

NYC Transparency Working Group

Citizens Union ▪ Common Cause NY ▪ League of Women Voters NYC
NY Public Interest Research Group ▪ New York Civil Liberties Union
BetaNYC ▪ OpenPlans ▪ Reinvent Albany ▪ Women's City Club NYC

NYC Open Data Law Progress and Challenges

November 19, 2013

1. Executive Summary

We strongly support the NYC Open Data Law and its intent. The NYC Open Data Law is widely considered one of the best in the world. We strongly urge the council and next administration to fully fund and support its implementation. To this end, we welcome this oversight hearing and urge council to schedule more, especially after the release of the annual Agency Compliance Plans in July 2014.

The Open Data Law is very important because it establishes standards for data formats and availability which the city's digital information vastly easier to reuse. In particular it defines what open data is – which is basically digital information that is findable on the internet and usable with desktop office software. Plus, the law calls for “data streaming” features like Application Programming Interfaces (APIs) which make data available in the same way that Netflix, or Amazon Instant or iTunes do for movies, compared to having to visit a video store.

Broadly, the Open Data Law has been a success. It has led to the release of long sought land use data sets like PLUTO and ACRIS in easy to use formats, as well as real estate databases and other data sets of great public interest like 311 complaints. The City's open data site is being used by members of the public, advocates, developers, and journalists – the NY Times and National Public Radio have regularly credited data from NYC's data portal. City agencies are starting to use

“The City's open data site is being used by members of the public, advocates, developers, and journalists.”

the city's Open Data Portal for their own easy data retrieval. The best example may be the TLC, which publishes data on medallion vehicles and drivers. Very importantly, NYC open data is also being used by the apps developer community in apps like Roadify, Yelp, or NYC Building Violations to make government information more easily accessible to the public.

Implementation of the Open Data Law needs improvement. Many city agencies did not submit a mandated Agency Compliance Plan. Of those that did, many gave themselves a multi-year extension on releasing data that was supposed to have been released on March 13, 2013. (Deadline for publishing data currently on city websites in a non-open format.) Agencies released about half of the data sets

that should have been released on March 13th. Additionally, important city data, especially rules and reports, are still only available in closed formats or for sale in apparent contradiction of the Open Data Law. Lastly, many agencies, including the NYPD, have done little to make basic data, such as crimes and traffic crashes, available in open formats – though most major American cities already do.

The Open Data Portal has not yet evolved to the point where it is useful to the average New Yorker. The portal itself is somewhat difficult to use. The site’s search function is not very good, and most people are not adept at using raw data. This is to be expected, because the basic idea behind the portal is to make data available to super-users who will convert it into mobile apps and easy to use reports or articles. The greater concern is that there is a mismatch between what’s on the portal and what the public wants. The city does not appear to be prioritizing data releases based on agency website usage, FOIL, or 311 requests.

2. Overview

The NYC Transparency Working Group strongly supports the NYC Open Data Law and believes it has, overall, been a big success. The broad intent of the law is being realized, and it is achieving its goal of pushing City Hall and agencies to make much more data available. Our groups very strongly believe that the DOITT and City Hall staff time needed to implement the Open Data Law should continue to be fully funded.

Prior to the Open Data Law, there was no mandate for city agencies to proactively share their data with the public. That data sharing mandate has led to the release of hundreds of new data sets, including the PLUTO and ACRIS data which have long been sought by planners and researchers, who use those data sets to further their research on affordable housing and transportation issues.

“New York City’s Open Data Law is widely considered a global best practice.”

Additionally, our groups appreciate the power of the API (Application Programming Interface) feature on the city’s data portal, which essentially streams data to public users, powering countless mobile apps. The potential of this feature to link data to users in other agencies, levels of government, and the public is barely be-

ing realized. However, it is being recognized. The Open Data Law is widely considered a global best practice, and has drawn government officials from Tokyo, Berlin, and the United Kingdom to visit and speak with the New Yorkers who helped create it.

