Empowering NYC Communities through the Use of NYC Open Data

Capstone Final Report
May 2016

Client: New York City Mayor’s Office of Data Analytics
Advisor: Patricia Davies
Group: Leticia Correa, Meike Radler, Michael Sedillo, Qiaoyi Chen, Xiao Liu
Acknowledgements

Most importantly, we would like to thank our faculty advisor Patricia Davies for her invaluable advice on client consulting, research design and for her overall guidance throughout this semester-long experience.

We would also like the thank all of the organizations who responded to our email survey, the thought partners with whom we had early discussions, as well as the organizations with which we conducted our in-depth deep-dive activities, including Marco Battistella & Jill Pierce of Community Education Council District 1, Joel Chapman of Students First New York and Neil Reilly of the Citizen’s Housing and Planning Commission as well as those organizations with which we conducted our deep-dive phone calls, including Gregory Bender of the United Neighborhood Houses, Olivia Nunez of the Go Project and Lilah Mejia from GOLES.

Additionally, Jessie Braden and Craig MacDonald of the Pratt Center for Community Development have been instrumental with their interest in, and support of, our work. Will Colegrove also provided important insights into the development of the Open Data Law in NYC and in engaging citizens. Moreover, Tira Schwartz and Pratap Jhadav from Socrata provided us with support and useful information throughout this process.

Finally, we would like to thank our clients, in particular Amen Ra Mashariki and Lindsay Mollineaux from the Mayor’s Office of Data Analytics for getting us involved in such an interesting topic, for their enthusiasm and their valuable feedback throughout our Capstone project.
New York City Mayor’s Office of Data Analytics (MODA) is searching for ways to promote their “Open Data for All” vision by supporting the use of Open Data, particularly through their Open Data Portal (ODP), among community-based organizations (CBOs) and non-technical individuals in NYC as a way to promote a more equal playing field among civic users.

Our goal was to understand the landscape of non civic-technology users in order to advise MODA on ways to promote the NYC ODP as a valuable tool among civilians. Our research methodology followed a three-step process: 1) background research and conversations with important thought partners to understand the open data landscape; 2) a survey of New York-based CBOs; to identify data that is relevant to them and their perceived usage barriers; 3) and in-depth interviews conducted over the phone and in person. Best-case studies on how to leverage and support community use of open data by non-technical users is limited at this point. Thus, our research is relatively groundbreaking in its focus on users outside of the civic hacking sphere.

According to our survey and deep dive interviews, NYC municipal data is in high demand, but it is not entirely clear for potential community based organizations that the ODP is a reliable source of such information. Thus, our recommendations focus on improving aspects of the data available on the portal, the portal interface, outreach to community members and operations of the portal and MODA. Our main findings which include related and detailed recommendations include:

1. General distrust in the accuracy or “cleanliness” of the data on the portal; data that accurately reflects current services in areas where users have domain expertise.
2. The user interface is complicated; tools for finding and analyzing data are not obvious to a novice.
3. Community groups see enormous potential benefits of open data; more outreach should be done in promoting ODP
4. “Lack of appropriate data” was identified as the number one obstacle to using the open data portal.
Part 1: Statement of Need & Background
Part 1.1: Statement of Need

While the availability of Open Data has the potential to be a catalyst for economic and social development while also serving as a tool for transparency and accountability, it is most effective when it is used by all levels of the community. MODA would like a Capstone team to develop an engagement strategy that would provide an understanding of the level of usage currently employed by selected neighborhoods to access relevant open data, determine what roadblocks exist for accessing and effectively using the information (access to technology/training/etc) and then provide recommendations for ways in which MODA can better support the use of open data in communities across New York City.

(from client Terms of Reference)
Open Data Movement / NYC Open Data Law

- Goal is to make some data be freely available to everyone
- Signed into law March 7, 2012
- Available online data of city agencies and departments for the best interest of NYC
- Transparency, effectiveness and accountability of the operation of city government

NYC MODA: Administrator of the Open Data Portal

The Mayor's Office of Data Analytics (MODA) is New York City's civic intelligence center, allowing the City to aggregate and analyze data from across city agencies, to more effectively address crime, public safety, and quality of life issues. The office uses analytics tools to prioritize risk more strategically, deliver services more efficiently, enforce laws more effectively and increase transparency.
Part 1.2: Background

Our problem definition arises from MODA’s worry about the mismatch between its “OD for All” vision and actual usage of the ODP.

“Open Data for All”: MODA’s vision that Open Data should benefit all social members and sectors

Worries: the possible lack of use/information of the ODP by the target population of project

Project goal: Identify pain-points and barriers that non-technical users face in using Open Data Portal and give insight into current levels of ODP engagement/use among CBOs.

Taking the population and the diversity of NYC into consideration, the team decided that it was unrealistic to target individuals one-by-one to demonstrate and promote the ODP. Thus we chose to use CBOs as proxies, hoping that CBOs would help to popularize ODP within communities.
Part 1.2: Background

For background research, we spoke to several thought partners and attended several relevant events in order to become familiar with ODP landscape

Thought Partners
- Phil Ashlock, Chief Data Architect, Dat.Gov
- Emma Truswell, Service Business Manager, Open Data Initiative
- Greg Bloom, Chief Organizing Officer, Open Referral Institute
- Emily Tavoulareas, Founding Member, VA Digital Services
- Jessie Braden, Director of SAVI, Pratt Institute
- Craig MacDonald, Assistant Professor, Pratt Institute School of Information
- Tira Schwartz, UX Researcher, Socrata
- Pratap Jhadav, Data Solutions Architect, Socrata
- Will Colegrove, Senior Technology Advisor, Manhattan Borough President’s Office

Events
- Beta.NYC Open Data Conference
- Pratt Institute SAVI Open Data Workshop
- Socrata Training
- Pratt Institute School of Information UX Class
Part 2: Research Approach and Methodology
Part 2: Research Approach and Methodology

We developed a two-part methodology that will allow us to gather both a high-level overview as well as deeper insights into the pain points for CBO use of the ODP.

