

District of Columbia Interagency Data Team

November 6, 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
- **Data Science & GIS Analysts Working Group**
Peter Casey, Senior Data Scientist, Lab @ DC and Office of the Chief Technology Officer
Eva Reid, Senior GIS Analyst, Office of the Chief Technology Officer
- **FY19 Agency Data Submission Process for Level 0 Datasets**
Mario Field, Data Curation PM, Office of the Chief Technology Officer
- **Enterprise Data Inventory – New System in Testing**
Michael Bentivegna, Data Visualization and Analysis PM, Office of the Chief Technology Officer

Meeting Notes

Best attempt to capture notable comments and questions from attendees (paraphrased). For complete presentation visit, <https://octo.dc.gov/node/1310836>

Presenter: Michael Bentivenga (OCTO) – Welcome and News

- Tableau Desktop & Server
 - Still on version 2018.1.1 and soon users will be able to use Google sheets as a data source in production
 - Expect a major upgrade to Tableau Desktop version 2018.2 however please do not upgrade just yet. It's a major change we need to test compatibility with server first.
 - DC gov now has advanced Tableau support that comes with Architecture Review sessions with Tableau tech admins. OCTO plans to schedule such a review.
- Tableau Public for DC
 - Hyper is not currently supported by informatica Tableau extension.

- Box plug-in will be tested as a public data source
- Microstrategy
 - Currently moving Linux over to Windows
 - Also creating public facing web servers
- ArcGIS Server and Basemap upgrades
 - Some issues were found when bringing in WMS but believed to have been resolved
- Tableau training
 - Only two seats left for December intermediate class
 - For agencies looking to train large groups of employees, OCTO could do on-site training
 - New training for January and February will be announced soon
- Microstrategy
 - Still have a good number of seats for December classes

Presenter: Peter Casey – Data Science Working Group

- Peter is a fellow at the Lab @ DC. He helps bring higher level analytics to help answer questions with dc government data
- This group is only for government employees. We meet to talk about,
 - State of data science within DC and/or industry
 - Specific project issues that other data scientists & analysts are running into
- Meetings usually held mid-Wednesday of month. Next meeting is on November 14th at the Wilson building. Email peter.casey@dc.gov for more information (DC gov employees and contractors, only)
- Working group started in August 2018 and so far has met three times
- *Question:* what topics have you discussed?
 - First meeting was an introduction, meet and greet, and chatted about challenges with DC data scientist work – especially having appropriate access and overhead of getting across legal barriers to data sharing.
 - Second meeting were two presentations. One for rat infestation. Department of Human Services did a presentation looking for potential biases in housing resource selection
 - Recently, sessions for ethical issues with machine learning. Tools by Johns Hopkins University to help us walk through what kinds of issues we need to prepare for that will come up. We actually had a guest speaker from John Hopkins
- *Question:* have you talked about data sharing templates?
 - We haven't approached that just yet, being a newer working group. Most of these folks are technicians rather than legal scholars. Though I have recently taken on the data sharing "hat."
- *Question:* how do the two groups relate? Interagency Data Team and data science working group
 - Think of it as a subgroup of Interagency Data Team. People that touch and work with the data. A special interest group. It is not open to the public however presentations from outside DC government organizations are welcome.

Presenters: Mario Field – Agency Data Submission Process for Level 0 Datasets

- First, performed an internal review of the 189 open data layers to clarify what's really open data.
 - 15 entries are just documentation and should not be considered as data
 - 32 entries are just duplicates of each other or already on open data
 - This actually leaves 142 Level 0 datasets to work with
- DDOT and OCTO have the most of these with many other agencies having 3 or less
- OCTO liaisons will be working with Agency Data Officers to capture these
- Why do we have a data submission process?
 - Provides transparency
 - Allows public and leadership to understand overall availability
 - Document requirements and guidelines
- We researched other jurisdictions and their data submission process. We modeled much of our process after New York state. San Francisco has a good one that is an online submission process. Both of which have details that are applicable to DC.
- An Overview,
 - We will ask agencies to submit a service ticket for tracking
 - Preparation will include a submission packet prepared by both OCTO and the agency
 - Consolation for completeness
 - Curation – OCTO will QA/QC data for: set metadata, transformation processes between systems, data dictionaries, abstracts, meeting field definition requirements...
 - [view presentation for more overview notes]
- Data submission packet
 - Making this more formalized
 - Making these elements required for us to publish
 - A submission form is needed which is now a Quickbase application
 - Data dictionary, ideal however this should be required
 - A data mapping document might be required when identifying joins/relates between agency systems and OCTO
 - Approval form might be needed from general counsel or agency director
 - Data review summary report which will include testing results
- Data submission form will be integrated with the Enterprise Data Inventory. Think of this form as getting into the weeds of EDI. It will have details like database instances, types and sometimes credentials (not passwords).
- Data curation team at OCTO,
 - Would perform processing such as running through the Master Address Repository and containing required MAR fields, address verification
 - Data rules such as how spreadsheets are formatted, appropriate column naming and aliasing, character limiting... some geographic rules such as topology.
 - Metadata documentation will have required and optional elements
- Considering and implementing security practices,

