District of Columbia Interagency Data Team

March 7, 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***
  Barney Krucoff, Interim Chief Technology Officer, Office of the Chief Technology Officer

- ***Using Data as a Strategic Asset***
  Amber Ivey, Senior Associate, Data as a Strategic Asset, Pew Charitable Trusts

- ***Public Publication of the Enterprise Data Inventory***
  Barney Krucoff, Interim Chief Technology Officer, Office of the Chief Technology Officer

- ***CDO's Annual Report to Mayor Bowser***
  Barney Krucoff, Interim Chief Technology Officer, Office of the Chief Technology Officer

- ***Open Discussion***
  Specifically, input for the Data Policy annual report including the inventory process and recommendations to Council

Meeting Notes

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

Presenter: Barney Krucoff – Welcome and News

- Major activities,
    - OCTO is reviewing datasets in the Enterprise Data Inventory for completion
    - We will be publishing the inventory March 11th
    - Also, we are finalizing the annual report. We will be publishing the annual report for the Mayor.

Presenters: Mario

- New Data and Updated
    - *Food Commercial Kit*
    - *Food Waste Drop Off*
    - Impervious Surface 2017
    - Liquor License
    - *Hunger Heat Map*
    - Bus Line
    - Bus Stop
    - Camera Enforcement
    - District Property
    - Community Garden
    - *Community Supported Agriculture*
    - *District Development (WDCEP)*
    - *Emergency Food Provider*
    - *Group Meal Center*
    - *Healthy Corner Store*
    - *Joyful Market*
    - *Low Food Access Area*
    - Notary Public
    - Office Aging Lead Agency Site
    - Other Bus Line
    - Planimetric Redaction Area
    - *School Garden*
    - Transportation Analysis Zone
    - *Urban Agriculture*
    - *Women Infant Child (WIC) Store*
    - Ward

- *Question*: are there any data for adding to network analysis tools including details such as speed limits
  - *Answer*: DDOT is working on deploying Esri's Roads and Highways

Presenter: Amber Ivey – Using Data as a Strategic Asset

- Per performs evidence based research – the team is Data as a Strategic Asset (DASA)
- Team's goal is to Initiative to increase knowledge of how data is being utilized in all 50 states and the District of Columbia.
- There is a full report available online at www.pewtrusts.org/strategic-data
- State and local governments contain a lot of data in their systems. Are states really using the data to its potential? They use it for reporting back to the federal government. How else? They're using data in 4 strategic ways,
  - States are crafting policy responses to complex problems – such as opioid and traffic analyses
  - Using data for improving service delivery
  - Managing existing resources. For example, Delaware is using GPS data to save millions reducing fuel cost and better resource allocation.
  - Examine policy and program effectiveness
- There are many factors posing challenges to state data work. See the trend in slide presentation
  - Staffing, data accessibility, data quality, governmental culture, funding, technology, interagency cooperation, leadership buy-in, infrastructure, legality and data security
  - Why can't states share information? State officials cite STAFFING as the greatest obstacle
- There are many factors that are supporting state data work. See the trend in slide presentation.
  - Same as challenges however states cite LEADERSHIP as the greatest support for analyzing data for decision-making.
- There are 5 key actions to promote data and decision making,
  - Plan ahead to build guiding goals. Developing a data strategy where everyone understands the future. Have you set up data governance structure? Are you doing an inventory? A good example is Virginia. They set up the Commonwealth Enterprise Information Architecture Strategy. They went through a year process of researching any mention of "data" and finding limitations of sharing.
  - Building capacity – mainly with stakeholders. Get highly skilled staff. Leverage partnerships. Find dedicated funding. Example the Lab @ DC. There is money and skilled researchers, data scientists. Staff use evidence from academia and industry to design policies and programs, and then apply methods like randomized controlled trials to evaluate effectiveness.
  - Sharing – ensuring data is accessible. Improving data quality and enterprise wide view. Example, Massachusetts. They have a problem with opioids. They put in place in law to require agencies to share administrative datasets relating to opioids. Where people are dying, where prescriptions are being filled? The whole idea of "my data" is no good create data sharing agreements.

- Analyze data- take the spreadsheet and turn it into something understandable. Even graphs are not always as impactful as they should be. Example, use visualizations like dashboards. Figure out how to tell the story. In Indiana's management performance HUB helped decreased infant mortality rates. They found that individuals weren't getting prenatal care. Massachusetts took their data and connected to the story. Brought awareness and built a taskforce. Be analysts and good storytellers
- Sustain – how do you sustain all the work into new administrations and leadership? Enhance leadership commitment? Virginia had a focus on homelessness. So just before the election VA started a 100 day challenge before the new administration came in. The challenge was a success. After this then the new administration "had" a "win." Also, enact legislation requiring keeping programs in place.
  - Change of data culture – Texas put in place a chief data officer (coordinator). During Hurricane Harvey, the state was able to use the framework put into place by the data coordinator and stakeholders to map GIS data.
- Comments from DC's Chief Data Officer. How is DC's performance on the above?
  - DC has a data policy however not entirely a strategy, nor data sharing agreement
  - DC has hired skilled staff at OCTO, Lab@DC and other agencies
  - Funding is unfortunately all over the place.
  - Sustaining hasn't been as difficult here and we are okay
  - Culture is good and trying to get agency analysts talking, closer together with committees like the Interagency Data Team

