

## District of Columbia Interagency Data Team

**June 27, 2018**

2:00 PM – 3:30 PM

Office of Chief Technology Officer

200 i Street SE

Washington DC, 20003

Conference Rooms 1001A & 1001B

### Agenda

- ***Welcome, News & Updates***  
Michael Bentivegna, Data Visualization and Analysis PM, Office of the Chief Technology Officer
- ***Levels 1-4 Data Sets, Developing a Data Sharing Governance Structure***  
Guy Hammond, Performance Analyst, Office of the City Administrator
- ***191 Level 0 Data Sets, Timeline and Data Submission Process***  
Mario Field, Data Curation PM, Office of the Chief Technology Officer
- ***Newly Published and Updated Datasets***  
Mario Field, Data Curation PM, Office of the Chief Technology Officer
- ***Master Address Repository 2.0***  
Julie Kanzler, Data APIs & Systems PM, Office of the Chief Technology Officer

## Meeting Notes

Best attempt to capture notable comments and questions from attendees (paraphrased).

### Presenter: Michael Bentivenga (OCTO) – Welcome and News

- Tableau Server Updates
  - Update to 2018.1 had significant problems but did not affect production
  - OCTO has recently resolved all issues
  - Change Advisory Board (CAB) tickets are submitted and once approved we will be updating to 2018.1 on July 5<sup>th</sup>
- Tableau Public (for DC) presentation and process is postponed but users should come to us if you are ready to create a new ETL process for regular updates
  - *Question:* what is Tableau Public?
    - Tableau Public (for DC) is a separate infrastructure that we've deployed for embedding into agency websites. It allows agency to share visualizations publicly. We can push data out but will have no backward connections to data sources. Big difference, Tableau has a free environment but it is not ETL based and they use their branding. Ours will have ETL and no Tableau branding.
  - *Question:* have you made progress on decision of what other BI tools DC Gov will use?
    - We decided to use Tableau and Microstrategy as our centralized tools. These are what we have resources for: training, server, tech support. We will not be supporting Power BI or Qlik.
    - We will put on the OCTO website how we've deployed Tableau and Microstrategy.
  - *Question:* what agencies are using Microstrategy?
    - Many agencies have access to view dashboards created by OCTO. For example, the CIO dashboard. Agencies with their own projects and dashboards are,
      - Department of Employment Services
      - Office of Attorney General
      - Department of Human Resources
      - Department of Behavioral Health
      - Department of Motor Vehicles
      - Department of Human Services (DCAS)
      - Department of General Services
- Microstrategy Updates
  - Upgraded to 10.11 and now finalized all of testing and rolled out to production
  - If you'd like to test connectivity we do have a place for you to do so
- ArcGIS Online
  - Mapping tools in Tableau and Microstrategy are good, but fairly basic. It just can't compete with a company like Esri that focuses on geospatial apps like ArcGIS Online
  - However you can embed ArcGIS Online maps into your BI apps
  - Nice way to visualize with analysis and/or create story mapping

- Find out more by visiting <https://dcgis.in.dc.gov> to learn more. Note that this is an internal DC website
- Tableau Training
  - July trainings are full which includes the first intermediate class
  - Newly scheduled trainings are,
    - Introduction classes September 19 to 20 and October 17 to 18
    - Intermediate classes August 28 to 29 and September 12 to 13
    - Sign up at <https://dcgis.in.dc.gov> and click [Tableau Training](#). Note that this is an internal DC website

## Presenters: Guy Hammond (OCA) – Levels 1-4 Data Sets, Developing Data Governance Structure

- First, mention that this project has not been rolled out yet. But it's also not new because you've been doing it. What *is* new is OCA's involvement coming up
- With the completion of the EDI we know what we have and it's impressive
- Now the challenge is getting hands on the data - shared.
- The default between agencies tends to be "no sharing." If one director can't get it from another. So what do we do?
- One approach... We use our personal relationships to get an extract of some sort and do the best we can.
- Another approach... the Lab@DC (within OCA) has created a new data sharing agreement workflow and documentation. It is essentially an MOU with a narrow focus and specific sharing purposes – only for the Lab@DC and can list names.
- Another approach... FOIA. Yes, some agencies have used the FOIA process within DC Gov
- What I would like to describe is an idea for a longer-term project
  - This would be only be for datasets in the EDI classified Levels 1 through 4
  - Our first step is to create a high level working group with volunteers from within this Interagency Data Team
  - We *do not* want this to be a policy exercise. OCA wants it to be actionable
  - We are now asking agencies to come to the working group with data willing to share – work through the process
  - Bring data that you see value and think other agencies would value and in return you get data shared with you.
- OCA can bring a little weight and authority to this project
- *Comment:* DDOT has participated in a similar project on a national level. National Committee of Sciences help put this project together on a national scale.
- *Comment:* we need to work on procedures for aggregation of data with privacy concerns. Consider when is it sufficient aggregation. Also create standard domains so that we can reconcile. Consider procedures for obscuring.

- *Question:* we don't like to share because we don't know where the data is going or when will it be destroyed. We already have the MOU process that works for us. What is the incentive for participating in this upcoming project?
  - What is your process right now? The incentive is creating a uniform process rather than current informal between agencies.
- *Comment:* as part of the building of this we have to go to agencies and ask agencies. At what point do you feel comfortable in sharing your data. The answer shouldn't be never.
- *Comment:* executive partnership is very important. Even the Mayor needs to have this a priority. Also non-mayoral agencies need to participate.
- *Comment:* example OCTO is currently trying to get census data from MPD. It's been tough but to help we are now showing examples of how we've safeguarded other sensitive datasets.
- *Comment:* there needs to be an education campaign on how/what for understanding
- *Comment:* DDOT is now finally getting data from DPW fleet but it took a long time. It could have taken less with more information on how data sharing could best work with regard to security.
- *Comment:* this new working group needs to review all antiquated laws out there that people may not be aware of. These sometimes surface without any knowledge; what are the laws out there.