3. Intent

Fundamentally, the Open Data Law had two important goals: First, to make the City's giant treasure trove of data available to the public – making all public data available in open formats to enable reuse. Second, to help propel our City government into the Data Age, by compelling it to inventory and assess its data in a methodical and timely way. To accomplish these, the city must settle on data standards and methods for sharing that data both with the public and with other agencies and levels of government. Ideally, sometime soon, agencies will become among the biggest users of the city's data, as they use data from other agencies or other parts of their own agency, or from state and federal data.

The Open Data Law has achieved much of what the Commission on Public Information and Communication (COPIC) aimed to do some decades ago, by forcing agencies to inventory their data, and creating an impetus for the Mayor's efforts to rationalize data management and sharing.

4. Highlights

Broadly, the Open Data Law has been a success in getting city data published in an open format. For example:

High value data sets are being released.

The Open Data Law resulted in the release of a number of data sets long sought by the public including PLUTO, ACRIS, and 311 complaints. This is a big win for open information advocates.

Agencies are using the open data portal for internal purposes.

It was hoped that the Open Data Law would lead to some city agencies using the data portal to publicly publish data which the agencies use in the course of their everyday work. The Taxi and Limousine Commission was the first to do this, and has published data sets on authorized medallion vehicles and drivers.

Journalists are using the open data portal.

The NY Times, National Public Radio and Guardian have been regularly crediting NYC Open Data for data sets which power their interactive maps. In particular, they both used flood and damage maps housed in NYC Open Data after Hurricane Sandy. Additionally, the NY World frequently uses the data portal.

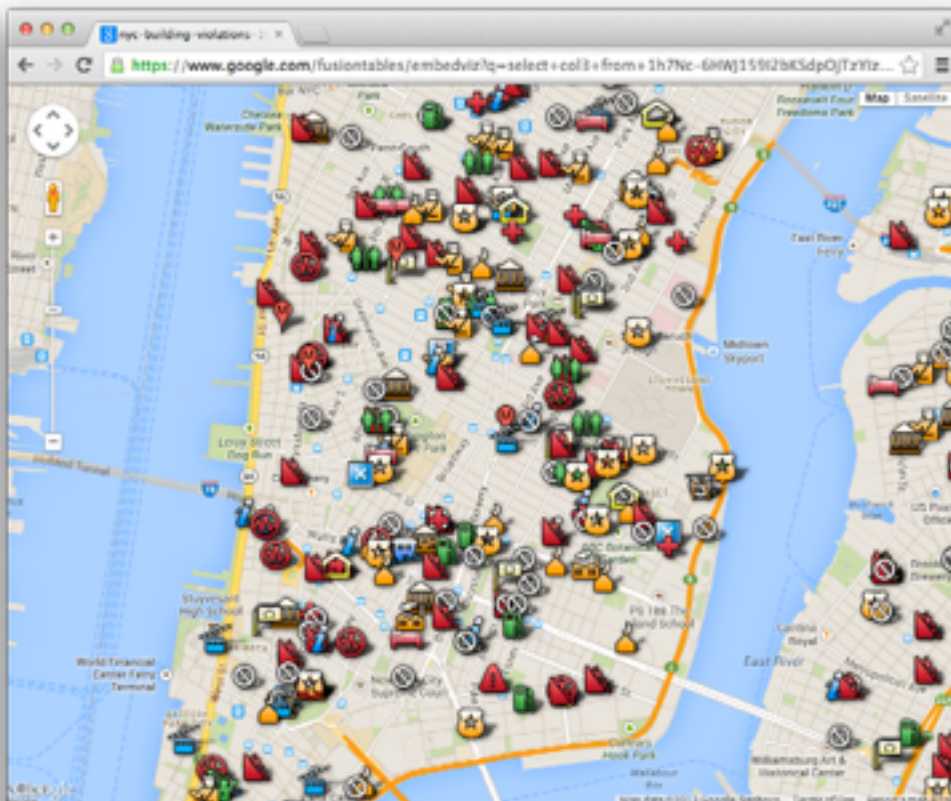
Academics are using the open data portal