<table>
<thead>
<tr>
<th>Survey</th>
<th>Deep-dive Interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Find out the current use level of data usage among CBOs</td>
<td>• Deep-dive into user experience</td>
</tr>
<tr>
<td>• Collect data on the landscape of data usage, knowledge of Open Data</td>
<td>• Identify the obstacles blocking users from fully using ODP</td>
</tr>
<tr>
<td>among different sectors</td>
<td>• Analyse case-to-case user practice</td>
</tr>
</tbody>
</table>

Rationale: Individual users are difficult to target one-by-one. Using CBOs as proxies would to popularize ODP in larger scale.

Rationale: In person interviews can help us to get first-hand users’ opinions and practice.
Part 3: Survey Method, Results and Trends
Part 3.1: Survey Overview

Purpose: To build MODA’s understanding of CBO use of data and the ODP in order to ultimately provide MODA with recommendations rooted in the problems identified.

Community Based Organizations

- To understand to what extent CBOs use “data” and for what purposes
- To understand CBOs’ familiarity with “open data” provided to the public by MODA on the ODP

NYC Mayor’s Office of Data Analytics

To understand how the Mayor’s Office of Data Analytics can better support CBOs’ use of open data in planning and decision making.
Part 3.2: Survey Structure

Part one: General background information on organizations

Get a general understanding about the sectors with which the CBOs are engaged, their relative size and annual budgets

Part two: Understanding CBO's use of data as a tool

To what extent CBOs are aware of ODP, their primary focuses, and the common sources for them to get the data they need

Part three: Follow-up questions

Whether the CBOs are willing to be contacted by our team for deeper dive

Survey questionnaire can be found in Appendix C
Part 3.3: Outreach Approach

- Created a 16-question survey and distributed via email using SurveyMonkey tool
- Began with an initial request to complete survey to approximately 141 organizations:
  - 10 CBOs in 9 sectors each: environment, recreation, business, arts, transportation, education, healthcare, housing, public safety. The 9 sectors are those which are represented on the ODP homepage
  - 36 Community Education Councils (CECs)
- Sent email follow-ups to encourage CBOs to complete the survey
- Emailed additional CBOs in four sectors (business, transportation, environment, and recreation) in which less than 2 responses had been received
Part 3.4: Survey Response Rate

Total complete responses: 48
Total organizations sampled: 141
Response rate: 34%

CECs: 36 (32 districts, 4 citywide councils), Environment: 10, Recreation: 15, Business/Jobs: 8, Arts: 14, Transportation: 9, Education: 11, Health: 10, Housing/Development: 11, Public Safety: 17
Part 3.5: Data Analysis Approach

- Some data was recoded (people who responded “other” were placed into relevant categories)
- Deleted several incomplete responses
- Multiple responses from the same CECs were consolidated into one observation
- Derived summary statistics for all questions
- Did cross-sectional analyses for some questions
- Used Excel graphs and charts to visualize the analyses
- Summarized the take-aways
The majority of CBOs surveyed were small, both in terms of their number of employees and annual budget.
Part 3.6: Survey Results

Of the nine domain areas listed on the ODP homepage, we received multiple responses for each.

The majority of the responses included education, of which Community Education Councils (CECs) were the most commonly represented.

The 3 organizations with an “other” focus, were specified as:

- Multi-issue civil liberties
- Poverty/Advocacy/Social Services
- Economic Development
All but one of the organizations said that they currently use data to make decisions.
Part 3.6: Survey Results

CBOs already frequently use data from government sources, including the NYC municipal government

- However, **only 32%** of respondents were familiar with the ODP (results presented in detail on pg. 25)
- Given the wide understanding of what the word “data” means, our survey thus interrogated the source and type of the data currently used by organizations
Part 3.6: Survey Results

The most popular type of data used by organizations was demographic data, with all other types relatively equal.

Descriptions of data types:
- Financial - related to your industry, funding opportunities, government procurement, etc.
- Demographic - about your constituents, clients or target users
- Landscape - general information about your field, community, clients, services, etc.
- Customer/client preferences - related to client or customer opinions/feedback on your services or services of others.
Part 3.6: Survey Results

The most common source of referral to the ODP for the 9 CBOs that said they had visited the ODP was through a colleague or City Government.

Of the 32% of respondents (15 CBOs) that claimed familiarity with NYC Open Data, 60% (9 CBOs) said they had visited the ODP.

Of these, the following number of organizations said that they or someone on their staff had ever visited the portal:
- Yes: 9 (60%)
- No: 2 (13%)
- Unsure: 4 (27%), likely because they could not say with certainty that nobody in their organizations had ever visited the portal

The most common source of referral to the Open Data portal was through a colleague or City Government.
Part 3.6: Survey Results

The most frequent barriers encountered when using the ODP were related to finding appropriate datasets and ensuring reliable content of datasets.

Several organizations mentioned the difficulties of finding data unless they knew the specific name of the relevant dataset they needed.

Some organizations are also unsure of how the data in the portal could be useful to them.

The main barriers (later called “pain point”) that CBOs encountered when using the ODP were further queried in our deep dives (see pg. 28).
The most common type of data sought from the NYC municipal government is demographic data.

Requirement of financial, demographic, landscape, client preference data are equally important but requirement of demographic data is a notch above others.

