Open Data Council

Draft Meeting Minutes

4/4/2016

1:03pm Meeting called to order

**Council Members present:**

* Rhonda Lehman - DTI
* Meredith Tweedie – Governor’s Office
* Mike Mahaffie for Bob Zimmerman – DNREC
* Steve Yeatman - DSCYF
* David Blowman - DOE
* Li Wen Lin – DelDOT
* Dean Stotler – OMB
* Dana Rohrbough - GIC
* Jill Fredel - DHSS

**Others:**

* James Collins – DTI
* Mike Hojnicki – DTI
* Sheila Hunton - GIC
* David Ginzburg – Open Data Delaware/Zip Code Wilmington
* Chris Williams - – Open Data Delaware/Technical.ly Delaware

**Agenda**

* Welcome/Introductions
* Review/Approve Previous Meeting Minutes
* Update/Review
* Presentation from Tyler Kleycamp, CT Chief Data Officer
* Approval Process
* Open Discussion
* Next Steps
* Public Comment

Welcome/Introductions

Review/Approve Previous Meeting Minutes

* The previous meeting minutes from 3/29/16 were approved.

Update/Review

* A purchase order for Socrata software license / Open Data Portal was issued in late March 2016
* On 3/15/16, Rhonda attended an Open Data Meet-up
  + DOE is very interested in open data and provides data to the public today
* Dana & Rhonda met with Dan Shortridge from the Dept. of Agriculture to discuss open data
* GIC is working on developing a website so others can suggest data to be uploaded to the Open Data Portal
* The Open Data Council will be discussing guidelines to determine what types of data should be posted to the Open Data Portal

## **Presentation: Tyler Kleykamp, Chief Data Officer for the State of Connecticut (Open Data Portal)**

* Connecticut’s Open Data Portal has been live for 2 years
* Per the Governor of Connecticut’s executive order, each agency must have a separate data officer
* Mr. Kleycamp was appointed as a data officer and no additional staff was hired
* Connecticut has data stewards who are existing employees

Connecticut’s Process for Building and Maintaining their Open Data Portal

* Each agency had 90 days to submit “low hanging fruit” data for their initial release
* The day they went live, Connecticut had 100 datasets to start
* Connecticut has a validation process and uses Pilot Fish, a custom proprietary open data tool to begin building their open data portal
* Data stewards have been appointed within each agency
  + This is in contrast to the state of New York’s process in which a specific team was hired to deal with open data

Question and Answer with Mr. Kleycamp

* Dean Stotler of Office of Management and Budgeting (OMB) inquired about Connecticut’s process for refreshing data. Mr. Stotler shared OMB’s process for refreshing their data.
  + Data is refreshed monthly per Mr. Kleycamp
    - Mr. Stotler also inquired how does Connecticut establish the frequency of refreshing data
    - Mr. Stotler shared that OMB, for example, publishes employee payroll annually (per FOIA requests) and other sets of data is published monthly and periodically while protecting Personal Information in accordance with OMB’s and other policies
      * Mr. Kleycamp responded, Connecticut does not currently measure FOIA & ROI inquiries.
      * Many of the ROI requests are from within state government
* Ms. Lehman inquired is there a data quality process?
  + Yes, after and the process is after the fact depending upon having resources
  + Mr. Kleycamp explained his experience has been that agencies review data quality
* Mr. Stotler of OMB inquired the specifics of ensuring accuracy and appropriateness of the data.
  + Mr. Kleycamp responded that it is the responsibility of the agencies to ensure the accuracy and appropriateness of the data internally first before it is distributed.
  + Agencies work to ensure that data is shared a level that does not personally identify individuals or groups of people
* Ms. Lehman asked about CT working with the community and local governments.
  + CT does not have civic hacking groups yet. There are people interested in policy issues on the state level. But, there is no shortage of people seeking data from the state.
* Mr. Mahaffie asked how do you deal with agencies who are already publishing data on their website?
  + CT will often just link to the data, for example XML of roadway incidents is not going to be forced into the open data portal. There is a bit of a struggle with how to handle this.
* Mr. Mahaffie of the Department of Natural Resources, (DNREC), are any of Connecticut’s datasets that are ultimately pushed to the Federal Government?
  + Connecticut carefully adheres to what the feds, states and other parties may share
  + Some sets of data are not authorized for use to be shared or in cases where there are no clear guidelines
* Mr. Blowman inquired about how geographical data is managed.
  + CT constituents are interested in what is going on in their neighborhood as opposed to seeking broader levels of data
  + An example is educational and neighborhood school data
* Mr. Mahaffie inquired about duplicate requests for data that is already published.
  + Connecticut may links the data if it is compliant with open data
* Mr. Blowman asked about web traffic coming to the portal.
  + Traffic has been consistent.
    - 1/3 of traffic is from state government
    - 1/3 of users are state employees, from there it varies
    - ¼ traffic is direct
  + CT’s statewide search - does include data sets
* Ms. Fredel asked about Connecticut’s addiction data
  + Mr. Kleycamp provided some visual examples

**Approval Process**

Ms. Lehman explained the range of dataset approval processes for publishing data, NY with a more stringent policy and CT with a decentralized approach.

