

Notes from the GISMO Open Data Event: Weds, March 6, 2019, 4pm to 6pm

Attendees:

Matsue Iwata, NYC MODA
Amen Mashariki, UrbInt
Amy Jeu, Hunter College, GISMO Board
Emily Goldman, BetaNYC*
Josh Lieberman, Open Geospatial Consortium (OGC)
Paul Schiffer, EDR (Environmental Data)
Richard White, EDR
Dara Mendeloff, Columbia University, GISMO Board
Jack Eichenbaum, GISMO Board*
Albert Webber, NYC DOITT
Asif Zaman, NYC DEP
Jigar Talati, NYC DEP
Marija Drobnjak, Citizens Committee for Children (CCC)*
Varun Adibhatla, UrbInt
Alan Leidner, GISMO President
*provided written comments in appendix

Unable to Attend But Provided Written Comments:

Aileen Gemma Smith, Vizalytics
Hector Tarrido-Picart, Vivacity

Opening Remarks: Alan Leidner

Most City data sets to be found in MODA's Data Portal and DCP's Bytes of the Big Apple have at least one spatial attribute (address, BIN, Parcel, x,y coordinates). Through the efforts of DCP, DEP and DOITT the City has created a robust geospatial foundation that provides accurate and comprehensive location identity to any point or address in New York. MODA and DCP have done a great job through successive administrations making thousands of datasets available to the public. The great volume of open data creates enormous opportunities for all sectors (private, academic, non-profit, academic, community groups, individuals) to use spatially enabled information in powerful ways. In the future we can see vast increases in geo-enabled data (sensors, IoT, social media) and corresponding increases in value created by the analysis of this data and by the development of new spatially enabled applications. The common spatial element in all the data bases is the key to making the data interoperable and therefore able to be used in combinations, across data silos. This Open Data Event is intended to examine the current use of NYC Open Data and to explore ways that this enormous asset can be made even more valuable. We would like to thank MODA for selecting GISMO as an Open Data event site.

Opening Remarks: Mitsue Iwata

The Open Data Portal currently has more than 2,000 data sets with more on the way. MODA is interested in knowing who is accessing Portal data and understanding the value being derived from Portal data use. ([Follow this link to a video of Mitsue's presentation](#)) ([Follow this link to Mitsue's slides](#))

Summary of Written Comments from Hector Tarrido-Picart, Vivacity

Vivacity, a startup company uses the NYC Data Portal to help customers navigate real estate related processes related to taxation, ownership and sales. NYC is ahead of the curve when it comes to data delivery and transparency, but much work still needs to be done. Vivacity believes weaknesses in the Portal can be solved by implementing the Los Angeles Hub model. NYC should aim to standardize all its data sets and include a common reference identification number across them. All data needs to be tied to a geospatial framework which will enable them to be represented on a map. LA has shown that this vision opens the door for startups to stand up better market solutions. Data is the foundation of our future. (Full text and diagram in appendix)

Summary of Written Comments from Aileen Gemma Smith, Vizalytics

Vizalytics, founded in 2012 following Superstorm Sandy, to solve problems faced by senior citizens and local shopkeepers. Open Data Portal was cumbersome and assumed a level of digital fluency and technical knowledge. Mind My Business app released in 2014 serving shopkeepers, and won a NYC Big Apps award. Vizalytics partnered with the Mayors Office of Technology to create neighborhoods.nyc to provide New Yorkers with realtime data at the neighborhood level. Ultimately funding for this app was discontinued. Open data is very promising but there is work to be done: data interoperability, better meta data, historical data and more reliable real time updates. Vizalytics has provided detailed technical comments and is available for further input. Hopes the City will consider the information needs of all New Yorkers. Not everyone has the time or interest in learning how to use raw data sets. Vizalytics is now operating in U.S. cities and internationally. It is an example of how an open data centric company can start in NYC and branch outwards. (Full text in appendix)

Discussion

Amen Mashariki, Urblnt: Skeptical about the value of the LA Hub. Feels it was implemented with no real use cases in mind.