**Vertical axis**: categories of data source

**Horizontal axis**: amount count of each data type under different data source

---

**Part 3.6: Survey Results**

Cross-sectional analysis of data sources/data types

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Financial</th>
<th>Demographic</th>
<th>Landscape</th>
<th>Customer/client preferences</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYC municipal</td>
<td>14</td>
<td>25</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>State government</td>
<td>10</td>
<td>20</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Federal government</td>
<td>10</td>
<td>16</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Multilateral organizations</td>
<td>12</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer feedback</td>
<td>15</td>
<td>17</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>Other non-profit organizations</td>
<td>10</td>
<td>15</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Within the organization</td>
<td>17</td>
<td>25</td>
<td>20</td>
<td>19</td>
</tr>
</tbody>
</table>
Summary of key survey results

- CBOs have a high level of awareness and usage of data. Data, especially municipal data, plays critical roles in daily operation and decision making of CBOs of NYC.

- Most popular types of data: financial, demographic, landscape, client preference data are equally important.

- Barriers in using ODP: the most frequently mentioned one was finding appropriate datasets and ensuring reliable content of datasets.

- Referral to the ODP: the most common source was through a colleague or City Government.
There are important implications for the ODP that arise from the responses of the 68% of respondents (32 CBOs) that had not heard of NYC Open Data

**Findings**

Of these, **67% (21 CBOs)** said that **data collected by the NYC Government** would be **helpful** to their organizations if it were made available.

Of the **31 organizations** that use NYC municipal data (pg. 20 on data sources), **less than 50% (15 CBOs)** are **familiar with NYC Open Data**.

Of the **16 CBOs** unfamiliar with NYC Open Data, **88% (14 CBOs)** say that **NYC government data** would be **helpful** to their organization if it were made available.

Additionally, **67% (21 CBOs)** of respondents that are **unfamiliar with NYC Open Data** said that they **currently use NYC Municipal data** to make decisions.

**Implications**

Represents a **possible growth area** for the ODP: people who are **interested in NYC government data** but are **unaware that it is readily available on the ODP**

This indicates that there is a **large degree of demand** for NYC municipal data.

Furthermore, this indicates that **Municipal data is valued and often used**, but that many organizations are **unfamiliar with the ODP as a reliable source** for such information.
Part 4: Deep Dives Method and Analysis
Part 4.1: Deep-dive Overview

The “deep-dives” methodology was designed to more deeply understand the specific issues users face when trying to find relevant data on the ODP.

Goals of the Deep-dives:

● To gain a better understand of the experience of CBOs -- with varying degrees of sophistication in data analysis -- who might use the Open Data Portal (ODP)

● To Identify pain-points and barriers in CBOs using the ODP, and

● To obtain a clearer sense of what interventions MODA could take to support the use of open data among CBOs.
Part 4.1: Deep-dive Overview

The selection of CBOs to participate on the deep dive was based on survey results and our previous relationships with the organizations.

- From the survey results, it is possible to classify respondent CBOs into two different levels of familiarity with the ODP.

<table>
<thead>
<tr>
<th>Level</th>
<th>Low</th>
<th>Medium/High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criteria for classification from survey</td>
<td>Unfamiliar with NYC Open Data Have never visited ODP</td>
<td>Familiar with NYC Open Data Have or have not visited ODP</td>
</tr>
<tr>
<td>Number of respondent CBOs</td>
<td>32</td>
<td>21</td>
</tr>
</tbody>
</table>

- Based on the results from the survey and on our previous relationship with each of the organizations, we selected one CBO in each category of level of familiarity (Low/Medium/High) with the ODP to do an in-person “deep-dive” (see description of methodology next).

- We also selected three other CBOs for which data plays an important role in their work (based on our previous relationship) with which to do the phone call interview.

- Only those organizations who indicated in the survey that they were willing to have a SIPA representative follow-up with them were contacted.
Part 4.1: Deep-dive Overview

The deep-dives were divided in two approaches, each of them conducted with 3 organizations.

Phone interviews (deeper dive into survey results) 

In- Person (Pre-Interview + ODP Exercise)

Identified six CBOs that participated in the original survey
### Part 4.1: Deep-dive Overview

#### Phone interviews (deeper dive into survey results)

<table>
<thead>
<tr>
<th>Organization name</th>
<th>Organization representative, title</th>
<th>Date (2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>United Neighborhood Houses</td>
<td>Gregory Bender, Co-Director, Policy &amp; Advocacy</td>
<td>April 14</td>
</tr>
<tr>
<td>Go Project</td>
<td>Olivia Nunez, Director of community Engagement</td>
<td>April 18</td>
</tr>
<tr>
<td>GOLES</td>
<td>Lilah Mejia, Disaster Preparedness Coordinator</td>
<td>April 21</td>
</tr>
</tbody>
</table>
## Part 4.1: Deep-dive Overview

**In-Person (Pre-Interview + ODP Exercise)**

<table>
<thead>
<tr>
<th>Level of familiarity with ODP</th>
<th>Organization name</th>
<th>Organization representative, title</th>
<th>Date (2016)</th>
<th>SIPA representative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Community Education Council 1</td>
<td>Marco Battistella &amp; Jill Pierce, Member</td>
<td>April 15</td>
<td>Michael Leticia</td>
</tr>
<tr>
<td>Medium</td>
<td>Students First New York</td>
<td>Joel Chapman, Policy Analyst</td>
<td>April 1</td>
<td>Michael Meike</td>
</tr>
<tr>
<td>High</td>
<td>Citizen’s Housing and Planning Commission</td>
<td>Neil Reilly, Senior Policy Analyst</td>
<td>April 8</td>
<td>Michael Leticia, Jessie Braden (Pratt Institute)</td>
</tr>
</tbody>
</table>
Part 4.2: Deep-dive Structure

The in-person interviews followed a three-step process designed to elicit insights into common barriers faced by users with varying levels of ODP experience

1. Pre-interview
   - Inquires CBO use of data
   - Investigates details of data sources and types used
   - Clarifies level of familiarity with ODP
   - Queries previous ODP issues

2. Prompt
   - CBO participant given a prompt relevant to organization's focus
   - Prompt intended to have participants find specific information and do simple manipulation of data in ODP

3. Observation
   - Participant observed until they reach a point at which they are no longer able to continue or have a question (seen as a “pain point”)
   - Socrata or SIPA representative helps participant overcome pain point

Technical set-up and logistics:
   - Participant(s) from a CBO and two SIPA representatives were present in room
   - One SIPA representative took notes, the other helped lead participant(s) through the process
   - Participant used a laptop to complete ODP prompt, laptop screen shared with Socrata UX Research representative, who was also available over conference call speakerphone line
   - Entire sessions were recorded via conference call system for further data analysis

N.B.: see appendices for copy of pre-interview questionnaire
### Part 4.2: Deep-dive Structure

Exercises were designed to observe the users while they were looking for specific information related to their organization on the ODP.