- Will data need to be encrypted?
- Database level with credentialing?
- Map service level with credentialing and token?
- Approval forms,
 - Sign off from Agency Data Officer and General Counsel, at least
 - Sign off from OCTO Data Curation program manager and OCTO General Counsel
 - All Level 2 data will need agency director and OCTO director sign off
- Other publication steps that may be necessary,
 - Scheduled future review sessions for datasets documented in Enterprise Data Inventory
 - Archiving of older data after replacing with new
- Action Items
 - Finalize formal workflow
 - Author training materials
 - Eventually develop an online submission workflow
 - Identify how agencies will submit errors found after publication
- *Question:* what kinds of errors will agencies be able to submit?
 - Example, HSEMA just told us a lot of churches are no longer active. Right now, I get an email requesting an update. Sometimes it is just a misspelling. We need a process that is not email based.
- *Question:* for preparation step, will agencies be working with OCTO liaisons or working directly with OCTO Data curation team?
 - It will likely be a mix and partnership between data curation team and liaisons. Example liaison and metadata analyst worked with OCTFME to publish their two datasets
- *Question:* when are we actually publishing the book?
 - It is not up yet. Our target is for end of December. All this will eventually be published online for the public to view.
- *Question:* we think that OCTO should be creating the ticket since they are driving this process. Agency should not have to submit ticket.
 - It will be a mix. Sometimes OCTO will create the ticket on behalf. But, the agency is the steward and should be documented so.
- *Question:* what is the timeline for this process? An example? I am concerned with staffing availability to complete this. For example, completing a data dictionary.
 - For GIS data, in many cases OCTO did the leg work for bringing data to a proper format and requirements for publication. OCTO liaisons can work with ADO (interview process?) but the agency is subject matter expert
 - The timeline depends on the type of dataset – spreadsheet versus enterprise database. Spreadsheets are easier. Enterprise datasets, OCTO connecting to agency systems take time to work through. Example, live crime data from MPD and live permitting data from DDOT & DCRA.

- *Question:* what about agencies with Level 2 data? Are we required to go through this?
 - No, we are currently prioritizing only Level 0 datasets. Agencies like Department of Human Services do not have Level 0. Though eventually OCTO would ask agencies to again revisit their Enterprise Dataset Inventory.
- *Question:* when will this plan be finalized?
 - Targeting for end of December. However, we will likely have to make edits as we go along as the first round.

Presenter: Michael Bentivegna – Preview of New EDI Tools

- Round of “raise your hand” questions
 - Who here is still an Agency Data Officer?
 - Who here actually entered data into the Enterprise Data Inventory?
- If you did enter data into the EDI, then you know it was not an easy system to work with. We have been working to create a new system that will be much easier. It’s currently in testing.
- Our new system is the result of your feedback from survey’s we sent out earlier this year.
- Mayor’s order requires us (DC agencies) to manually review the EDI and make sure nothing has changed or make needed changes. And to also document any new datasets.
- The new system was tested and will be in production by December 1st.
- The next round inventory must be done by OCTO and Agency Data Officers and reviewed by council by February 11, 2019
- *Live Demo of the data entry tools for agencies. Note that there will eventually be a public “view” version
 - For documenting applications, we are requesting less information. Remember our previous version asked for too much technical detail most ADO would not know. Also, you will be able to see all applications listed easily without hunting
 - For data center we are also requesting less information – less technical
 - Getting general counsel to sign into the application and approve was not easy. For approval process, we have included an ability for ADO to approve on behalf of general counsel. However we still prefer general counsel. They will receive an email notification that you have finalized a dataset and ready for approval.
 - Color coding per records. For example, last year’s data will be in green to highlight “please review.”
 - You’ll have the ability to retire a dataset but must enter a reason. It will not be present in next year’s inventory.
 - The system will have a “notes” section that tracks changes over time. You’ll see who has made changes and when it occurred. Also, when general counsel has approved.
 - The new app will allow for in-line edits from list rather than opening up the form for each record individually. This will speed up data entry.
 - As you work with the system, changes you make will be highlighted in yellow. It helps you keep track of several changes and then you can click save or save all.

- General counsel will receive weekly updates at the end of the week. So ADO should make sure to get your updates in before every Friday
- *Question:* will there be ability to filter?
 - Yes, you'll be able to filter off things like classification. Filtering options appear next to the field name
- *Question:* is the back end still Quickbase?
 - No, this is totally new. It was too difficult to work off of Quickbase.
- *Question:* do you already have the names and roles for agencies?
 - We have last year's ADO list however OCTO liaisons will be contacting agencies to update any changes in staffing.
- Also, this system will be accessible outside of DC private network. It will be in Microsoft Azure and allow for those agencies not on dc.gov to better participate. Example, DC Water.
- *Question:* will this allow for changes after EDI publication?
 - No, there is an eventual freeze at which point we deliver to Council and Mayor. The freeze date is February 11, 2018, when agencies must complete their inventory.
- Remember that this redesign does come from feedback received from surveys sent to all ADOs after last year's inventory.

General question for attendees... are there topics that you would like us to discuss here? If so, send them along to us. OCTO wants to open up the agenda to other agencies too.

Meeting adjourned

Next meeting tentatively scheduled for January 2019