The PLUTO land-use data set was immediately put to use by academic researchers at CUNY, Pratt Institute, and elsewhere. Prior to its publication in the open data portal, PLUTO data cost thousands of dollars a year in license fees. Additionally, the ACRIS real estate data set was also recently liberated and has been heavily used by NYC Furman Center and other researchers.

Developers are using the open data portal to build apps for public use.

Numerous mobile apps have been built using the Open Data portal. There are also many commercial applications for open data powering publicly useful tools. For example, a third party map of current building violations issued by NYC DOB:

Source: BuildVio, Building Violations History Reports. <http://buildingviolation.com>



Additionally, NYC's BigApps, an annual software competition held for developers of online and mobile applications using official City data sets, has grown every year. As of the fourth iteration of BigApps, nearly 300 apps have been created using data from the NYC open data portal. The NYC Transparency Working Group believes BigApps serves the dual purpose of encouraging the tech startup community, as well as highlighting demand for open data from City agencies.

5. Needs Improvement

While the Open Data Law has been broadly successful, its implementation has areas which still need improving. For example:

The Big Picture

Use 311, FOIL, and other public requests for information to determine what data the public wants. To date, data published on the open data portal has been released to fulfill the March 2013 mandate, but more in response to Big Apps competitions than to public interest. Moreover, it appears to that DOITT and the Mayor's Office have not even analyzed public requests from FOIL, 311, and agency website forms and analytics to determine what data is in greatest demand. DOITT and the Mayor's Office should be mindful of the public's interest in certain data sets during their management of the city's digital property.

The Bigger Picture

Performance and AgencyStat data is still not published on the open data portal. The Bloomberg administration has worked hard to collect performance data and put it to work improving agency operations. However, very little of that highly-refined data has been made available to the public or City Council via the open data portal.

“Little highly-refined agency performance data has been made available to the public.”

For instance, the Mayor's Office has gathered and mapped all of the data on sewer backups, a big issue in Queens in particular; this map was shown at a recent conference at NYU, but it's not in the open data portal. Additionally, AgencyStat data such as HousingStat, NYCHA's performance management process (which includes real-time updates of broken door locks and broken elevators in public

housing) is not available to either the public or council. This data is some of the most highly sought after, and the Open Data Law cannot succeed without data like this being published.

Agencies must stop selling data and make it free

There are 32 publications for sale at the NYC City Store that are easier to purchase than to download from the open data portal. These include: 10 Surveys and Reports and 22 Rules and Regulations. Contained in these reports is a large amount of tabular data which is neither machine-readable nor reusable. This includes NYC's Building Codes, which are composed digitally, published to paper, and finally posted online with the information locked in PDF documents.

However, on November 7, 2013, the NYC Rent Guidelines Board started selling the latest edition of *HousingNYC: Rents, Markets and Trends*. This publication is sold in the NYC CityStore for \$30. (The soft bound book is available for purchase for an extra \$20.) As if adding insult to injury, the Board specifically states the data "is available for sale exclusively from CityStore in both book and CD format, but is not available as a free download."

Agencies must publish data in a machine readable format

The Open Data Law clearly requires agencies to produce machine readable data in a CSV or XLS format that allows for easy use. However, some agencies are publishing data in ways that make it impossible to use. For example, the NYPD is publishing spreadsheets that carry over the same malformed layout as their PDF versions. The NYPD states this is to protect the integrity of the data, but their choice of formatting makes this data effectively impossible to use without a large amount of time and manual data processing.

"Some agencies are publishing data in ways that make it impossible to use."