Paul Schiffer and Richard White, HDR: Their business depends upon tax parcel related data provided by the NYC Data Portal. Curious about the difference between the data released and NYC "official" data. Mitsue reports that the data made available is used operationally by the City. There is provision on the Portal to request changes and improvements. Josh Lieberman of OGC noted that the State of Massachusetts has standardized parcel data across the State. Alan

Leidner stated that NYS has done the same, and has also created a state-wide data portal that includes state data about NYC which City users can tap into. <https://data.ny.gov/>
Alan Leidner, GISMO President: Wondered whether MODA/DOITT's Open Data Portal and DCP's Bytes of the Big Apple <https://www1.nyc.gov/site/planning/data-maps/open-data.page#pluto> are becoming part of a unified open data environment which might also include NYS data and other government open data portals in the region.

Marija Drobnjak, Citizens Committee for Children (CCC), <https://www.ccnyc.org/> :
Wonders which data is more accurate – agency data or portal data. Mitsue and Albert Webber respond: Data is pushed up from agencies which often have their own websites. DOITT and MODA are now working on improving data quality and increasing the frequency of update.

EDR <https://edrnet.com/> : The packaging of open environmentally related data is the foundation of EDR services. Knows about four or five other startups thriving on the use of Open Data. Open data is creating jobs. Wonders how, after creating a service based on data from one jurisdiction, you then go to other cities, if each city has its own data models and its own rules for sharing data. Alan Leidner suggests that the Open Geospatial Consortium (OGC) creates data standards. Government adoption (Municipal, County, State and Federal) of OGC standardized data formats would help companies more easily expand nationally and internationally. Mitsue mentions that the Open Portal also offers analytics and tools for users.

Amen Mashariki, UrbInt <https://urbint.com/> : Against governments trying to raise revenue by selling their data. Understands that some governments are unhappy about giving their data away for free but believes that the economic and social benefit of sharing data provides significantly greater benefits. Selling data for a price inhibits its use. Jack Eichenbaum noted that Nassau County continues to charge for data which discourages use and sharing.

Emily Goldman, BetaNYC <https://beta.nyc/> : Discussed the difference between data made public and the data that the City continued to hold privately. Wanted more guidance on how decisions are made about which data sets could be made available and which could not. Remarkd that there is a significant learning curve when teaching students to use Open Data. Even college students find it difficult. Wondered how you could get non-technical citizens to feel comfortable. Currently, using the spatial data on the portal requires the knowledge of a data scientist and so is only useful for professionals who can “download the shapefiles” into GIS software.

Azif Zaman, NYC DEP: Has taught at Bronx Community College which promotes GIS learning. Has found that students in middle school and high school pick up data use practices quickly. Wonders who is responsible for Open Data sets and what happens if there are mistakes in the data: who is liable should a mistake be made based on inaccurate data. Also concerned about

making sure that sensitive data sets such as infrastructure data, are kept secure. Wonders if there are laws to protect agencies from having to share their data.

Jigar Talati, NYC DEP: Agencies hampered by wanting to retain ownership of their data. This leads to difficulty sharing data between agencies. Right now there is a time consuming bureaucratic process that must be followed to share data between City agencies. Often, the Open Data Portal, serves as an easy way to enable interagency data exchanges. Believes that the security features of state-of-the-art RDBMS' can provide different levels of data access permission greatly reducing security concerns. Often the need for security is based on scale/level of detail. This can be modified depending on requestor. (Leidner Note: For example, historically, NYPD allowed NYC imagery to be made public in 1999 when the pixel resolution provided was increased from 1' to 4'.)

