for Children's Services	
Background	This is a directory of preventative agency location centers. Preventive service providers work with at-risk families to ensure that children remain safe in the home and to prevent children from entering foster care. When ACS issues requests for proposals seeking qualified vendors to provide preventive services, the providers who meet the proposal submission criteria and are awarded contracts are the ones that are included in the directory.	
	Column Names	Column Definitions
	PROGRAM_LST_ID	A unique identifier generated by ACS' system to identify each provider that has received a contract to provide preventive services.
	PROGRAM_CODE	The preventive program's vendor ID or facility ID that is registered in the State's systems
	PROGRAM_NAME	The name of the program, which is within the agency and will carry out the contract.
	ACTIVE_FLAG	A field that helps the user identify whether or not the program in question is active. Programs whose contract is current and has not been terminated will have an ACTIVE_FLAG equal to 'Y', while closed programs will an ACTIVE_FLAG equal to 'N'.
	AGENCY_NAME	The name of the agency, i.e. the provider organization that responds to

		procurements and are awarded contracts by ACS.
	DESCRIPTION	<p>Family Description - Family Treatment/Rehabilitation (FT/R) is designed to support families whose children are at imminent risk of foster care placement or replacement because of prevalent effects of parental and/or child substance abuse and/or mental illness. FT/R also serves families with an infant with a positive toxicological screen.</p> <p>General Preventive - General Preventive (GP) services are provided to families with children at risk of foster care placement in order to ensure the safety of their child(ren), to reduce risks to children, and to preserve, support, and strengthen the family, when appropriate. For children who have been placed into foster care programs, GP services programs provide community-based assistance that will help promote timely family reunification or adoption, and prevent replacement into foster care</p> <p>Intensive Preventive - Intensive Preventive and Aftercare Services for Adolescents provides intensive family support services for at-risk adolescents using evidence-based treatment models such as Multisystemic Therapy (MST) and Functional Family Therapy (FFT).</p> <p>Medically Fragile - Medically Fragile serves families with children with severe physical disabilities and/or chronic illnesses.</p>
	BOROUGH	The borough where the program is



		located.
	PROG_CONTACT_LAST_NAME	Program contact's first name
	PROG_CONTACT_FIRST_NAME	Program contact's last name
	PROG_CONTACT_PHONE_NO	Program contact's phone number
	PROG_CONTACT_FAX_NO	Program contact's fax number
	PROG_CONTACT_EXTENSION	Program contact's phone extension
	PROG_CONTACT_EMAIL	Program contact's email address
	PROGRAM_ADDRESS_1	Program's street address
	PROGRAM_ADDRESS_2	Program's street address
	PROGRAM_CITY	Program's street address - city
	PROGRAM_ZIP	Program's street address - zip code
	OFFICE_MON_FROM_TIME	Program's office start hours for Monday
	OFFICE_MON_TO_TIME	Program's office closing hours for Monday
	OFFICE_TUES_FROM_TIME	Program's office start hours for Tuesday
	OFFICE_TUES_TO_TIME	Program's office closing hours for Tuesday
	OFFICE_WED_FROM_TIME	Program's office start hours for Wednesday
	OFFICE_WED_TO_TIME	Program's office closing hours for Wednesday
	OFFICE_THURS_FROM_TIME	Program's office start hours for Thursday
	OFFICE_THURS_TO_TIME	Program's office closing hours for Thursday
	OFFICE_FRI_FROM_TIME	Program's office start hours for Friday
	OFFICE_FRI_TO_TIME	Program's office closing hours for

		Friday
	OFFICE_SAT_FROM_TIME	Program's office start hours for Saturday
	OFFICE_SAT_TO_TIME	Program's office closing hours for Saturday
	OFFICE_SUN_FROM_TIME	Program's office start hours for Sunday
	OFFICE_SUN_TO_TIME	Program's office closing hours for Sunday



**THE LEAGUE OF WOMEN  
VOTERS** OF THE CITY OF NEW YORK

4 WEST 43rd STREET, SUITE 615, NEW YORK, NY 10036  
PHONE: (212) 725-3541 • FAX: (212) 725-3443  
WWW.LWVNYC.ORG • OFFICE@LWVNYC.ORG

Testimony before the Committee on Technology

October 1, 2015 at 1:00PM

Good afternoon. My name is Catherine Gray, one of the Vice President of the League of Women Voters of the City of New York. As a multi-issue, non-partisan political organization we encourage informed and active participation in government, work to increase understanding of major policy issues, and influence public policy through advocacy and education.

For over 96 years the League has been interested in government operations and how it effects every person in the community. The path to information is now electronic data and its access is paramount to successful operations, oversight and transparency in a Democracy. We appreciate this opportunity to comment today.

The League of Women Voters of the City of New York is here to support the testimony from the Transparency Working Group (TWG). The League supports reforms to the Open Data Law (Local Law 11 of 2012), because after three years it has become apparent that the Law needs adjustments, as can be seen in the seven open data bills before this committee today.

I will reiterate the need for the three key factors that TWG has identified :

1: No legal mechanism to compel agencies to publish public data or to meet their own data publishing schedules.