<table>
<thead>
<tr>
<th>Level of familiarity with ODP</th>
<th>Organization name</th>
<th>Prompting questions: “Find…”</th>
</tr>
</thead>
</table>
| Low                          | Community Education Council 1                                   | ● HS attendance rates for the 2014 - 2015 school year for schools in district 17 in Brooklyn"  
                              |                                                                 | ● # of UPK seats for PS 89 Cypress Hills  
                              |                                                                 | ● Math test results from district 17 in year 2008...avg from years 2009 - 2011 |
| Medium                       | Students First New York                                          |                              |
| High                         | Citizen’s Housing and Planning Commission                       | ● Housing Maintenance Code Complaints from Community Board 4 in 2014  
                              |                                                                 | ● How many vacant units in Manhattan?  
                              |                                                                 | ● How many NYCHA residents live in community district 13? |
Part 4.3: Deep-dive Experience - Phone Calls

United Neighborhood Houses (UNH) supports its members through policy development, advocacy and capacity-building activities. Gregory Bender, UNH Co-Director for Policy & Advocacy, primarily uses NYCHA data for his work. He is familiar with the NYCHA website, and he is satisfied with its tools. He is aware of the existence of the ODP, and even though he knows that the data he uses is probably available on the ODP, he does not usually go to the portal.

The Go Project provides academic enrichment and family support for public school students in need. Olivia Nunez, Director of Community Engagement, works with NYC education data and also with data collected by Go. She usually gets data from the Department of Education website, which she does not find to be user friendly. She was not aware of the existence of the ODP and she believes that some kind of support for new users would be helpful.

G.O.L.E.S. is a housing advocacy organization. We spoke with their Disaster Preparedness Coordinator, Lilah Mejia. In her role, Lilah regularly uses data. Even though she uses NYC data, she was not aware of the NYC ODP. She believes that a centralized source of NYC data would be very important to reduce the time that is wasted looking for the data in several municipal departments.
Marco Battistella and Jill Pierce were not familiar with the ODP, and they were not sure that the website would have helpful data. They usually get the data they need from the DOE website and they find it pretty user friendly. Marco and Jill experienced the ODP with us and they were excited by its potential, but had difficulties finding the data sets they were looking for because the description of the file was misleading. Both of them said that they would go back to the portal but while Marco was very happy to be able to access the raw data, Jill got more excited about the state open data portal that has more clear visualization tools.

Survey Snapshot:
- Uses data to make decisions
- Not familiar with NYC Open Data
- Data collected by NYC Gov’t would be helpful to organization

Community Education Council’s (CEC) are community advisory boards charged with “promoting student achievement, advising and commenting on educational policies, and providing input to the Chancellor and the Panel for Educational Policy on matters of concern to the district.” They represented a significant portion of our survey responses, so we wanted to make sure to do a “deep dive” with one. CEC D1 uses data to compare budgets and student enrolment rates among schools. If they could have access, they would be interested in neighborhood demographic data that confirms the segregation of schools. This data would allow them to structure their recruitment efforts in such a way as to bring diversity to their school.
StudentsFirstNY (SFNY) is a non-profit education advocacy organization in NYC. We selected SFNY based on their basic understanding of open data and the open data portal. They know about open data, but hardly ever use the portal (however, they frequently use the State’s data portal) even though much of their policy research focuses on municipal education data. We met with their Policy Analyst, Joel Chapman, who often looks at different data sources to compare and contrast different schools in the city.

Joel has tried using the ODP, but has not had success looking for datasets. He seems to be brand-loyal to the state education data portal, which to him seems way more user-friendly and able to anticipate the data he needs. Data plays a huge role in his day-to-day, but as he is under constant time and resource constraints, he doesn’t have time to gain fluency in using another dataset. He is not sure if the data he needs is in the ODP, and past search failures keep him from investing more time in using the ODP.

Survey Snapshot:
- Familiar with NYC Open Data, user interface and lack of appropriate data sets is a barrier in using the site
- Data collected by NYC Gov’t would be helpful to organization
Citizens Housing Planning Council (CHPC) is a housing research organization in NYC. We selected CHPC based on their strong knowledge/experience using the ODP. We met with their senior policy analyst, Neil Reilly, who said that data plays a very important role in the daily work at CHPC. Neil is interested in demographic data available from the American Community Survey (ACS-CENSUS) as well as housing & development data.