Clarify which agencies are subject to the Open Data Law

There is some confusion over which State chartered authorities in the city are subject to the Open Data Law. It would be helpful to have a decision from the Law Department about whether, for example, the NYC Board of Elections, Rent Guidelines Board, NYC Housing Authority, or the Roosevelt Island Development Corporation fall under the law.

Likewise, private companies or non-profits that operate as city chartered franchises, e.g. the NYC Bike Share Company (“CitiBike”) and the NYC Technology Development Corporation, are neither clearly subject to nor exempt from the Open Data Law. The murky legal status of all these entities should be cleared up by an official position from the Law Department.

Create accountability for statutory deadlines

Agencies largely missed the March 2013 deadline for publishing data sets already on their websites in PDF or other closed format. We estimate they only liberated 40% of the data sets that should have been posted by this deadline. Additionally, many of the data sets scheduled for future release under the Agency Compliance Plans should have been released in March 2013. Essentially, the agencies have given themselves an extension.

Many agencies failed to participate in the agency compliance plans, including the following: Board of Standards and Appeals, Campaign Finance Board, Civilian Complaint Review Board, Office of the City Clerk, and the Commission to Combat Police Corruption, possibly among others.

Improve the portal’s technical deficiencies

There are widespread complaints about the Search function on the open data portal. Even technologically proficient users find it very difficult to locate the data they want. Technologists have repeatedly pointed to deficiencies in the metadata, (the descriptive information attached to each data set) which greatly reduces the usefulness of the search function.

“A public right of action would hugely increase the power of the Open Data Law.”

We strongly recommend that DOITT convene a public Search/Metadata Working Group to help address this problem. There are many technologists among the public who would welcome a chance to help make the platform more useful.

Strengthen the Open Data Law

The law would be enormously strengthened by a public right of action. As part of the negotiation over the Open Data Law, the Law Department had a public right of action removed from proposed legislation. This means that when city agencies

miss deadlines, there is no public recourse except applying pressure in the press or via City Council. A public right of action would hugely increase the power of the Open Data Law by actually compelling agencies to comply with it.

6. Future Concerns and Questions

1. Loss of funding for DOITT staff implementing the Open Data Law.
2. End to leadership from City Hall from the Mayor's Office of Analytics.
3. Few automatically updating data sets. Too many data sets must be updated by hand, which is expensive and laborious.

Our groups will continue to develop questions along these lines:

1. When will the Open Data dashboard, which is called for in the Technical Standards Manual, be made available?
2. Does the administration look at website analytics, FOIL requests, and 311 requests when deciding which data sets to prioritize for release?
3. Has the administration used the Open Data Law to reduce the number of FOIL requests city agencies receive? Does the City know what FOIL requests agencies are receiving? Do their responses contain tabular data suitable for publication on the Open Data Portal?
4. Is the Open Data Law helping to reduce 311 requests?
5. Are agencies using the open data portal in the course of their everyday work? Are other levels of government using the city's data portal?
6. Are there plans to publish VENDEX data and the Doing Business With data in the open data portal where it can be accessed via API? Why or why not?
7. Why do City agencies, the latest being the Rent Guidelines Board, continue to charge for data that taxpayers already paid for?
8. Can the Law Department clarify which City and State authorities and boards are subject to the Open Data Law?