2: Some of the Data sets have been found to be riddled with errors and there is no formal method of reporting the errors so that these errors can be fixed.

3. Not all data is presented in a uniform and useful manner. This addressed in Int. 898 by Council Member Vincent Gentile requiring a data dictionary for every data set on the open data portal. "The data dictionary accompanying each data set, or a link to such data dictionary" Could be improved by changing the word "or" to "both". The idea of flipping between a database dictionary and data is burdensome.

The concept of every data set having the same format for dates and addresses, and those addresses that had been geocoded with latitude, longitude, community board, and council district would be revolutionary.

The Open Data law would greatly benefit from have all six Council Members, members of the Open Data Community and the Mayor's Office of Data Analytics working on a comprehensive legislation that would make the Open Data Law more convenient to users and City Agencies.

The League applauds the Council Members: James Vacca, Ben Kallos, Vincent Gentile, Fernando Cabrera, Annabel Palma and Richie Torres for taking action to improve the Open Data Law.



**TESTIMONY OF JOSH MUMM  
OUTREACH and ADVOCACY MANAGER, COMMON CAUSE/NY  
BEFORE THE NEW YORK CITY COUNCIL COMMITTEE  
ON TECHNOLOGY  
RE: INT 916 – 2015, INT 914 – 2015, and INT 908-2015  
October 1, 2015**

Good afternoon Council. My name is Josh Mumm and I am the Outreach and Advocacy Manager at Common Cause New York. We provide a voice for citizens in support of open, honest and accountable government at all levels. We work to strengthen public participation and confidence in our institutions of government. Thank you for holding this hearing. There are a lot of bills on the table today and I am going to briefly comment on three of them.

New York City's Open Data Policy was made law at the end of 2012 and has already been implemented under two different mayoral administrations. Common Cause New York recognizes that a lot of work goes into coordinating with the city's numerous agencies and we thank the Department of Information Technology and Telecommunications (DOITT) and the Mayor's Office of Data Analytics (MODA) for their continued work in wrangling information out of agency silos.

However, the law still has a long way to go in terms of ensuring universal compliance. Common Cause NY supports [Intro 916 of 2015](#), which would mandate the Department of Investigation to perform an open data audit on nine departments, including the Department of Corrections and the NYPD.

Conducting an audit will not only be helpful in unraveling the confusion that an individual agency may have in complying with the law – it will also be a concrete way to delineate future compliance with other agencies. We strongly suggest that the New York City Economic Development Corporation and the Department of Small Business Services, also subject to FOIL, should be included in the agencies whose compliance is included in the audit. Thus far, there has been no comprehensive open data audit and it is now long due as we are almost four years into the implementation of the law.

We are also in support of the spirit of Intro 914 of 2015, though we are unsure of how it will be carried out. The lack of responsiveness to questions posed by the public on the Open Data Portal is a real cause for concern and there must be a way for concerns and queries to be answered in a systematic and trackable format.

Additionally, we want voice our support of Intro 908-2015, which would place datasets that have been successfully requested by FOIL into the Open Data Portal. Easily allowing other members of the public to view FOILED information is not only a sensible move from a transparency and good government perspective, it will also drastically cut down on repetitive FOIL requests and thus save time for the agency's FOIL officers.

We share the concern of our colleagues at Reinvent Albany for the enforceability of the Open Data Law. While we are supportive of the idea of creating a private right of action to gain compliance, we are mindful of the burden and cost associated with litigation, which makes it a seldom used tool in these situations. Perhaps an additional enforcement mechanism might be to empower the Department of Investigation, upon request, to determine whether there has been a failure of compliance and issue such a finding, in order to foster agency compliance.

As for the other bills being heard today, Common Cause NY supports their passing as they will bring added clarification to the law in terms of mandated compliance.

In conclusion, we echo our colleagues' testimony for the need to improve over-all data quality that is made available to the public and we hope that more resources will go towards staffing the Open Data Portal. Thank You.

**Date:** Thursday, 1 October 2015

**From:** John Krauss, GovLab Tech Fellow <[john@thegovlab.org](mailto:john@thegovlab.org)>

**To:** NY City Council's Committee on Technology.

**Subject:** In regards to Int. Nos. 0890-2015, 0898-2015, 0900-2015, 0908-2015, 0914-2015, 0915-2015, 0916-2015, and the oversight of the Open Data Plan.

Dear Chair and Committee Members,

Thank you for requesting that I testify today about these proposed amendments to the city's open data law. My name is John Krauss, and I am the technology fellow at the GovLab at NYU, whose mission is to "improve people's lives by changing the way we govern," including through the "increased availability and use of data."

New York City is one of the leading open data cities in the United States. By sheer number of datasets, New York City publishes more than twice as many on its portal as Chicago, which is second in number of datasets, and three times as many as San Francisco, which is third.<sup>1</sup>

The numbers are impressive, and New York, spurred by 2012's open data law, has published a large number of datasets.

What we're discussing today are a handful of improvements to that law, many of which were inspired by prior testimony from the open data community that brought flaws in the published data to light.