The ODP is the main source of municipal information for Neil. He is very enthusiastic about the concept of open data and believes that it gives an equal and fair access to information for everyone. His main critiques of the portal have to do with the reliability of the information: the data is not clean and there is no data dictionary. He never uses the website to manipulate data or to create charts because he is proficient in STATA and other statistical programs that have better options to data manipulation.
We identified eight categories of pain points in using the ODP

**Pain point categories:**

1. Lack of awareness of ODP and potential contents
2. Lack of a user-friendly interface
3. Inefficient search process
4. Ineffectual portal functionalities
5. Lack of clean data
6. Unclear and unreliable mechanism for questions and feedback
7. Absence of subject specific training
8. Ineffective outreach regarding relevant datasets

**Rationale:**

- Recorded user test
- Selected key quotes represent important pain points
- Identified recurring themes
- Categorized quotes by pain point type
Part 4.4: Deep-dives Pain Points

1- Lack of awareness of ODP and potential contents

“Things might be on the portal but I end up getting it because someone e-mails it to me.” “…” “But I am sure a lot of data sets I am using are actually on the portal” (UNH)

“First barrier, I didn’t know about it before you approached me. I think now, the next step would be to have time to go through it.” (GO)

“I really don’t remember what came from the portal and what came from other places” (UNH)

“I don’t think the data is there, but it could be there” (SFNY)

“Honestly I don’t know” (GO)

“come out to us, meet out to us, … there are so many programs out there but if you don’t go to the people they don’t know.” (GOLES)
Part 4.4: Deep-dives Pain Points
2- Lack of a user-friendly interface

“About is very helpful…I want to know what city agency posted it and when it was posted...metadata means very little.” (CHPC)

“It’s a little intimidating…just try to simplify it, simple is better.” (SFNY)

“I don’t know what that means...there’s no explanation of what that means.” (CHPC)

“What am I looking at first of all…” (SFNY)

On the NYSED website:
“Give me more information...what do these columns mean? I’m coming in blind and I’m trying to figure out what I’m looking at.” (SFNY)

“The data dictionary is really important…I shouldn’t have to open a .PDF in a new tab to get the information.” (CHPC)

“I don’t know what that means...there’s no explanation of what that means.” (CHPC)

“Our is very helpful…I want to know what city agency posted it and when it was posted...metadata means very little.” (CHPC)
Part 4.4: Deep-dives Pain Points

3- Inefficient search process

“The times that I go on the ODP, it’s never the case where I search for a search term and the first one I click is the one I want…what happens most often is that the one I want is the 7th or 8th one down…never like, ‘OK, found it let’s move on.’” (CHPC)

“What is this? What is PMMR and why is it the first result?” (CHPC)

“The description of the file is misleading” (CECD1)

“To me, it was very disappointing that after the search, the first title was the right one and we skipped it because the description in the information that we had right there seems to be not including that piece of information that we actually wanted.” (CECD1)

“DOB results...that seems relevant and that’s the 15th thing down...” (CHPC)

“I would definitely not use the dataset for this...because I don’t think it’s here” (it was there) (SFNY)

How to make the search better: “Whenever you have the list view, you have the quick way to see this are the columns present in this data set, of course you also might need the data dictionary” (CECD1)
Part 4.4: Deep-dives Pain Points

4- Ineffectual portal functionalities

“Data banner” refers to the ad-like dataset suggestions that take up nearly half the page when users search for data sets.

When asked about the data banners: “it’s kind of an annoyance because it didn’t allow you to see much of the data” (CECD1)

“The fact that we have these pictures -- these icons -- taking up half of the screen, I don’t appreciate that” (SFNY)

“I didn’t look at them at all...are those data sets? I guess they are...none of those are what I’m looking for...they don’t look like something I would take seriously... they look like ads.” (CHPC)

“I haven’t tried making a map in here before and I probably won’t again.” (CHPC)

When searching through data sets: “Popularity is not relevant to me.” (CHPC)
Part 4.4: Deep-dives Pain Points

5- Lack of clean data

“I don’t trust the cleanliness of the data…” (CHPC)

“I’d rather have data that is six months old but answers all the questions I have than data that goes up quickly and is incomplete... that might mean meeting with users of the ODP to talk about what they’re looking for. I imagine Socrata knows what I just searched for... agencies should know that, too.” (CHPC)

“I would not trust data on the ODP with respect to NYCHA.” (CHPC)

“There should be no errors -- it should be complete, it should be accurate…” (CHPC)
Part 4.4: Deep dives Pain Points

6- Unclear and unreliable mechanism for questions and feedback

On giving feedback: “It wasn’t clear for me where I would go if I couldn’t find something” (CECD1)

On giving feedback: “I try to find someone at the relevant agency...I didn’t know you could [pose a question on the site]...I don’t know who would read it, so I would rather send an e-mail to HPD because who reads it?” (CHPC)

Feedback by Email: “Either my question could be misunderstood or it would be a lot of back and forward” (GO)

“On giving feedback: “It would be nice to be able to ask question, and get a little clarity about how they collect the data, if there is any filter and how they are coding it.” (GO)

“On giving feedback: “Being able to call someone would be the most effective or even scheduling it” (GO)
Part 4.4: Deep-dives Pain Points

7- Absence of subject specific training

Interest on training: “If I felt that I would go for different information than I was going to the DOE website for” (CECD1)

Interest on training: “It is definitively something that I think it would be interesting, probably the whole police team would want to go” … “But not as something that is part of the work” (UNH)

“If there were groups, or people that I could connect about how they use data or some type of ongoing training or whatever it could look like, that could also be helpful” (GO)

Interest on training: “Yes, as long as it is not all day” “When you give us a space where we can get data, I think everyone will want to be involved” (GOLES)
Part 4.4: Deep-dives Pain Points

8- Ineffective outreach regarding relevant datasets

“When asked about being notified about relevant data: “It would be very helpful for me” (CECD1)

“If you got one email a day, telling you some new data set, maybe not understanding what it is about,... maybe it will go direct to the junk. But if it comes once a month… there is actually a new data set…” (CECD1)

“It doesn’t have to be a monthly task force, but getting survey information from users to understand what they’re looking for…” (CHPC)

“If MODA suggested datasets: “Yeah - that’d be interesting…” (CHPC)

“Datasets that I didn’t know existed...if it’s relevant to what I’m looking for...I would take a look” (CHPC)