Appendix I: Data Sets Requested for Priority Publication

Title	Description	Online	Agency
Doing Business Database	The Doing Business Search provides access to information on entities and individuals that do business with the City of New York.	http://www.nyc.gov/portal/site/DBusinessSite	Mayor's Office of Contracts
Suggest a Dataset Dataset	A list of datasets suggested/requested by users of the NYC Open Data Portal.	https://data.cityofnewyork.us/nominate	DOITT
Health Department Licenses	Licenses in 11 major categories including child care providers, food vendors, facilities that use radioactive materials	http://www.nyc.gov/html/doh/html/licenses/licenses-home.shtml	Department of Health and Mental Hygiene
Citywide Facility Inventory Database required by Local Law 26 of 1988.	includes information on 8,800 facilities citywide that store hazardous substances (this is separate from and in addition to the federal Toxic Release Inventory program);	http://www.nyc.gov/html/dep/pdf/derta/dep_hazmat_annual_report_2012.pdf	DEP
Inspections by DEP of hazardous substance storage facilities	Over 9k such inspections in 2012 alone, which resulted in 672 Notices of Violation (NOVs) for noncompliance with reporting requirements;	http://www.nyc.gov/html/dep/pdf/derta/dep_hazmat_annual_report_2012.pdf	DEP
Risk management plans submitted to DEP	Over 300 based on Local Law 92 of 1993	http://www.nyc.gov/html/dep/pdf/derta/dep_hazmat_annual_report_2012.pdf	DEP
Incidents of hazardous material releases	- 2,430 in 2012 of which 33.7 percent were responses to reports of chemical odors; 25.6 percent were petroleum spills; 15.3 percent were abandoned chemicals 10 percent were responses to chemical spills and 9.3 percent were responses to special investigations.	http://www.nyc.gov/html/dep/pdf/derta/dep_hazmat_annual_report_2012.pdf	DEP
NYPD City Wide Crime Statistics "Compstat"	Currently, the most recent weekly report on criminal statistics in NYC is available by PDF, with no useful archive to look back through. All statistics recorded for every week since the inception of the program should be open to the public in a real machine-readable format.	http://www.nyc.gov/html/nypd/html/crime_prevention/crime_statistics.shtml	NYPD
NYPD Traffic Crash Statistics			NYPD
NYC Lobbyist Search	The NYC Lobbyist Search provides detailed information about lobbyists, registered with the City of New York, and their clients. It is currently only available as a search portal.	http://www.nyc.gov/lobbyistsearch/	Office of the City Clerk
Legislative Research Center	The activity of the New York City Council. It is provided only via a portal, with no data format offered other than page-views.	http://legistar.council.nyc.gov/Calendar.aspx	The New York City Council
Vendex data on public contractors	This information is extensive and complex, but elements of it would be interesting and useful for application developers and the general public, such as contracts given to businesses registered in the city vs. outside the city, types of contracts by vendor, vendors by agency, etc. At this point, you need to visit the "Public Access Center" at 253 Broadway to view the info		MOCS

Data on facilities and sites that have may environmental impacts	City, state, and federal regulations require DEP to maintain a wealth of data on each of these, including location, type of environmental impact, amount of material such as hazardous waste or wastewater discharge, history of environmental impact, etc. Though state and federal environmental agencies also collect and maintain similar data, DEP often has unique data not included in these state and federal files such as the City's "industrial pretreatment program" or CSO locations. DEP has made some information available through the Open Data Portal, but not the info described above.		DEP
Full BIS database	The Department of Buildings Information System (BIS) provides a wealth of information of keen interest to the public and to anyone hoping to understand development trends, building safety, and how the City interfaces with the real estate industry in general. BIS includes information on building applications, certificates of occupancy, inspections, (final construction, plumbing, electrical, and elevator), complaints, violations, safety reports, licensing data for contractors, and more. DOB provides excerpts from BIS in spreadsheet format for certain items, but not the detailed BIS information.	http://www.nyc.gov/html/dob/html/codes_and_reference_materials/statistics.shtml	DOB
NYC Landmarks: individuals sites and districts.	The landmarks have been part of the Open Data Portal for a couple of years, but they haven't been updated on the Portal since at least 2010 (some of the data sets at the Portal indicate they're from 2009). Yet the Landmarks Preservation Commission has designated many new landmarks and historic districts since then.		NYC Landmarks
The New York City Loft Board Interim Multiple Dwellings database	Although there are relatively few of these – 328 as of October 2013 – loft living is a subject of interest to residents, urban planners, and others. The Loft Board provides a list of IMDs on its website, but only in PDF format and the list omits other information such as violations. The list and related data should be publicly available in electronic format.	www.nyc.gov/html/loft/downloads/pdf/imd_buildings.pdf	Loft Board
Checkbook NYC 2.0	The newly released Checkbook NYC 2.0 website provides unprecedented access to view and track how New York City government spends its nearly \$70 billion annual budget. With a maximum of 200,000 records available for download from 46,048,876 available records (as of today), it is not useful outside of the dashboard provided by the Office of the Comptroller.	http://www.checkbooknyc.com/	Office of the Comptroller - City of New York

