

## District of Columbia Interagency Data Team

**May 17, 2017**

2:00 PM – 4:00 PM

Office of Chief Technology Officer  
200 I Street SE, Washington DC, 20003  
Conference Room 1001A & 1001B

### Meeting Notes

### Agenda

- ***Welcome, News & Updates***  
Barney Krucoff, Chief Data Officer, Office of the Chief Technology Officer
- ***Demonstration of New Open Data***  
Alexandre Santos, GIS IT Specialist, Office of the Chief Technology Officer
- ***Demonstration of MicroStrategy & Esri Integration***  
Martha Garcia Villalon, MicroStrategy with Matthew Viverito, Esri
- ***Data Inventory Support for Agencies***  
Barney Krucoff and Michael Bentivegna, OCTO
- ***New Crimemap.dc.gov user interface***  
Julie Kanzler, OCTO

### Meeting Notes

Capturing notable comments and questions from attendees.

#### Presenter: Barney Krucoff – House News

- Round of introductions so that we can meet new ADOs
- Under the new policy there will be two groups – ADO and AISO. This meeting will primarily be for ADOs
- Some attendees are appointed ADOs however not everyone is an official representative. We want everyone who is interested to participate on the team.



Presenter: Alex Howard – Remarks on DC’s Data Policy from the Sunlight Foundation  
(Paraphrased Notes)

- Sunlight Foundation contributed greatly to the creation of DC’s data policy. They have also worked with other jurisdictions.
- Sunlight is a 10-year-old national nonpartisan nonprofit working to get governments to be open by using technology.
- Proud that our city [Washington, DC] is stepping up. We are being competitive with other cities worldwide.
- It is up to [DC] government to properly disclose data and be good stewards of data that should not be in the public domain. Things that shouldn’t be on internet and sometimes end up.
- The [DC] data policy is not perfect but laws never are. It is a good policy where government will be doing the heavy lifting and we [the public] will be using.
- Across the country, governments are acting as stewards of data not as owners. The data belongs to the public.
- As you [Agency Data Officers] move through iterations and create machine readable data you will be getting/sharing data easier with each other. 80% of the time is spent cleaning up the data.
- Hold onto helping each other as your north star. Don’t lose track of each other to help navigate implementing this [DC data Policy].
- I am here to say...
  - that I appreciate the effort in getting you all in the same room.
  - having data be open by default unless there is a strong reason not to.
  - And that you have colleagues outside of government to assist.
- I am not here to tell you what to do
- When you think about this work, think about the experience you will create around the services that you [agencies] provide.
- Question: what does failure look like when it comes to open data?
  - When data is taken offline
  - When you put up PII data like SSN
  - Putting up data that increases risk to residents
- Question: what is your estimate of the number of datasets that the DC Enterprise Data Inventory will yield?
  - Thousands. Most credible estimates of agencies in cities is that a minority of their data is digital. DC is in a good place when it comes to digital data and would say is leading the nation.
- Question: is poor quality data even an excuse not to publish? Where do you draw the line?
  - Unfortunately, yes. But government should work to make it better and engage others [agencies and local community]. Organizations like Code for DC [<https://codefordc.org/>].

You will find willing partners. Do not stonewall.

- Question: in your experience, what kind of soft or hard skills might we pick up accidentally along the way and what do you recommend we pick up?
  - Curiosity, good judgement and skepticism.
  - Clearly, people that know how to use spreadsheets along with automation in data creation, publication and analysis.
  - You could learn more tools... like Learn R, Python,
  - A good understanding statistics
- Question: what are your thoughts on governments copyrighting information?
  - I wish that governments do not do it. It is a block in progress. It is locking up work that should not be.
  - I recommend that you think of “open by default” not just in the sense that something is machine readable. But also, keeping away restrictions that are licensed where ever you can. It will affect you internally too – amongst agencies.
  - So, stay away if you can. Work with your lawyers to avoid getting into contracts.

#### Presenter: Barney Krucoff – Data Policy Updates

- The clock started may 4<sup>th</sup> for all of us
- Some big changes are,
  - Copyrighted material exempted. An agency’s copyright will outweigh the policy and is noted in the policy. Hoping these will be few.
  - General counsel will need to review and sign off on data publication and will be tied/related to FOIA.
  - There will be training for all agency legal counsels. The CDO will reach out to them.
  - There will be a check box in the inventory asking if the dataset has been reviewed by counsel. However, we still want agencies to enter it into inventory.
- Make sure you know what “enterprise dataset” means. This will help you know what data to enter in the inventory. Page 7 of the policy.