“I found very interesting what you were suggesting earlier about learning about data set that I haven’t thought about.” (CECD1)
Part 4.5: Deep-dives experiences: User Feedback

Key Takeaways

● Many users do not know a) what open data is, b) where/how to access the NYC open data portal or even why they would try and c) that much of the data they need is already on the ODP

● Users do not like the ODP interface because it presents excessive tools and overwhelming information

● Searching for appropriate datasets frustrates users
  ○ Icons, views, popularity, etc. are not helpful signifiers when choosing the appropriate datasets
  ○ The website does not prioritize well the search results

● It helps when users know that the data they are looking for exists on the portal
  ○ User experience difficulty understanding the data that they are looking at: data dictionaries and clear columns labels would be helpful
  ○ Unclean and outdated data is very frustrating and makes it less likely that the user will return

● Users find interesting the idea of having their needs anticipated
  ○ They would like weekly emails regarding relevant datasets being released
  ○ New York State Education Department cited as an example of a website that does a good job of predicting what information users want and informing them about new and relevant datasets
Part 5: Recommendations
Part 5: Data Recommendations

Brief summary of the problem
While our survey results indicate that CBOs are hungry for NYC government data, many of them do not trust the accuracy or “cleanliness” of the data on the ODP. Indeed, survey respondents listed “lack of appropriate data sets” as the number one barrier to using the ODP. CBOs looking for sector-specific data know their fields well, and they can tell when data is funky, which risks MODA losing credibility. Basically: unclean data is frustrating, makes it much less likely that a user will return, and undermines the portal’s purpose and credibility. Further, new users need to quickly know what they are looking at via a data dictionary as they are often confused by various acronyms used within datasets.

Recommendations
- Work with other government departments to ensure clean, accurate and more reliable datasets posted in ODP
- Provide user-friendly and easily-accessible data dictionaries
- Make data source, timeframe and contents more apparent for users
- Integrate data dictionary within dataset with, for example, a “hover over” function
Part 5: Website Recommendations

Brief summary of the problem:
The user interface and the lack of appropriate dataset (or inability to finding it) were mentioned in the survey as barriers in the use of the ODP. These survey finds were corroborated with the deep-dive analysis. During the exercise, users complained that the website presents excessive tools and overwhelming information. They also mentioned that the search process is frustrating. Beside presenting irrelevant information (such as icons, views, popularity), the search process does not effectively prioritize results. Users thought that the ODP should be able to learn more about their preferences and predict/suggest datasets that could be interesting for them.

Recommendations:
- A cleaner interface
  - Collaborate with Socrata to provide a more user friendly interface
- Make a more efficient search process
  - Improve the datasets description and the search prioritization tool
- “Netflixification” of ODP - knows what users like
  - Evaluate the possibility of implementing a tool that permits the website to collect information about the users and predict/suggest relevant datasets based on their search history
- Improve data manipulation functions
  - Create tutorials and improve the data manipulation usability
- Provide video or other tutorials (i.e. Austin’s How-to wiki\(^1\)) to help new users

\(^1\)http://atxdataportal.wikispaces.com/.
Part 5: Outreach Recommendations

Brief summary of the problem:
With the main source of referral to the ODP coming from colleagues or City Government, the ODP and NYC government data have an enormously potential growth of awareness which was indicated by the survey. Deep dive users also felt it necessary and useful for themselves and other conceivable users to be more informed of existence of the website, relevant datasets and usage of data.

Recommendations:

● Engage more people and CBOs of using ODP
  ○ Cooperate with other government agencies to publicize ODP

● Inform relevant CBOs of relevant dataset releases
  ○ Develop web feed and send out email blasts to inform updates of ODP or datasets to CBOs which sign up to the sectors that they are interested in

● Provide specific training workshops in each sector
  ○ Collaborate with Socrata, city governments, academic and research institutes, NGOs, civic hackers to develop and furnish training workshops

● Build up and promote active online community of ODP
  ○ With online community which is affiliated with ODP, people could communicate and solve problems of data usage with help from fellows
Part 5: Operations Recommendations

**Brief summary of the problem:**
Survey results indicated that a lack of appropriate datasets is the most common barrier for CBO use of the ODP. Additionally, our deep dives led us to believe that CBO users do not understand how they can provide feedback on outdated or incorrect datasets, how they can have their questions about specific datasets answered as well as how to request relevant data in an efficient and transparent manner.

**Recommendations:**
- **Feedback process:**
  - Work with departments to provide a channel through which users on ODP can get questions answered about specific datasets
  - Develop an efficient reporting process so that users can report issues with outdated or incorrect data and have their issues dealt with quickly
- **Dataset requests:**
  - Provide more information about the ways in which dataset requests are handled and a specific timeline for such requests
Appendix A

Important Abbreviations

● Mayor’s Office of Data Analytics (MODA)
● Open Data Portal (ODP)
● Community Based Organizations (CBOs)
● Non-governmental Organizations (NGOs)
● Community Education Councils (CECs)
Appendix B

Terms of Reference
Project Title: Value Mapping NYC OpenData for MODA
Client Organization: New York City Mayor’s Office of Data Analytics
Client Website: http://www.nyc.gov/html/analytics/html/home/home.shtml
Client Project Director and Contact Information:
Dr. Amen Ra Mashariki
Chief Analytics Officer
AMashariki@cityhall.nyc.gov

Background:
The Mayor’s Office of Data Analytics (MODA) is New York City’s civic intelligence center, allowing the City to aggregate and analyze data from across City agencies, to more effectively address crime, public safety, and quality of life issues. The office uses analytics tools to prioritize risk more strategically, deliver services more efficiently, enforce laws more effectively and increase transparency.

