Delaware Open Data Council

# Meeting Minutes

# Feb. 29, 2016

## Opening

The first meeting of the Delaware Open Data Council was called to order at 1pm on Feb. 29, 2016 in DTI by Ms. Lehman, Chair.

## Attending

The Honorable Governor Jack Markell, James Collins, (DTI); Rhonda Lehman (DTI) – Chair; Dana Rohrbough, (DOS/GIC); Steven Yeatman, (DSCYF); Li Wen Lin, (DelDOT); Dean Stotler, (OMB); Jill Fredel, (DHSS); Meredith Tweedie, (Governor); Robert Zimmerman, (DNREC); David Blowman, (DOE); Charles McLeod, (DOS); Bob Rhodes, (DTI); Mike Mahaffie (DNREC); Sheila Hunton, (DOS/GIC), Baljinder Kamboj (DTI), Rhonda Ramsburg (DTI).

Governor Markell welcomed the Open Data Council, explained to Council members how important open data will be, not just today, but in 10-20 years from now too. This Council will lay the foundation for many years to come.

Chair Lehman introduced the agenda.

Introductions.

* Ms. Fredel – Department of Health and Social Services – Interested in publishing data on restaurant inspections. An enterprising person could look at the most common errors that are found during inspections and make recommendations to restaurateurs.
* Mr. Yeatman – Department of Children, Youth, and their Families – interested in allowing data sharing to happen among agencies and with DHSS.
* Mr. McLeod – Department of State – 19 divisions, and many opportunities within the Department.
* Mr. Zimmerman – Department of Natural Resources and Environmental Control (DNREC) – the agency publishes a lot of data, and DNREC is upgrading their website and document management technology right now.
* Ms. Tweedie – Governor’s Office – interested in decreasing agency time and resources around accommodating FOIA requests because data is instead published on the Open Data Portal.
* Mr. Blowman – Department of Education – Personal Identifiable Information (PII) concern, and interested in what is considered public data and what is private data.
* Ms. Lin – Department of Transportation–DelDOT has a lot of data, so the challenge will be the volume of information.
* Mr. Stotler – Office of Management and Budget (OMB)– interested in what is public information versus what is not.

Secretary James Collins welcomed the Council. He explained that we are building on what has happened already, and that we have an opportunity to do work that hasn’t been done on an enterprise level before. We need to be consistent about our format and machine-readable and start with low hanging fruit. Cross pollination is another priority among divisions and agencies. Not every agency is a part of this Council, but we can invite others into this group.

Ms. Lehman talked about the vision of Open Data.

Open Data is data that is freely available for anyone to use, and you can do what you want with it (no permission, no copyrights). We will have limitations of liability and disclaimers on the data. Users of the Portal do not need to log in or identify themselves. Open data is machine readable in raw format, with metadata and data definitions. Open data has no cost to the user.

What’s not included? Open data does not include personal identifiable information (PII).

Why Open Data?

* Data sharing among agencies
* Help with FOIA requests, reducing staff time to fulfill requests
* Transactional Data Pulls
* Collaboration with Private Sector
* Public Good

Federal Government – data.gov and President Obama has issued two Executive Orders on Open Data. There is an opportunity to take information from Data.gov about Delaware.

Data Set Examples

* Restaurant Inspections – DHSS
* Where is my bus, GPS bus routes – DelDOT
* Environmental Sanctions – DNREC
* School profiles – Education
* Budget – OMB
* Licensed Child Care Center – Kids
* Licensed Professionals – State

Ms. Lehman presented various site designs of other open data portals within other states.

Mr. Blowman asked how many datasets are other states publishing.

Ms. Lehman said that Utah has 645 and Hawaii publishes a few thousand, for two examples.

Ms. Lin asked if we can find out which site design gets more traffic. Ms. Lehman said that might be hard to figure out since some states are publishing lots of data, where others are publishing less. But, making the data accessible is easy –API’s are built into the platform.

Mr. Stotler said the State already owns our data in our contracts, but we may need to amend contracts to reflect machine-readable formats. Need to look at Terms and Conditions in the business case process.

There was discussion about the site designs – the Council will need to make a decision about the Portal design, and should consider usability and user experience in that decision. The site designs will be presented at a future meeting, after the contract with the vendor is completed.

Mr. Stotler asked who is going to be responsible for fielding requests for datasets. If it is online interaction, then The Government Information Center (GIC) could field the requests. The agency themselves may also field such these requests, as they are doing currently.

Mr. McLeod asked would it be worthwhile to have a conversation with one of our counterparts in another state. Ms. Lehman agreed and said the Council will be doing a videoconference with the Open Data Chief from Connecticut.

