

## District of Columbia Interagency Data Team

**January 10, 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, Chief Data Officer, Office of the Chief Technology Officer
- **How Code for DC Uses Government Data Working with Agencies**  
Kate Rabinowitz and Lizzie Ellis, Co-Captains, Code for DC
- **Oracle Service Oriented Architecture (SOA) Suite**  
Julie Kanzler and Siva Abbineni, OCTO

*OCTO has several technologies to assist with the exchange of data between systems. These include Extract Transform and Load Tools (Informatica, and Safe Software), Service Oriented Architecture Suite (Oracle), and API Gateway.*

*This is the second in a series of presentations covering these technologies.*

- **Feedback to the Enterprise Dataset Inventory Process**  
Barney Krucoff and Eva Stern, 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

- OCTO is reviewing datasets in the Enterprise Data Inventory.
- We're seeing many datasets without general counsel review. Please work with your agency general counsel to sign off at the bottom of each dataset.
- Descriptions and classifications are best if citing FOIA or specifics of what is contained in the dataset to assign Level 3 classification.
- DC Council usually asks agencies for what datasets they're producing. The EDI can help agencies answer these questions.
- Note that when agencies are FOIA'd, they are not required to create documentation or data. They are required to share what already exists.
- Barney Krucoff was named Interim Chief Technology Officer to DC government.

Presenters: Kate Rabinowitz and Lizzie Ellis – How Code for DC Uses Government Data Working with Agencies

### Complete Speaker Notes

- Code for DC is the DC Chapter of the Code for America Brigades. We bring together community members and subject matter experts with data enthusiasts and techies to help make a better DC for all of its residents.
- Goal is to help make our city a better place with some kind of technology and open data. You don't have to be a coder.
- Some of our more recent projects include,
  - DCPS
    - Per student funding project of DCPS: <http://dcpsbudget.ourdcschools.org/>
    - App that shows where dcps students are coming from: <http://edu.codefordc.org/>
    - *Question:* is the data where students' residence publicly available?
    - *Answer:* no, it's provided by OSSE. But we understand the limitations and complication with sharing such data. We work to de-identify the data. Largely done in partnership with OSSE.
  - Urban Institute & CHNED
    - Last year Code for DC had its first ever grant to create Housing Insights: <http://housinginsights.org/>

- Partnered with the Urban Institute and Coalition for Nonprofit Housing & Economic Development (CHNED)
    - [More Sample Projects: http://bit.ly/2ErPkTC](http://bit.ly/2ErPkTC)
- Continuing project for rat abatement by helping to build a model to predict rodent issues. And also a web portal to view 311 requests by area.
- Code for DC Open Data catalog: <http://data.codefordc.org/> is our open data portal with over 500 datasets. Includes things from all over the internet including derivatives.
  - The data portal includes a Data Wish List where anyone can submit
  - Built using CKAN. Challenge in hosting large datasets.
  - *Question:* what are some ways you're scraping pdfs for data
  - *Answer:* we use <http://tabula.technology/> but it only works with table formats in pdf. Not narratives. However, we sometimes question whether the data really open if it's only in pdf format.
  - *Question:* do you have a count of datasets in your data portal that aren't in <http://opendata.dc.gov/>?
  - *Answer:* we do include many layers, source information and credit for who found it
  - *Question:* for layers that you have that come from DC government's portal are you linking to them directly or creating static copies? Concern is stale data.
  - *Answer:* looks like it comes through as a static link.
- Code for DC is Here to make the extra improvement to data. We're not here to just scrape all data.
- We would like for the DC government data portal to be better so that ours doesn't need to exist. Connect
- We would like to see subject matter experts noted with the datasets
- Code for DC would like to partner with everyone to make the best use of DC data and app. Happy to provide user testing and feedback for any applications you'll be launching
  - *Question:* you mentioned that you can provide some services to agencies. Is this voluntary or some kind of paid? It's actually harder to get free work because sometimes DC gov is not allowed to ask for free work.
  - *Answer:* we usually do free work. The grant was the first where we got paid for it. We'll likely not be doing formal paid contracting of our volunteers.
- Code for dc shares data as complete as possible without going over that "line" – security and privacy
- For apps we build, we prefer open source – build it once and use it many times with minimal customization effort
  - *Question:* do you collect data on how much your apps and tools are being used?
  - *Answer:* it's something that we should be looking for however our target audience can be fairly targeted rather than average users.