Objective:
While the availability of Open Data has the potential to be a catalyst for economic and social development while also serving as a tool for transparency and accountability, it is most effective when it is used by all levels of the community—MODA would like a Capstone team to develop an engagement strategy that would provide an understanding of the level of usage currently employed by selected neighborhoods to access relevant open data, determine what roadblocks exist for accessing and effectively using the information (access to technology/training/etc) and then provide recommendations for ways in which MODA can better support the use of open data in communities across New York City.

What is the experience of a community that uses NYC Open Data and what is the value of NYC Open Data to a group of citizens in finding local solutions or answers to issues within its community?

Tasks:

First, students will explore the NYC Open Data Portal – what’s there, how does it work?

Students will select a community to work with as a case study for examining the ways in which open data could be used to solve problems within the community. MODA will assist by suggesting several neighborhoods that would make good candidates for this study and allow students to choose ONE neighborhood for study.

Define the neighborhood: demographics, opportunities, challenges, realities.
May begin with Pratt Data Portal   http://prattcenter.net/neighborhood-data-portal

Why this neighborhood? Identify key stakeholders. Identify key partners – those who can facilitate access to community groups for interviews and focus groups.

Convene community stakeholders and work with them to identify key issue or questions they would like to address.
• First need to understand how can the Open Data Portal be used. What is its potential? Limits? How does it work?
• Who is currently using it and how?
• Second, what data is available and relevant to community issue?
• Provide examples of innovative technologies that are using urban data to address quality of life issues (NYC e.g., BigApp contenders).
• Set up a pilot project: define community, community defines an issue to explore and parameters; students navigate NYC Open Data with community and document the process of using it as a tool.
• Conclusion will include findings and recommendations for improving experience and outcomes.

**Deliverables and timetable:**
The Capstone team will provide analysis of the ways in which Open Data could be used in a community to solve public policy issues, as well as recommendations for—what works well, what is needed to ensure user success and positive experience, challenges and assessment of community accessibility in terms of technology, physical access; also accessibility in terms of demonstrated comfort with data and massive amounts of raw data.

General Timeline:
• Detailed work plan: first week of February, 2016
• Draft report outline: First week of March, 2016
• Draft report: Mid-April, 2016
• Presentation of key findings: End of April, 2016
• Final report: First week of May, 2016

**Preferred Qualifications:**
Data analysis – comfortable using large data sets. Basic Excel and statistical skills, possibly skills teaching or instructing others.
Comfortable with written and verbal communication; interviewing and focus group management.
Data visualization – graphics and design a plus!

**Logistics:**
Regular communication with NYC MODA (suggested a standing check-in call scheduled every 2 weeks during the spring semester). Additionally, students will meet with the MODA team at the outset of the project for a ‘Kick-Off’ meeting, a mid-term presentation and a final presentation. This meeting can take place at NYC MODA’s Office.
Appendix C

Survey
Community Based Organization Survey of NYC Open Data

**Introduction:**

We are a group of graduate student consultants from School of International and Public Affairs, Columbia University working with the New York City Mayor’s Office of Data Analytics (MODA) and we need your help!

We are targeting non-profit and community-based organizations that advocate for issues on behalf of New Yorkers.

Our goal with this survey is to understand to what extent your organization uses “data” (however you define this) and for what purposes. Additionally, we would like to know if you are familiar with “open data” and more specifically, the data that is made free and available to the public by the NYC Mayor’s Office.

The survey should take less than 5 minutes to complete and will be enormously helpful to us in providing recommendations to the Mayor’s Office in improving awareness of the value of NYC Open Data to different organizations and communities to improve inequities in data usage.

Even if you do not think that any of this is relevant to your organization, we would love to hear from you about the reasons why!

Thank you for participating in our survey. Your feedback is important.

**Section 1 Introduction**

Purpose: General background information on organization

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer choices</th>
<th>Answer type</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How many employees does your organization have?</td>
<td>a. &lt; 5</td>
<td>categorical - select one - required</td>
<td>Move to question 2</td>
</tr>
<tr>
<td></td>
<td>b. 5 - 15</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. 15 - 30</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. 30 - 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>e. 100+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. What is your average annual budget?</td>
<td>a. $50,000 or less</td>
<td>categorical - select one - required</td>
<td>Move to question 3</td>
</tr>
<tr>
<td></td>
<td>b. $25,000 - $200,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c. $200,000 - 500,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>d. $500,000 - $2million</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>e. $2million +</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Which area is your organization’s primary focus?</td>
<td>a. Arts</td>
<td>categorical - select all relevant - required</td>
<td>Move to question 4</td>
</tr>
<tr>
<td></td>
<td>b. Business/Jobs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Section 2: Questions related to Data Usage