Jack Eichenbaum, GISMO Board: Data accuracy and completeness is extremely important. It is critical that data creators and users get out into the field to verify data first hand. Increasingly funding goes into technology and not into ensuring that data is high quality and well maintained. The data quality of Open Data and its currency is of great concern.

Amy Jeu, Hunter, GISMO Board: In a crisis like 9/11 or Sandy you need reliable data to come together rapidly. Knowledge of data quality should begin in the school system.

Dara Mendeloff, Columbia/CIESIN, GISMO Board: We need a follow-up meeting on data and data work flows. We need standards and we need data quality enforcement. Data creation, quality and stewardship should be taught in school. Agrees that middle graders are highly receptive. Notes that different databases use different names for the same school. Wonders whether MODA feedback loop is getting these kinds of comments. We need more of a dialogue and strategy to improve data.

Amy Jeu: We also need a Chief Geospatial Officer who takes responsibility for data quality procedures across agencies. The City does not currently have a CGIO.

Varun Adibhatla, UrbInt: Has worked with CUSP. Also works with immigrants, helping them get jobs by mastering Open Data and offering access to Open Data as a job skill. Believes he has assisted as many as 50 people get jobs in this way. More citizen engagement about Open Data is important. We need ambassadors to go out into the City. We also need to make sure that data collected in the field is accurate and true. Alan Leidner suggested whimsically that perhaps what is needed is an Adopt-A-Data-Set program where community groups, schools and individuals work to support increasing accuracy and completeness of open data sets.

Marija Drobnjak, CCC: Her child advocacy organization interfaces with City social services agencies. Often negotiates to obtain unpublished data. Interested in a dialogue about making more data public.

Albert Webber, DOITT: There is an open data coordinator in every City agency. Data sets are made public even if they are not perfect with the belief that they should be published quickly and then improved it over time. Users warned about possible data inaccuracies. BetaNYC brought in to teach data literacy. Makes a distinction between data accuracy and data errors. Data accuracy issues occur when data collection methods are less than optimal. Data errors occur when people make mistakes. Enforcement of data quality is left to the individual agencies.

Amen Mashariki: Training on how to use data is important. Government transparency is improved when the public understands their data. Many reasons to teach younger students about data and how agencies do their job. Empowers communities to monitor services and to identify possible inefficiency and even corruption.

Dara Mendeloff: Data providers should be obliged to take responsibility for the quality, completeness and currency of their data. Research has shown that there can be massive data errors that lead to bad public policy and operational errors. There is a need for data workshops to educate the public, organizations and businesses. Believes there should be an advisory committee composed of representatives from each sector, to provide input to MODA, DOITT and DCP on open data governance.

Josh Lieberman, OGC <http://www.opengeospatial.org/> : The Open Geospatial Consortium is the standard setting body for the spatial community. OGC develops tools to foster data interoperability across agencies within a jurisdiction and between jurisdictions and sectors. The Open Data movement is an opportunity to implement standards. Believes jurisdictions should aim for an “Open Data Eco-System” with dynamic links between data sets and organizations. OGC has a document on best practices for spatial data on the web.

<https://www.w3.org/TR/2017/NOTE-sdw-bp-20170928/>

Jurisdictions need a federated model and should be able to share sensitive data at different scales for internal use and for public consumption. Open data should be supplemented by various kinds of tools to make analysis easier and more user friendly. The current NYC Open Data Socrata platform is a good one but could be augmented with additional tools to make access to and use of the data easier.

See Josh’s PPT at this link:

<https://docs.google.com/presentation/d/103VSVYhyva0xu1qdbtLky3280kzyiZi3PqVi0hiMUQ8/edit?usp=sharing>

Albert Webber: Socrata has recently been linked to CartoDB, a platform which will enable easier data visualization without having to have a GIS at the desktop. DOITT and MODA are working towards geocoding those datasets where the City's geosupport application has not yet been used. This will substantially improve location accuracy and spatial context. Pratt Institute is providing Open Data classes for the Brooklyn Community. BetaNYC is also training community residents and students. The Portal website has a gallery featuring various projects enabled by open data.