*“a dataset that directly supports the mission of one or more public bodies. Typically, an enterprise dataset is stored in a named information technology system. For example, the District’s general ledger is a dataset hosted in the “System of Accounts and Records.” Typically, such named systems and the datasets they contain are accessible to multiple workforce members.”*

- OCTO is looking for volunteer agencies to work with as a starting point.
- Agencies have 30 days to name their ADO. If you are here, then you are assumed to be it.
- Inventory will be updated once a year comprehensively. This doesn’t mean we won’t update layers as updates come in.
- June 5<sup>th</sup>: we will begin entering data.
- How will we measure success? Here are measures that will likely end up in the Mayor’s dashboard,

- Inventory breadth: How many public bodies have listed at least one dataset?
  - # of # eligible bodies
  - # of # mandated bodies
- Inventory depth: What is the number of datasets inventoried?
  - #
- Other measures,
  - Is OCA reinforcing policy through the Stat and KPI processes?
  - Have Public Body Directors designated ADOs?
  - Have Public Body Directors designated AISOs?
  - Did OCTO release the data inventory tool within 60 days?
  - Has OCTO released a list of eligible and mandated public bodies?
  - Has OCTO provided public bodies training on the inventory process?
  - Has OCTO publicly released the Enterprise Dataset Inventory within 270 days of the release of the inventory tool?
  - What public bodies have not inventoried at least one dataset?
  - Percent of all Inventoried Enterprise Datasets reviewed by General Counsels or their designees.
  - Are there observed gaps in the data inventory?
  - What is the breakdown of datasets by classification 0 to 4?
  - Is the percentage of Level 0 datasets increasing year over year?
  - Percent of eligible agencies with accounts on tracking/processing system (currently FOIA Express)
  - Are there datasets that have been holistically released under FOIA not classified Level 1 or below?
  - Is FOIA request metadata being released as open data in a timely manner as prescribed by the policy?
  - Is FOIA request metadata being released as open data in a safe manner respectful of privacy and security concerns?
  - What datasets have been classified as level 1 and stated reason?

#### Presenter: Kevin Wilson – Security Procedures & Templated Data Sharing Agreements

- A presentation of The Lab @ DC's security procedures for data sharing between it and its agency customers.
- What is The Lab @ DC?
  - New group in Office of the City Administrator
  - Bring academic findings into government
  - Help agencies measure the effects of policy changes
- Often our projects are of a sensitive nature. Everything requires data and usually with multiple agencies.
- What the new data policy “does”
  - Establishes data taxonomy based on sensitivity and some legal requirements
  - Describes general transparency and sharing principles

- Describes general security principles
- What the new data policy “does not do”
  - Template interagency data use agreements
  - Explicit security protocols for data
  - Publication and disclosure protocols
- We [The Lab @ DC] had to develop security procedures and data sharing agreements that tailor to project needs. But, we built a template Data Use Agreement and Security Policy.
- You can’t share data the same way: education versus health data. Therefore, the data policy does not state publication principals and procedures. It would be too long.
- The Lab treats all data as Level 3. Level 4 requires negotiation with you [the agency]. HIPPA being the most restrictive of all.
- Our policy principles
  - *Access Controls* – two factor authentication, centralized, explicit authorization
  - *Encryption* – full disk encryption on PCs and cloud, transactions over SSL/TLS, NGINX config limits algos
  - *Deletion* – upon completion of project, all level 3 data deleted. Exceptions for replicability and anonymization
  - *Employee Restrictions* – all access VPN, avoid wifi access, WPA2 encryption on home net
- Template features include
  - explicit enumeration of purpose,
  - data security policy,
  - custodians,
  - code to produce data shared along with data,
  - unauthorized disclosure provisions
    - Defines OCA’s responsibilities when data is disclosed not pursuant to agreement
    - Requires OCA to make reasonable efforts to mitigate effects
  - publication restrictions
    - No raw Level 1+ data
    - Agency reviews analysis plan
    - Open Science Framework
- Question: you mentioned mostly working in the cloud. Would office 365 work for you? Would you consider AWS?
  - The question is where you process the data. OCA currently does not have 365. But we need computational power. We are using Azure.
  - Specifically OCA is using OCTO contract for Azure.
- Question: really appreciate the thought that’s gone in the data sharing agreement and security principals. My question is, in terms of district and adopting standards is this an on-ramp to get there?
  - From Barney Krucoff: we need to make this more uniform throughout the District. It is up to us in this group to start this work. The Lab’s example is good however we are not

yet ready.

- Question: internal handoff – what is that exactly?
  - If you create a space for an agency to upload data to. Then you are the only one who can access it. But you may designate others *specifically* to have access. The agency will have explicit purview. Comfortability with sharing data with others. There are some requirements to audit access (this is an attempt to try this).
- Question: what is the process for requesting data from agencies?
  - We request data because of a project. And we would involve the agency as a partner. So, it is not out of the blue.

Meeting Adjourned