Mr. Stotler asked how the Council should prioritize dataset publishing since the contract with the vendor will be for only 150 datasets. Mr. Zimmerman asked if the datasets should be what is already currently on agency websites or if it should be what is not yet on the website. Ms. Lehman said that we are leaving that to agencies, and that for the initial launch agencies should consider low hanging fruit that are already published.

Mr. Stotler said that Department of Finance is not named on the Council, and they have a lot of data such as checkbook.delaware.gov and credit card transactions. He asked if the Department of Finance datasets will be considered in the 150 number and does the Governor’s office have a priority on which datasets to make available. Ms. Tweedie said that the Governor’s office does not have a priority in mind, and that the priority is up to the council.

There was discussion about refreshing data – the example of the City of Los Angeles publishes their financial data and it is refreshed every single day. The Council discussed whether Delaware could consider something like this. Right now, much of the data coming out of a database/system will still need to have quality control to make it usable. But, should approval (legal/or publishing approval) still be needed if the data is just being refreshed? The Council will need to consider this.

Data definitions and metadata will be important to making sure that data is not misunderstood. Mr. Mahaffie suggested that the Open Data Portal should have a disclaimer that explains the Council/the State does not stand by someone else’s interpretation of the data- i.e., there is no warranty of the data presented beyond the Portal.

The Portal will be a centralized place to find data and users will not need to know which agency to go to. Ms. Lin said that there is a misnomer that the State is not currently publishing data, in fact, there is a lot of data already available online, and we should centralize datasets, but we should also consider work that has already been done.

Mr. Blowman asked who would be responsible for the content of the information in the datasets. Ms. Lehman said that the Council will have to work out a data approval process. There needs to be a review process and an approval process. NY has a data officer in each agency. Delaware may want a singular individual in each agency that will approve each dataset for publishing. NY has a stringent data approval process and CT less stringent. Discussion continued to if the Deputy Attorney Generals at each agency will need to approve each dataset. Ms. Lehman said that each agency does not currently go through their Deputy Attorney Generals for publishing information on their websites. Ms. Rohrbough said there are different approaches; NY involves a legal review and CT does not. CT is open and collaborative, and it depends on how the Council wants to approach this. Mr. Mahaffie said that there should be a high level review of whether the data should be public and published, and then when the data is refreshed it does not need a legal review again.

Mr. Stotler asked about if data is housed in multiple places, which is the dataset of record? Ms. Lehman said that the Open Data portal will host a copy of data, it will not be a data warehouse approach. There are ways to automate the refreshing of data, so multiple copies are not an issue.

Mr. Stotler said that agencies will have to be careful that various data sets cannot be triangulated to identify individuals. He further asked about data exchanges between agencies. Ms. Lehman said that at first, we should focus on public datasets, but data sharing in private view can be considered in the future.

Ms. Fredel gave an example of a public data mashup – there is a heroin and prescription drug epidemic in Delaware. Taking data related to overdose related deaths, state police drug arrest data, and health and treatment related data – this mashup would be of great benefit for the public. All three datasets would have personal information that would need to be scrubbed. This could be a test case for data mashups between agencies.

There was s discussion of refreshing data – keeping multiple copies of data could be problematic for agencies. But, the portal can set up automatic data pulls and refreshes. There was a concern that real-time data coming directly from a database could overwhelm the Open Data Portal’s bandwidth. Ms. Lehman said that we do not need to consider ramping up hardware at this point.

Mr. Blowman asked if Ms. Lehman or Ms. Rohrbough have done a scan of major agency websites for data set examples. Ms. Rohrbough said yes and that she will send out the list before the next meeting.

Mr. Zimmerman said that DNREC reports data to other organizations, such as the federal government, and sometimes third parties host data for their agency. Ms. Lehman explained that some of the federal data may be on data.gov and we could pull data about Delaware down from there.

The Council has the short-term goal of providing a list of initial datasets within 120 days, which is mid-May. Agencies are encouraged to consider more datasets rather than less, even though we only have purchased 150 initial datasets.

Ms. Lehman mentioned the Data Governance Council who is working on data sharing among criminal justice agencies. Executive Order 57 replaces the Data Governance Council with the Open Data Council. The Open Data Council will be above the Criminal Justice Subcommittee.

The next steps for the Council members are to consider initial dataset lists, look at data already on your agency’s website, and to talk with FOIA coordinators about the most requested datasets. The file format is anything machine readable – .csv, .xml, Excel, etc. PDF’s are not machine readable. There was a question about if the Portal will be accessible to those with disabilities. Ms. Rohrbough said she will find out.

## Adjournment

Meeting was adjourned at 3:05pm by Ms. Lehman, Chair. The next Council meeting will be at 1pm on April 4, 2016, in Wilmington.