Presenter: Julie Kanzler and Siva Abbineni - Oracle Service Oriented Architecture Suite

- In the last Interagency Data Team meeting we talked about API for sharing data across systems. This time we're talking about SOA – another service for exchanging data across enterprise systems.
- Here is an analogy, let's say your employer offers to subsidize your transit to work – any mode. But, then you have to interact with each individually to sign up and manage accounts or payments– bus, boat or car. This gets complicated for the user. Well, let's say you get a one transit pass by your employer to use anywhere. You just worry about the transit pass. The transit pass would be "SOA." However, this is an oversimplification of SOA but it's a good way to conceptualize.
- It gets a little more complicated across enterprise systems.
  - Spaghetti infrastructure – when systems have lots of unique connections.
  - SOA – removes the spaghetti and is more configuration. A more repeatable, flexible process where we're not worried about upgrading systems.
- Current challenges – when you make a small change in a service used by different applications, that change could affect many without realizing. SOA suite provides flexibility to deploy new versions while allowing older apps to continue with older service while they update to the new version.
- Why use middleware, SOA? – look at slide for answers
  - Maximize cost efficiency
  - Remove complexity
  - Support heterogeneous environments
  - Mask differences in OS/Language
  - Provides flexibility
- There are many adapters that SOA provides for applications, databases, mainframe, cloud, SaaS and more
- *Question:* since you already have established with PASS are you open to adding new fields for other agencies and sharing?
- *Answer:* we can but would have to know or broker an agreement with other agencies who own the data.
- *Question:* is there a catalog of services similar to what was in your slides?
- *Answer:* it's a work in progress. You as the ADO come with your needs.
- *Question:* is this a technology already developed and can we leverage or is there a cost for us?
- *Answer:* both, but it's a lot cheaper. Just because it's there doesn't mean it's totally finished. It is a robust system that is well utilized but could be more.
- *Question:* it would help to let us know where these robust integrations would make sense and where they would not, where other integrations make more sense.
- *Answer:* our presentation series is covering other forms of integrations like last week's API gateway. Next, we'll have Extract Transform Load services. We should probably have a wrap up at the end of the series where we can do comparisons.

## Presenters: Eva Stern and Barney Krucoff – Feedback to the Enterprise Dataset Inventory Process

### *Enterprise Dataset Inventory*

- So far there are 839 Datasets were entered by December 11 (entry deadline). The numbers are increasing as agencies continue to enter datasets
- There are a lot of acronyms that need to be spelled out
- Each OCTO liaison is meeting with barney and Turna to run through issues and clarifications
- Expect to hear from your liaison with edit suggestions and questions
- Final edits due January 26<sup>th</sup>
- Participant survey – so far have 5 responses. Would like to have more so please fill it out.

For those of you who participated in the Enterprise Data Inventory, OCTO is interested in learning about your experience with the EDI tool and the process of adding to the inventory.

Please take a moment to [share your experience using this questionnaire](https://www.surveymonkey.com/r/YW85H7G). This will help plan for the FY 2019 inventory. <https://www.surveymonkey.com/r/YW85H7G>

### *Annual Report to Council and Mayor*

OCTO is looking for ideas to include in the Data Policy annual report. Thus far,

- Critical infrastructure captured in FOIA
- Identify Level 0 datasets, but not yet in Open Data DC
- Improvements to the EDI tools and process
- Flag Level 3 datasets for possible redaction and publication
- What are your ideas? Please send to [barney.krucoff@dc.gov](mailto:barney.krucoff@dc.gov).

From the policy:

*Publishing, in coordination with ADOs and the OGAG, an annual report to the Mayor beginning November 1, 2017. At a minimum, the report shall include recommended changes to this policy and other relevant policies, recommended legislation, a list of datasets opened during the prior fiscal year, and a list of datasets planned to be opened during the then-current fiscal year.*