Alan Leidner: Thanked attendees for a great session. Suggested that an Open Data advisory committee might be helpful to MODA and DOITT.

Appendix: Written Comments for GISMO's Open Data Week Event

Aileen Gemma Smith, CEO Vizalytics Technology Incorporated

Background

Vizalytics was founded in 2012 in the wake of Superstorm Sandy. In my hometown of Staten Island, I saw firsthand the digital divide impacting senior citizens and local businesses. The opportunity to help solve problems existed, through publicly available data, but reasonable means for the folks who needed that data and insight the most did not exist. For a senior citizen, or for a local shopkeeper an open data portal was a cumbersome point of frustration that assumed a level of digital fluency and technical knowledge to get anything meaningful out. Folks were not willing to invest the time or skill building needed to get what may be perceived as limited or unknown utility. We saw this as an opportunity to serve and make open data a critical part of our business.

Example use cases

Vizalytics first mobile app, Mind My Business was released in 2014. The app served thousands of shopkeepers throughout the five boroughs of New York as well as the city of Chicago. Mind My Business was an NYC Big Apps Winner in 2014. Business owners were grateful for all of business specific alerts, suggestions and contextually relevant information that was provided to them, simply by registering for our app with their business name and address. The rest was powered by open data from the city, state and federal level.

Vizalytics partnered with MOTI in 2015 to create neighborhoods.nyc to provide New Yorkers with realtime data at the neighborhood level for every single neighborhood in NYC and in multiple languages. The neighborhoods.nyc project won multiple awards including a Webby Honoree and AWS City on Cloud, Partners in Innovation. Regrettably, due to lack of NYC providing funding to support the project after three years of complete sponsorship from AWS for all of the compute requirements the neighborhoods.nyc site has been sunset. We gratefully note Alby Bocanegra (interim CTO, City of New York) made a concerted and dedicated effort to try to find alternate paths for support for over six months, and we again thank him for his work in trying to secure further funding for the project.

Ongoing work

Vizalytics has continued to expand its proprietary knowledge graph to serve public and private sector clients in transit, tourism and economic development. Our footprint is global, as is that of open data and open data portals with clients hailing from the US, Canada, Europe and Australia.

Due to the business opportunities to work and partner with public sector clients in Australia we have opened a Sydney office where myself and our CTO are currently based. The Vizalytics team acknowledges other city open data portals such as city of Sydney, city of Adelaide, city of Melbourne and more in their ongoing work.

Recommendations

While open data provides much promise, and there are other examples of govtech companies throughout the globe using open data to power their mission critical applications, there is still more work to be done. Data interoperability, better meta data, historical data, and more reliable and realtime updates are just a few of the ongoing areas where additional growth is needed. Collaboration with other cities and municipalities who are further along in their data journey may help to make the business case for city agencies that are reticent to release data, or who find the process of doing so burdensome. Vizalytics has previously provided detailed technical comments at the data set and metadata level to DoITT and would be happy to do so again in the spirit of supporting open data initiatives in NY.

Lastly, I hope you do consider the needs of ALL New Yorkers in ongoing data strategy and initiatives. Not everyone has the time or interest in learning how to use raw data sets, please understand those user's context and do not exclude them from your conversations. Accessibility means different things to different users.

Many users, such as seniors, would like to be included in these outreach initiatives, but are at times **not even aware** that they are happening because the outreach does not reach them in their local neighborhoods (especially more residential ones) or via their local library, senior center, local newspaper and so forth.

Closing comment

The team at Vizalytics is always open to serve and support data initiatives and welcomes ongoing discussion. Thank you for the opportunity to provide these comments, and I hope everyone is having a wonderful open data week.