Data on historical property sales in the city.	Recently DOF has provided free downloadable data for recent sales, but only back to 2003. The Department's internal data systems maintain this information at least back to the 1980s, and the full set of historical data should be made available.		The Department of Finance
LPC Released Data: Landmark List, Historic Districts, and Scenic Boundaries data	Released but should be put on Open Data Portal sooner than six months to produce the data to the City's open data portal		LPC
Public Library's catalogue and request records	Unreleased		NYPL
Museum Attendance and Revenue	Unreleased		Mayor's Office
Transit Adjudication Bureau Data	Appeals for violations including but not limited to SBS Fare Evasion, Fare Evasion on the system, and or other violations related to use of the MTA system. Rate of cases successfully appealed. It is unclear if TAB is NYS or NYC - all tickets are issued by NYPD transit cops.		TAB/MTA
Roosevelt Island Dev Corp.	Unreleased		RIDC
Battery Park Dev Corp.	Unreleased		BPDC
Parking regulation website	Released in open formats	http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#ParkingReg	DOT
listing of NYC Executive social media outlets		http://www.nyc.gov/html/dot/html/about/datafeeds.shtml#ParkingReg	MOME
Business Improvement District Data	Released	http://www.nyc.gov/html/ops/html/data/street_bid.shtml	SBS
City Record Online	Unreleased. Should be released before 31 Dec 2014		DCAS
Office of Citywide Purchasing	Unreleased. Should be released before 2015/2016 respectively		DCAS
DOT Released Data	Released but continues to host a number of data files on their website without the basic element of linking them from the NYC Open Data portal to their own. Additionally, DOT then states that they will update or add this "data" to the data portal by 1 Aug 2014.		DOT
NYPD Unreleased Data:	Precinct Captians, Community Affairs Officers		NYPD
NYPD Unreleased Data:	Community Affairs programs or calendars.		NYPD
"Murder in NYC" Report	Released but as a written report and not on the City's Open Data portal. The data submitted in this report should be in an open format and not in a paper/PDF'ed report.		NYPD

Compstat's weekly records.	Released but needs to be published in open, "un-pdf" copy. This is not mentioned in the "Open Data Plan."		NYPD
RFP Data	Unreleased RFP should be out sooner than 31 Dec 2014		SCA
City chartered franchise agreements (e.g. Citibike),	those entities should comply with local law 11 of 2012	http://citibikenyc.com/system-data	
Enrollment Capacity Utilization from School Construction Authority	Currently released as a PDF. There are currently three different reports which are just the raw data organized by different columns. Release data for all years and all columns in open format.	http://www.nycsca.org/Community/CapitalPlanManagementReportsData/Pages/EnrollmentCapacityUtilization.aspx	BOE
Community Board Districts and Numbers	Released	http://www.nyc.gov/html/cau/html/cb/manhattan.shtml	Mayor
Community Board Contact Information	Released	http://www.nyc.gov/html/cau/html/cb/manhattan.shtml	Mayor
Business Improvement District GeoFiles	Released	http://nycdataset.pediacities.com/dataset/nyc-business-improvement-districts/resource/47512ee9-711d-467f-a692-b5d38530e65b	SBS
Comments on the Data Portal	A listing of registered users comments on the Data Portal		DOITT
Street cleaning files	Cleaning routes are not published though alternate side of street parking calendars are		DOT
DOT Unreleased Data	Unreleased. Already crunches crash data around its SafeRoute for schools. This is the same data that would help community boards and safe streets advocates plan and advocate for safer streets for all.		DOT
Community Board District GeoFiles	Unreleased		DCP
NYCEDC Data	Unreleased		NYCEDC
Polling Locations	Released but not in machine readable format		NYCBOE
Bike counts	Released but recommended to be done quarterly		DCP
DCP Released Data	Released but should stop making apps (aka geosupport) desktop applications and start producing universally accessible web data tools.		DCP
Greenbook	Released. Data Portal release date in 2014 too late		DCAS
Entry Point LCR Monitoring Results	Released but recorded on a daily basis but will only be released as an annual data set. They should be released in real time.		DEP