Within the open data community, there has been agreement for years that unanswered questions and unresolved problems in the data are impeding the use of the open data portal. To understand this issue, we downloaded all 293 comments from the portal 2011 to the present and analyzed them.<sup>2</sup> We broadly divided the comments into four categories:

1. Problems needing a fix, such as, "70 poll sites on this list have blank entries for the LOCATION field." (70 comments)
2. Questions needing an answer, such as, "Will this be updated regularly?" (106 comments)
3. Comments that need no response, such as: "Hard to trust any of this data. Very few respondents." (32 comments)
4. Responses, such as "Thank you for your feedback." (82 comments)

<sup>1</sup> Count of datasets on municipal Socrata open data portals, from <http://www.opendatacache.com>: as of 10/1/2015, New York City has 1,363 datasets, Chicago has 573, and San Francisco has 438.

<sup>2</sup> All data can be viewed at <https://docs.google.com/spreadsheets/d/1ZqIEf17h-CXnbHcOX-i4I3fp3VcPkOwrqKMB4B9bQrg/edit?usp=sharing>

Slightly less than half of the questions (51 of 106) were answered. On average, half a year (180 days) elapsed between question and answer. The average unanswered question on the portal is over a year old.

Actual problems fared worse. About two thirds (46 of 70) haven't been fixed. On average, unfixed problems were reported over a year ago. Over half (28 of 46) never even received a confirmation that the problem was being looked into. Of unfixed problems that were confirmed (18 of 46), on average over a year has elapsed since confirmation without a fix.

Int. 916 (data compliance audit), Int. 915 (portal must be up to date with website), Int. 900 (geospatial format), Int. 898 (data dictionary), and Int. 890 (data retention) all address ~~problems~~ problems brought up by open data portal users. As others testifying today will tell you, simple compliance with the Technical Standards Manual would fix most of these issues. New legislation requiring data conform to standards it was already supposed to conform to is redundant.

What is needed are people who can respond to comments by actually fixing the issues brought up in them. We need people to fix this data, not more requirements it be fixed.

The current arrangement, where DoITT manually ferries reports of problems to agencies that may or may not have resources to resolve them, does not work. We need open data expertise on the front lines, responding to these problems and empowered to fix them. This means funding for staff and resources within an organization dedicated to open data. It also means a private right of action by open data users to put them in the same class as FOIL applicants, whose requests must be responded to.

I want to emphasize that DoITT is doing their best to resolve questions and concerns. Due to deficiencies in the open data portal, keeping track of comments is difficult. Without a clear mandate to fix problems at the source, oftentimes there is nothing more that they can do than tell the user that the agency has been informed.

Thank you for your time, and please do not hesitate to contact me for more information at [john@thegovlab.org](mailto:john@thegovlab.org).  
John Krauss

## Appendix

Below are specific modifications we would recommend to the proposed legislation.

### Int. 914:



- Comments that merit a data fix should require one, or an appropriate response explaining why a fix is impossible.
- All comments and requests, as well as their responses and resolutions, should be made available in a machine-readable format.

**Int. 890:**

- Large amounts of data will still be lost if archival is only required when "a majority of the data contained" changes or is removed.
- Many datasets are updated monthly or annually. Even in cases where only a small percentage of data changes, archival copies of all these should be kept.
- For datasets updated more frequently, a changelog of removed data should be maintained using free software such as dat<sup>3</sup>. This would allow reconstruction of older data without keeping a prohibitive number of redundant copies.

---

<sup>3</sup> <http://dat-data.com/>



**THE COUNCIL  
THE CITY OF NEW YORK**

Appearance Card

I intend to appear and speak on Int. No. 908, 914, 916 Res. No. \_\_\_\_\_

in favor  in opposition

Date: 10/1/15

(PLEASE PRINT)

Name: Josh Mumm

Address: \_\_\_\_\_

I represent: Common Cause NY

Address: 80 Broad St., NY, NY 10007

**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: 10/1/2015

(PLEASE PRINT)

Name: Catherine Gray

Address: 936 PRESIDENT STREET

I represent: League of Women Voters

Address: Transparency Working Group

**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: 10/2/2015

(PLEASE PRINT)

Name: Albert Webber

Address: \_\_\_\_\_

I represent: DOJ TT

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: 10/1/15

(PLEASE PRINT)

Name: Amen Ra Washariki

Address: \_\_\_\_\_

I represent: Mayor's Office of Data Analytics

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: \_\_\_\_\_

(PLEASE PRINT)

Name: John Kaehny

Address: \_\_\_\_\_

I represent: Reinvent Albany

Address: \_\_\_\_\_

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

890,898,  
914,915,916

**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: 10/1/15

(PLEASE PRINT)

Name: Tara Datschen

Address: 2 W 120 St NY NY 10027

I represent: Columbia University

Address: 1255 Amsterdam ave ny ny 10027

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: \_\_\_\_\_

(PLEASE PRINT)

Name: JOHN KRAUSS

Address: 993 CARROLL ST APT C2 BROOKLYN NY

I represent: The GovLab 11225

Address: 2 Metropolitan Roadway 9th floor BROOKLYN

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