Respectfully submitted,

Aileen Gemma Smith

Written Comments Received from Emily Goldman, BetaNYC

These were my two main thoughts for the record:

1) There should be more guidance around what qualifies as public versus private datasets, as per the Open Data Law.

Agencies need guidance on this, and should be held accountable to some standard, so that it is not entirely on their discretion.

2) Still, "every day users" find the portal "the environment for a data scientist." This may be fine, but if the desire is to create a tool that every day users can use, the interface needs to be continually adjusted to meet that goal. Perhaps there can be a more instructional version of the portal with short explanations or mini-tutorials.

Sincerely,

Emily Goldman

Comments from Jack Eichenbaum, GISMO Board of Directors and GISMO Founder

1. Along with **accuracy** of data it is important to understand how it is defined. **Good metadata is critical** and must be read. There is also much "insider" info that is known but is not made explicit. A good example is **race data** in the US census.

2. "Bad" maps should be publicly criticized. Perhaps GISMO (and others?) can start the NYMP, the New York Map Police

that comment on *publicly circulated maps* that clearly are based on moot data or are formulated poorly.

Jack

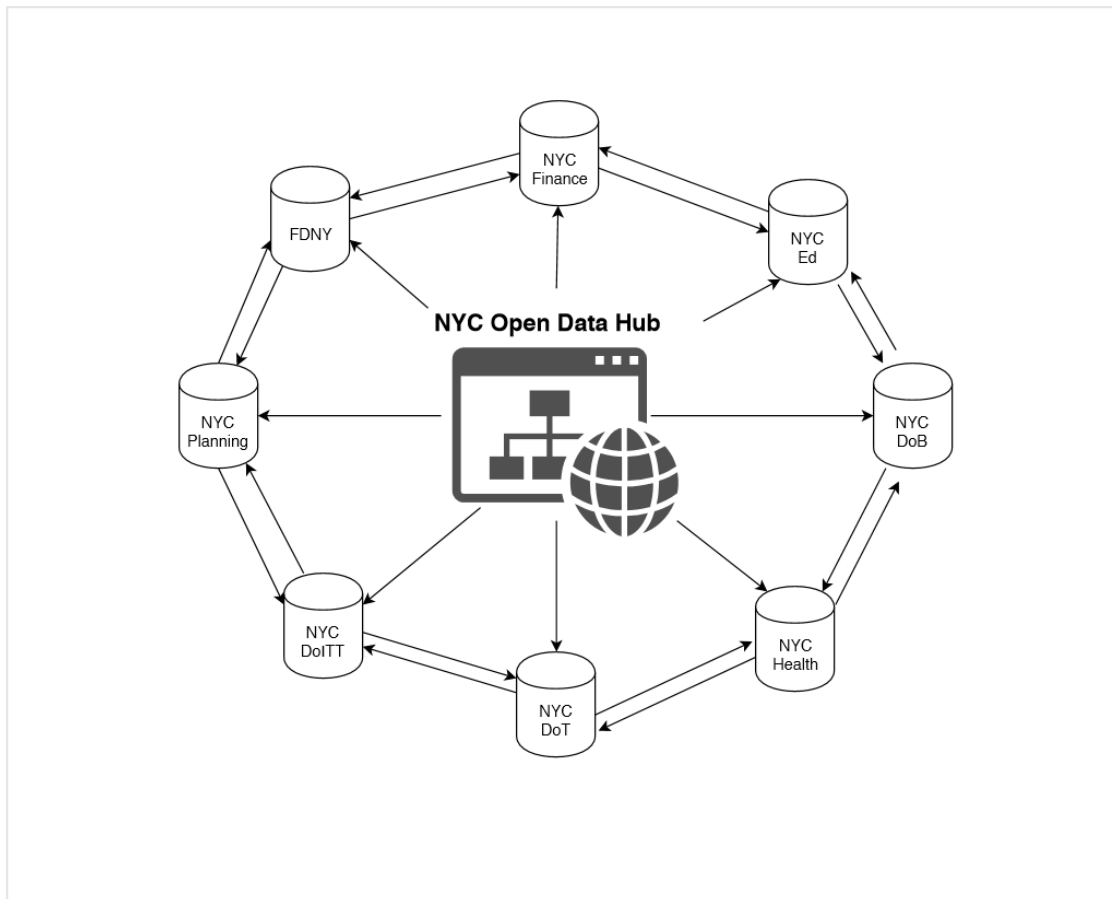
Written Comments Received from Hector Tarrido-Picart, Vivacity