Mr. Stotler said that the agency is the business owner, and agency should have the responsibility of determining if the data is eligible for publishing. Further, when there is an interest from the public but data cannot be released the agency has an obligation to protect personal identifiable information or other confidential information like health information. Mr. Stotler said it seems like NY has a lot of sign offs in their data approval process, how does that pass from administration to administration?

Ms. Lehman explained it is left to the agency to determine they could change which data to publish if they want. In CT, agency staff have been given the keys to publish direct to the open data portal.

Mr. Yeatman talked about the role of the council – the council is to prioritize datasets. The council will need to know about what should be public vs. non public. Further, what is “a dataset”?

Sec. Collins explained that “a dataset” is one file, so take for example Professional Regulation data, this would be one dataset.

Mr. Mahaffie spoke about the need for agencies to do work on the file in the beginning to make the data clean, and combine files if possible, to make them “one dataset”.

Ms. Fredel since the vendor license allows us up to 150 datasets the first datasets we should prioritize and publish should have a good variety of topics included. Variety is important.

Ms. Lin what does executive order define as transparency?

Ms. Rohrbough said that the executive order does not further define the word transparency.

Mr. Stotler asked what should be in the inventory? Such as, location data – should this be standard geography, rolled up data, census track, or town? Sometimes the law requires data collection at the census track.

Mr. Blowman is “personal identifiable information” defined at the federal or state level? A clear answer was not immediately available.

Sec. Collins talked about funding going forward, and that agencies who want to publish even more datasets than what our current contract allows, then those agencies will be asked to support the licensing costs.

Ms. Lehman explained that the Council will be asked to do a prioritization of initial datasets they would like to make available. Each Councilmember will need to fill out the excel prioritization of their initial datasets.

Sec. Collins asked Councilmembers to consider including datasets in the prioritization that constituents want to see available.

Mr. Yeatman gave the example that the Kids Department deals with FOIA requests about an individual or a facility and some of this might be confidential data.

Sec. Collins then said, consider how do we wrap that up to publish about a facility? That way, it takes the confidential data out of the dataset.

Ms. Fredel explained DHSS would like to use datasets like drinking water reports, nursing home inspections, and restaurant inspections and make them easier to use online. She further asked, it’s about viewing data as an opportunity or a problem. DHSS views data as an opportunity.

Mr. Mahaffie said DNREC is interested in data such as the environmental navigator data. He is talking with IT staff to see what has already been set up like First Map. He is also interested in finding out more about which data is FOIA’d most frequently.

**Public comment**

Mr. Ginzburg encouraged the Council to look at the open data handbook (<http://opendatahandbook.org/>) and to look at the work Joshua Tauberer is doing in Washington DC on GovTrack <https://www.govtrack.us/> which is a website that presents analytics on the US Congress.

Mr. Williams inquired about the cost for the open data platform. How many datasets, is there a specific data size limit, are you charged based on bandwidth, and what about getting constant updates – will the data be replaced?

Ms. Lehman explained, it will cost $5000 per month for 150 datasets. There is no limit on size, and no bandwidth restrictions. Data will be replaced as there are updates to those datasets.

Mr. Williams suggested for federal datasets that it would be better for Delaware to link to it rather than republish using our 150 datasets. He suggested it would be great to have a list of where to go for those federal datasets.

Mr. Mahaffie explained that in the case with DNREC, the state agency reports to the federal government with data and then the data changes or is added to at the federal level.

Mr. Williams asked if the State Police are part of this effort?

Ms. Lehman said they are not named in the Executive Order, but that doesn’t mean they cannot participate.

Ms. Fredel said that they are very interested in police data – for example, taking addiction data and law enforcement data – that data mashup can help figure out a policy problem.

Mr. Ginzburg asked if the open data portal will have a copyright?

Ms. Lehman said, there is no copyright.

GInzburg mentioned he read an article two weeks ago about copyright data, he suggested to look it up. Creative commons license is not preferred because the data should have no restrictions.

Ms. Lehman said data should be identified as public domain.

Mr. Ginzburg – he will look up the article – might have referenced public domain.

Ms. Lehman thanked the Council and the members of the public for their participation.

3:11pm adjourned