Distribution Sites LCR Monitoring Results	Released but recorded on a monthly basis but will only be released as an annual data set. They should be released in real time.		DEP
Open Balance Detail	Released as "mainframe file". Should be in commonly accessible format and be published at a greater frequency than "Semi-Annually"		DOF
Licensed cigarette wholesale dealers	Released as "mainframe file". Should be in commonly accessible format and be published at a greater frequency than "Semi-Annually"		DOF
stamping agents	Released as "mainframe file". Should be in commonly accessible format and be published at a greater frequency than "Semi-Annually"		DOF
vending machine operators	Released as "mainframe file". Should be in commonly accessible format and be published at a greater frequency than "Semi-Annually"		DOF
DOE Unreleased Data	Unreleased. Besides what has been published does not plan to publish another data set until 31 Dec 2018.		DOE
Digital Payments Received data files	Unreleased. Slated to be published on 31 Dec 2018 and their publish frequency has yet to be determined. Should be released sooner and a reasonable frequency decided		DOF
Mobile Food Vending			DOHMH
Rodent Data			DOHMH
Radiation Data			DOHMH
Childcare Inspections			DOHMH
Beach Water Quality			DOHMH
Air Quality, Pool Quality			DOHMH
DOT Released Data	Released but should be posted in machine readable feeds		DOT
Fire Incident Data	Unreleased. Scheduled to be produced quarterly starting 1 Jan 2016. This seems to be a fairly generous date since the FDNY is currently producing Community Board Incident Counts. Release sooner.		FDNY
Notify NYC Notification Data	Released. Already producing daily public notifications, why would they produce their notifications on a quarterly basis?		OEM
ReadyNY events	Released but currently batched in a quarterly package. They should be daily and be linked to the current Calendar feed that the City is already producing.		OEM
311 request	On Data Portal		311
Potholes	Already relaeased		311
List of High Schools attended by Middle Schoolers	Used to be online. Currently unreleased?		DOE
District Map Files for all Levels	Unreleased		NYCBOE
NYC Voter File	Released but must be FOILed		NYCBOE

Borough President Community Events Calendar	Released	http://www.queensbp.org/clients/queens_events/list_queens_free2.asp	BP
Borough President Events Calendar	Released	http://www.queensbp.org/clients/queens_events/queens_calendar.asp	BP
Borough President one-off reports	Released	http://www.queensbp.org/content_web/Health_Human/healthreport_2006.pdf	BP
Public Advocate's Office Report Data	Released	http://pubadvocate.nyc.gov/	PA
Queens Borough President Business Resources	Released	http://www.queensbp.org/content_web/econ_dev/rg_biz_officespace.asp	QNS BP
listing of NYC Legislative social media outlets	Unreleased		City Council
Borough President List of Budget (Capital and Expense) and Grant applications and disbursements	Unreleased		BP
Community Board Applicants and Appointments	Unreleased		BP
Borough President Topographical Records	Unreleased		
Assets Under Management	This is information on the NYC Pension investments but it is not very good, and there does not appear to be good or common data about the five major funds.	http://comptrollernyc.com/bureaus/bam/assets.shtm	Comptroller's Bureau of Asset Management ADVISOR TO NEW YORK CITY PENSION FUNDS
Community Board Data: files, votes, records, etc	Status Unknown		Community Boards
Released MTA Data	Point to / linked to the city data portal		MTA
Released state or multi-state agency datasets; specifically port authority, and private transit operators that provide services in the city, port authority data (bolt bus, ctown bus)	Point to / linked to the city data portal		Multiple
City council discretionary funds	Online on City Council Site, not in Open Data Portal		City Council