Purpose: Understanding their use of data as a tool

Introduction: More than ever before, information is now widely available, free and accessible through all kinds of internet-based websites and applications. More and more businesses and government agencies use data as a tool to improve services and better understand their customers.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer choices</th>
<th>Answer type</th>
<th>Action</th>
</tr>
</thead>
</table>
| 4. Does your organization currently use data in order to make decisions? | a. Yes  
b. No                                                          | Required          | If YES → Move to question 5  
If NO → Skip to question 7   |
| 5. What is the most common source of the data your organization uses?    | a. Data collected from within the organization  
b. Data collected from other non-profit organizations  
c. Customer feedback  
d. Multilateral organizations (i.e. UN, World Bank, IMF, etc.)  
e. Federal government  
f. State government  
g. NYC municipal  
h. other: ___ | categorical - select one - required | Move to question 6 |
| 6. What kind of data is most useful to your organization in planning or decision making? | a. Financial - related to your industry, funding opportunities, government procurement, etc.  
b. Demographic - about your constituents, clients or target users?  
c. Landscape - general information about your field, community, clients, services, etc.  
d. Customer/client preferences - related to client or customer | categorical - select all relevant - required | Move to question 8 |
<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
<th>Type</th>
<th>Next Question</th>
</tr>
</thead>
</table>
| 7. Why does your organization not use data as a regular planning or decision making tool? | a. Do not know where to start  
b. Lack of organizational capacity/understanding  
c. Data is unavailable  
d. Other: _______________ | categorical - select all relevant - required | Move to question 8                        |
| 8. Are you familiar with NYC Open Data?                                  | a. Yes  
b. No  
c. Other: _______________ | categorical - select one - required | Move to question 9  
If NO → Skip to question 12 |
b. Colleague  
c. Media  
d. City Government  
e. Other: _______________ | categorical - select all relevant - required | Move to question 10  
Move to question 11 |
| 10. Have you or anyone on your staff ever visited the NYC Open Data Portal for the purposes of finding data for your organization? | a. Yes  
b. No  
c. Unsure  
ed. Other: _______________ | categorical - select one - required | Move to question 12  
Move to question 13  
If NO → Skip to question 14 |
| 11. What is the main barrier you encountered in using the open data portal? | a. User interface  
b. Lack of appropriate data sets  
c. Lack of “clean” data  
d. Other: _______________ | categorical - select all relevant - required | Move to question 12  
Move to question 13  
If NO → Skip to question 14 |
| 12. Would data collected by the NYC government be helpful to your organization if it were made available? | a. Yes  
b. No  
c. Unsure  
ed. Other: _______________ | categorical - select one - required | Move to question 14  
Move to question 14 |
| 13. In general, what kind of data would be helpful to your organization? | a. Arts  
b. Business/Jobs  
c. Education  
d. Environment  
e. Health  
f. Housing/Development  
g. Public Safety  
h. Recreation  
i. Transportation | categorical - select all relevant - required | Move to question 14 |
|                                                                         |                                                                        |                                     | Move to question 14                       |
### Section 3: Follow-up

**Purpose:** Follow-up question

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
<th>Type</th>
<th>Next Step</th>
</tr>
</thead>
</table>
| 14. May we follow up with you for some additional questions to better understand how to improve the Open Data Portal for users (or potential users) like you? | a. Yes  
b. No | categorical - select one - required | If YES → Move to Question 16.  
If NO → Move to Question 15. |
| 15. What is the name of your organization?                             | a. Prefer not to say  
b. Name of organization | required | Finish survey when completed |
| 16. Please provide name and contact information                         | Name:  
Organization:  
Email:  
Phone: | | Finish survey (make sure to say thank you) |
Appendix D

Deep dives pre-questionnaire
Record Interview:

Let’s keep goal of Capstone in mind
- Help MODA ID barriers;
- Support MODA’s “Data For All” vision

Pre-Exercise: Questions should focus on:
- Can you give us an example of how you use data in your org? We are looking for examples that can inspire other organizations. This can be helpful in the engagement strategy.
- Have you ever experienced a situation in which you needed a data set but you couldn’t find the data? We need understand if the data was not available or if they couldn’t find it.
- Are the data you need/use available at ODP?
  - If yes: Do you have any complaints about the way the data is presented?
  - If no: We need to know if this is a data set that should be at ODP. If it is not the case, there is nothing we can do.
- Amen’s mall metaphor
  - Do you use other websites to get access to municipal data?
  - Why do you seek other venues for finding data? They mention in the survey problem with the interface and lack of appropriate data sets. Are these the only reasons?
  - Where are you going to find other data sets?
- What has been your experience using other data sets? Not only the ones with municipal data.
  - How does it compare to ODP? We are looking for examples of better interface, data cleaning, support,...
- How could the data sets you’re currently using be improved? Both ODP and other sources. We are looking for what would be the perfect experience for the user.

Observing them using NY Open Data Portal:

Mid-Exercise: Question should focus on:
- How did it feel to get started searching for the data? Did it feel similar to other data sets you’ve explored? How easy is to navigate in the web site? Is it intuitive?
- Talk about the ODP’s usability and ease with which you could find your data set? We are looking for barriers: for example if it was difficult to find the correct data set among all data sets shows up.

Post-Exercise: Questions should focus on:
- Evaluate the experience:
  - Is the website user-friendly? Why or why not?
  - Was the data set good? Do you think you can use and interpret the data easily? We are trying to access the problem with poor data, with missing information or with a lot of codes that can not be interpreted.
What are your *Recommendations* to make the experience better?

Would you come back to the ODP for more data sets in the future?
  - Why or why not?

If you could ask MODA/Socrata some questions or give them some feedback what would you say?

Other comments
### List of client meetings

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Location</th>
<th>In attendance</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 28th, 6-7pm</td>
<td>SIPA</td>
<td>Amen, Tricia, Leticia, Meike, Michael, Qiaoyi, Xiao</td>
</tr>
<tr>
<td>February 18th, 3-4pm</td>
<td>Conference call</td>
<td>Amen, Lindsay, Leticia, Meike, Michael, Qiaoyi, Xiao</td>
</tr>
<tr>
<td>March 10th, 3:30-4:30pm</td>
<td>MODA Offices, 253 Broadway</td>
<td>Amen, Leticia, Patricia, Meike, Michael, Xiao</td>
</tr>
<tr>
<td>April 7th, 3-4pm</td>
<td>MODA Offices, 100 Gold Street</td>
<td>Amen, Lindsay, Patricia, Leticia, Meike, Michael, Qiaoyi, Xiao</td>
</tr>
<tr>
<td>April 21st, 3-4pm</td>
<td>MODA Offices, 100 Gold Street</td>
<td>Lindsay, Patricia, Leticia, Meike, Michael, Qiaoyi, Xiao</td>
</tr>
<tr>
<td>May 12th, 3-4pm (Final Briefing)</td>
<td>MODA Offices, 100 Gold Street</td>
<td>Amen, Lindsay, Patricia, Leticia, Meike, Michael, Qiaoyi, Xiao, MODA team</td>
</tr>
</tbody>
</table>