Presenters: Peter Casey – Enterprise Dataset Inventory Process

The numbers and details below were draft at the time of presentation (March 7, 2018). The finalized statistics and inventory are now published on http://opendata.dc.gov/pages/cdo-annual-report

- Over 90% of mayoral agencies participated. The ones did not are very small under Executive Office of the Mayor itself. We may take a second look at whether these agencies should have been included this first time.
- Non-mayoral agencies had little participation because they are not required. Less than 1/3.
- DCPS, DDOT, DOH, OCTO, DOEE were biggest contributors. Median contributing datasets were 14; so half of the agencies that participated had 14 or more.
- We were trying to get a sense or metric to compare. So, we used agency budgets as a proxy. As expected agencies with larger budgets had more datasets. We left out DC Public Schools and Department of Health Care Finance because their budgets are huge and really skewed the analysis visualizations.
- Classifications breakdown – over a 1/3 were classified as Level 0 Open and includes data in data.dc.gov. The least common was Level 1 Public No Proactively Released (under 10%). Most common was Level 3 Confidential. This included federally protected data by law. Lots of health data governed by HIPPA.
- Approval by general counsel. Over 90% of datasets were approved. Over 80% of agencies have all datasets approved by general counsel.

- 65% of data is already on data.dc.gov. Leaving 40% that we will be prioritizing for redaction (if needed) and release.
- Government operations is the most common category. Then education and health.
- *Question*: was there a common reason as to why non-mayoral agencies didn't participate?
- *Answer*: we believe it is the lack of a requirement that played a part. It was a lot of work and it was a lot to ask of them. Sometimes independent feel a need to exert independence.

Presenter: Barney Krucoff - Data team goals for the coming year

- Goals
  - There will be an annual report that must come out at the same time as inventory. This year is awkward because of when policy was signed. Future dates of inventory should be November 1st.
  - OCTO would like to extend the eMOU system. It could be used for multilateral data sharing agreements. Usually OCTO is middleman. We need to adapt and have templates for agency-to-agency sharing agreements (along with OCTO).
  - Develop a modernized data backend to support analysis for programs such as smart cities. Use big data.
  - Provide more assistance to agencies for proactive FOIA releases. we will "modernize" agency websites to include datasets that have already gone through FOIA.
  - Formalize the process for agencies to submit data and eventual publication. DC has an informal process right now between the OCTO Data team and subject matter experts
- Draft Recommendations (check slides)
  - For city administrator – develop KPI to encourage agency participation next year
  - For the mayor –
    - Change the annual report date from November 1st to Monday of sunshine week
    - Adopt reasonable and uniform retention policy for email. OCTO has been saving emails from 1998 in terabytes. We have no policy for managing them, space.
  - For council – encourage agency participation in EDI and have council get data from the inventory rather than ask agencies about data for performance hearings
  - For mayor and council –
    - legislation to strengthen and clarify FOIA exemptions for critical infrastructure information
    - legislation to mandate participation of independent agencies so we can have a true representation of all DC data.
- *Question/Comment*: legislation mandating participating isn't exactly going to be successful without the proper resources. If there are no resources, then they won't do it. I recommend that if this is something we are going to advocate for, we need to make sure we're not just stopping with words on paper
- *Question*: agencies also have a parallel effort to catalog an application for DC council. How old, who maintains, etc. Can this be consolidated along with the data inventory?

- *Answer*; the EDI currently asks for system and application names that data exist in. However we are going to hide this from the public publication of EDI on open data.
- *Question*: will you address removing application and system names from the descriptions of the inventory and also the name of the datasets?
- *Answer*: This is a new idea to remove systems and applications because of security risk. Yes, we are currently working through this.
- *Question*: are you also considering completeness with the public facing descriptions? Is there an assessment for the descriptions being enough or too much?
- *Answer*: our descriptions are fairly good. What we are doing now is looking at URLs to make sure they are not pointing to internal systems or applications

Presenter: Michael Bentivegna – Tableau Server and Microstrategy
- Clustered Internal Tableau Servers are operational Production
- Clustered External Tableau Servers have been in operation, ETL update process testing was delayed due to a need to update the Informatic server version, which is complete. Will be determining best work flow and testing the update process with DPW over the next month.
- We will be scheduling a quarterly Tableau Agency Site administrator meeting over the next month.
- Microstrategy connectors in place: DC Open Data site, XML, Tableau (From Tableau to MicroStrategy), Elastic Search, SalesForce, QuickBase, OneDrive, Hadoop
- Microstrategy dashboards in place: QuickBase Usage dashboard, Public Litter Receptacle Dashboard and Mobile App (with Write-Back functionality), FOIA Employee Dashboard
- Clustering in place: Intelligence Server clustering, Web clustering (In progress)

Meeting adjourned
Next meeting scheduled for May 2, 2018