Business Improvement District List with Websites	Released		
Department of Consumer Affairs Specific Fine Data (finee name, location, amount; other data already released)	Unreleased		DCA
	would love to see data around waste and waste collection and removal in NYC. where it goes. how trucks travel. waste characterization.		
DNS Data Query Error Log	<p>The DNS Data Log. The DNS Data Query Error Log</p> <p>Every request to be connected to a domain name within the .nyc top level domain will result in a notation in a Data Query Log, part of the Internet's Domain Name System or DNS. These log entries offer the potential to be a twitteresque pulse of the city. But caution is advised here as issues of privacy and security are associated with this big data trove.</p> <p>See more on this data at http://www.coactivate.org/projects/campaign-for.nyc/dns-data-query-log</p>		DoITT
Police reports	Arrests, age, gender, race, location		
CUNY Enrollment	some enrollment data are available in PDF, but it is mostly totals by campus	http://www.cuny.edu/about/administration/offices/ira/ir/data-book/current/enrollment.html	CUNY
CUNY Financial Aid Awards (TAP & PELL)	how many students are getting financial aid, what is the average award		CUNY
CUNY Student Loans	the breaddown of private and federal student loans. How many students get loans, what is the average loan amount, what are the loan rates; debt load after graduation.		CUNY
CUNY Graduation Rates	graduation rates by campus, length of study		CUNY
CUNY drop out rates	drop out rates by campus, length of study before drop out, reasons of drop outs		CUNY
Lead poisoning screening data			Dept of Health
lead inspection			HPD
lead safe work practices certifications			DOH
lead remediation work			DOH
Fire data, including fires suspected as caused by cigarettes.			FDNY

CSO Overflow Data	Overflow maintenance, rate of overflow, development and design benchmarking, implementation and distribution of load across water facilities.		DEP
DCP hi-res population data	Hi-resolution population data from the Dept of City Planning Population Division		DCP
Appointment Calendar	Appointment calendars of high-level city officials. Similar to how the President's whitehouse calendar is shared (http://www.whitehouse.gov/schedule/complete)		
Website Analytics	Website analytics of major NYC properties.		DoITT?
CB meeting transcripts	Community Board meeting minutes/transcripts as machine-readable data		Community Boards
CitiBike GPS traces	CitiBike GPS traces. During the inaugural CitiBike hack night, a CitiBike rep confirmed that GPS traces are captured by the system. It would be nice if anonymized traces are made available as open data.		DOT/CitiBike
TLC trip data	TLC trip data. Some of this data has been released to research/academic institutions. It would be great if its released as Open Data as well		TLC
NYC Green Book	Greenbook machine-readable data. For the first time, an online version of the Green Book was published in 2012. But the data is trapped in a MS-backed CMS. It would be nice if its released as machine-readable data.	http://a856-gbol.nyc.gov/GBOLWebsite/home.aspx	Citywide Administrative Services
Hpd and city agency real estate loan agreements	A constantly updated list of all buildings and their landlords that are given loans through hpd or other city agencies.		Hpd and other
Historical PLUTO dataset	Historical PLUTO data. Steven Romalewski during BetaNYC PLUTO night confirmed that he has PLUTO archives going back decades. However, they are still under the old licensing regime which prohibits him from releasing it. Either Steven is given a waiver to release the datasets of the City itself release all the historical data.		DCP