

Open Data Advisory Group October 25, 2017 Meeting Minutes

Introduction

A meeting of the Open Data Advisory Group (ODAG) was held on October 25, 2017 from 6:30-8pm via teleconference. The following Members and County Staff Liaisons were in attendance:

Members

David Bain
Duke Banks
Christopher Hedrick
Jason Thomas (Chair)

County Staff Liaisons

Jack Belcher
Emily Hughes
Dharmesh Jain
Jaime Lees (Staff Liaison)
Katie Pegoraro
Girish Rajput
Sanmeet Sidhu

Administrative

The logistics and etiquette of teleconference calls were shared and any access issues resolved or worked around.

The Staff Liaison introduced a new County staff representative to the ODAG. Emily Hughes, Budget Analyst for the Dept. of Management and Finance, was appointed to the Advisory Group by the County Manager in response to a request for DMF representation in the Group.

Availability for the January meeting was shared. Attendees preferred January 3 6:30-8pm for an in-person meeting. The Staff Liaison will send an invitation out to the entire Group.

Open Data Portal Dataset Catalog

At the request of the Group, a [Dataset Catalog](#) was developed and made available on the Open Data Portal. The Dataset Catalog is a quick reference to all datasets on the Open Data Portal. It is automatically updated every evening and, like any other dataset on the Portal, can be readily sorted and exported. The Group reacted positively to the new feature.

Two enhancement suggestions were put forward:

- Make the URL links active so users can single-click to be taken directly to the dataset
- Allow users to sign up to be automatically notified when a new dataset is posted

County staff will research whether these requests can be accommodated. It was noted the Dataset Catalog is a good base on which additional information and tools could be built.

Prioritizing and Finalizing Initial Recommendations

The Group discussed several initial recommendations proposed in prior meetings. The Chair directed the Group to discuss and finalize these recommendations not just as a potential to-do list for the County but

as action items that could be taken on by Members, both in order to better inform the recommendation and to perform work that does not require County staff to complete.

- 1. Converting PDFs of data files on the public website to open datasets.**
- 2. Build an inventory of County data assets that could be published as open data**

The Group discussed recommendations 1 and 2 together.

The Chair asked for more background on the PDFs on the public website. Duke Banks clarified he was referring primarily to budget data which is presented by the County Manager in the form of responses to community member questions in the form of summary reports that include data tables and charts/graphs. These reports are pulled from a variety of systems, including the budget system, and typically are manually generated from pieces of multiple systems' reports. The recommendation was refined from simply presenting the data tables and source data for charts and graphs in the PDFs as open datasets to providing a contextualized version of the data that brings all the tables, charts and graphs together to help understand the reasoning behind financial decisions and recommendations. Some concerns were raised whether this type of recommendation, with which there was consensus it was of clear value, was in the scope of the ODAG. Duke Banks indicated he would provide background material from similar work done in Chile and tie that back to what he felt was needed in Arlington.

The Group wants to pursue a data inventory to facilitate the task of identifying and prioritizing datasets for publication. In addition, a data inventory will help the Group monitor the County's progress on publishing all open datasets over time. The Chair also suggested a related item that there be included in the open data policy that new system procurements include a requirement to evaluate the data collected for open data publication. Members noted that published datasets currently include some data inconsistencies, making it hard to know which dataset on which to rely. The Chair suggested we focus initially on what datasets we want to see and then dive deeper into data quality recommendations.

The Chair asked again what data does the Group wish to see, to think about that from the perspective in which you work or connect to government. The Group acknowledged it is hard to think about pressing questions without the data to help drive the questions. The Chair agreed and charged the Group with practicing both the development of the questions and identifying the broad data areas in which they would like to see datasets published. The Staff Liaison and Members offered to share examples of data analysis performed using open data and the Group agreed that would be helpful.

- 3. Evaluate and update the Open Data Portal for greater usability and self-guided exploration**

The Chair initiated discussion by noting the Group has identified several enhancements to the Open Data Portal in this and previous meetings. For these smaller tweaks, no recommendation to the County Manager is needed however if the Group recommends substantive updates/replacement of the Open Data Portal then this recommendation needs to be fleshed out. The Staff Liaison noted that most jurisdictions with an Open Data Portal use one of two main software platforms and both are substantially similar in functionality; the County uses the

simpler/less expensive one of the two. The Group felt the current platform meets current needs and this recommendation can be removed.

The Group did feel that better marketing of the Open Data Portal might yield more Open Dataset Requests and increased usage of the Open Data Portal by the community. A suggestion was made to publicize whenever a new dataset was posted to the Open Data Portal.

4. Standardize data definitions Countywide for greater usability of open datasets.

The Group discussed this recommendation briefly, enough to determine that it should remain on the recommendations list and be discussed more in the next meeting. Particular focus in the next discussion should be on evaluating the level of effort against the intended yield.

Action Items

1. Send meeting invitation for January 3 6:30-8pm. (Staff Liaison, by Nov 8)
2. Research enhancement suggestions for feasibility and provide update. (Staff Liaison/Sanmeet Sidhu, next meeting)
3. Provide background on budget data contextualization based on Chile example and how it ties to Arlington (Duke Banks, next meeting) – COMPLETE
4. Identify pressing questions and broad areas of data that are of interest to Members and community members with which Members associate. (All, next meeting)
5. Present examples of data analysis using open data (Staff Liaison and Christopher Hedrick, ongoing, beginning with next meeting)
6. Coordinate with Communications to see if dataset posting could be highlighted on the public website. (Staff Liaison, next meeting)