At Vivacity we use the NYC Open Data Portal to help our customers navigate the complicated due diligence process in New York City.

There is no doubt that NYC is ahead of the curve when it comes to data delivery and transparency but we understand that much work still needs to be done.

We believe that the weaknesses we've found in the NYC Open Data Portal can be solved by following the LA Hub model (show diagram). New York should aim to standardize all its data sets (FDNY, NYPD, DOB, Finance, etc) and include a common reference identification number across them. These steps alone would go a long way towards navigating the relationships between data models. Finally, we feel that if all data were tied to a geospatial framework they could be represented on a map and reach far beyond the limits of the traditional "spreadsheet view". Los Angeles has shown that this vision is both achievable and impactful, opening the door for startups like us to stand up better market solutions.

Big problems are solved with strong foundations and there is no mistaking that data is the foundation of our future.



Written Comments from Marija Drobnjak, Citizens Committee for Children (CCC)

It was a real pleasure to participate in the GISMO Open Data event last Wednesday. Thank you again for inviting CCC and I am looking forward to upcoming events. You offered an option to email you key suggestions/concerns as it relates to NYC Open Data and providing constructive feedback to MODA, and here I am including our top three. I haven't included any comments about data quality, currency and accuracy, as many people in the room addressed that issue.h

1. ***If and how NYC Open Data and MODA can be used as an avenue to strengthen and support existing efforts among advocates, nonprofits, service providers in negotiations with city agencies for more public data?***

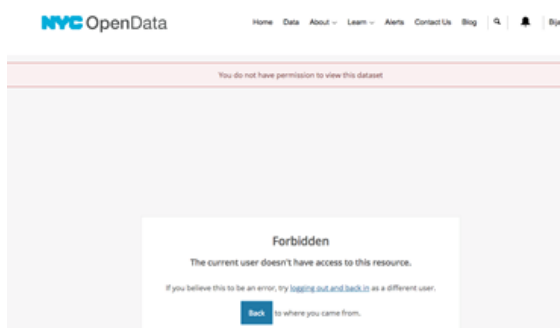
Background: CCC has policy/advocacy relationships with many city agencies and elected officials and leverages these when there is a need to obtain data that is not publicly available. In many instances, we are able to get unpublished data and make it public on our online platform Keeping Track Online (e.g. Administration for Children's Services subsidized child care data; Department of Homelessness Services' data on families entering homeless shelters, etc.), but public conversation about child well-being would benefit from more data that exist, and is not publicly available. Some of the CCC's most recent efforts are related to:

- Early Intervention data
- IEP data
- WIC enrolment data
- Data on screening and utilization of mental health services
- Data related to children and families in homeless shelters

2. ***Provide more clarity regarding the relationship between data available on NYC Open Data and on agencies' websites – which data source is more up to date and current?***

- Example: [Discover DYCD portal](#) vs. multiple datasets on [DYCD Open Data](#) page

3. Provide more information for users when ***publicly available datasets on NYC Open Data, aren't public anymore***. Right now, users get this message:



Thank you again and have a great day!

Marija Drobnjak, M.S. Data Analyst
Citizens' Committee for Children of New York
14 Wall Street, Suite 4E, New York, NY 10005
212.673.1800, ext. 12 | www.cccnewyork.org

EVERY CHILD | HEALTHY | HOUSED | EDUCATED | SAFE