#### Presenter: Barney Krucoff – New Mayor's Executive Order

- New order is out 2018-50
- Due date for future EDI publications is now March 10th
- March 10, 2019 will be based on datasets from December 2018
- We got a lot of feedback from Agency Data Officers (ADO) about the first Enterprise Data Inventory (EDI) process
- We will be building a new simplified application for EDI separate from IT Assessment tool
- Soon OCTO Data liaisons will be working with you and ADO to publish/update EDI

#### Presenter: Mario Field – Level 0 Data Sets, Data Submission Process

- We reviewed all 191 open data sets not in <http://opendata.dc.gov>. However some may not fit as a defined *enterprise* data set and/or is a duplicate
- We identified 140 as appropriate Level 0 data sets to work with
- DMV and DPR highest with 13
- Everyone should expect your OCTO Data liaison to reach out to you soon for publication
- One of the things we want to do is make the OCTO data publication process transparent.
- After comparing other jurisdiction, we found that New York State had the most comprehensive submission process. We are going to be modeling ours after this.
- Our pilot, DCPS lottery data sets
  - First, the agency should initiation by submitting a OCTO Helps ticket via remedy. Found at <https://start.dc.gov> or <https://octo.onelogin.com>. A DC.gov email account is required

- Second, a consultation for building a comprehensive requirements document for the dataset. This will include the expectation of the data and an assessment of security/sensitivity. This document will actually be a part of the data set.
- Third, the curation process OCTO performs (black box). Includes steps like checking against current domains. Defining how the data will be loaded based on source. Example, running datasets through the MAR and appending MARID fields (and others). Identifying column name limitations and creating plain language aliases. Metadata will come from EDI too.
- Publication, based on classification level
- *Question:* is this a one-way process? Are you just going to clean it up or will you provide us a process for cleanup?
  - It's always two-way. Hoping to alleviate clean up by having the requirements set prior to curation
- *Question:* who owns the metadata when it's done?
  - The agency will be the contact and known as the originator. OCTO would be the steward. If you are the originator, you should consider and think about how you want the contact information to appear to the public.
- *Question:* how would you know if the data is accurate? The agency is providing but OCTO will not know if it is correct from the source.
  - There should be an authoritative sign-off process from the agency
- *Question:* you are doing a test, dry run with the DCPS lottery data set. So then will you present the progress and trial/error by testing from DCPS?
  - Yes we will present at the next IDT
- *Question:* can you publish a schedule as to when OCTO Data team will be reaching out to agency to publish?
  - Our timeframe is based on the Mayor's order of getting this out there by March 10, 2019

### Presenters: Mario Field – Newly Published Data

- The following are new or updated on <http://opendata.dc.gov>
  - ATMs
  - Adopt-A-Block
  - Bank
  - Bridge
  - CBEs including SBE designation
  - Check Cashing
  - Clean Teams Area
  - Crash
  - Credit Union
  - DC property
  - Independent School
  - Moving Violation

- Notary Public
  - Parking Violation
  - Polling Place
  - PASS Contract Data
- Now integrated with Sales Force to get Certified Business Enterprises (CBE). These are registered DC businesses. DC law states that any DC Gov purchase under 250k must be with a CBE. Contracts over 250k should have SBE participation.
- We have p-card and purchase orders already... we are now working on PASS contract data
- *Question:* what is a adopt-a-block.
  - Data set from Mayor's Office of Clean City
- Next data publication during maintenance window are,
  - Circulator Route
  - DVD Product Archive – (open data) – these are old DVDs of DCGIS database we used to distribute publicly
  - Fire Alarm Area
  - Main Street
  - Planimetric – Archives (map service, open data, and database) – this building footprints, street centerlines, road polygons, sidewalks...
  - School Lottery Results

#### Presenter: Julie Kanzler – Master Address Repository 2.0

- A presentation recapping the MAR and myths about it
- Myth 1 – it is only about address
  - It's not, it's about LOCATION
  - It's also about the context of where the address is located (point-in-polygon) – Ward, ANC, Police District, Voting Precinct...
  - MAR 2.0 will have better integration with property data. You can ask location of an address and will also include more such as snow zone, watershed, enterprise zone... You can find any area boundary that you may find relevant to your address.
- Myth 2 – it is strictly a GIS tool
  - It's not. It can be integrated into lots of applications for address validation
  - Even the Geocoder is not just GIS. It's a search engine, taking imperfect data and standardizing it
  - MAR 2.0 will have a backend that is 100% a search engine. It will be elastic search - fast
- Myth 3 – It only contains active locations
  - It's doesn't. There is a whole life cycle to addresses
  - MAR 2.0 will have historic addresses at the request of several power users
- MAR is super smart and it's getting even smarter
- The geocoder will be re-written to use MAR 2.0
- We will also remove those Microsoft dependencies
- *Question:* will the new API allow specifying a date range?

- *Question:* what is timeline?
  - We will have testing capability by next GISSC on July 18

Meeting adjourned

Next meeting tentatively scheduled for September 2019