IDENTIFYING VALUABLE DATASETS FOR OPEN DATA

Introduction

This document is for the members of the Open Data Alliance to help with identifying valuable datasets for possible publication as open data. Included are general categories from the Washington State Office of the Chief Information Officer, popular categories identified by a study at the Sunlight Foundation, categories identified by our analysis of public records requests from the cities of Seattle, Port Orchard, and Olympia, as well as a set of recommendations to identifying valuable datasets.

Steps Needed for Identifying Valuable Datasets

Know Your Users – *Connect with the public*

Although connecting with the public may seem obvious and you might be doing so already, there may be other tactics to explore. However, one can only derive so much information on trends from public records requests. It is vital that as open data projects begin or continue, a connection and deep understanding of the needs of the public is clear and present when deciding datasets to choose for publication. Many citizens may not realize that their public records requests are information requests in truth, and it is these opportunities that provide space for possible valuable dataset identification. With such a community engagement tactic, as supported by resources in the Growth & Outreach document, any group or person will begin to identify the immediate needs of the public that could be addressed by specific datasets. As a side effect, there could also be the identification of a portion of the community that is not always addressed or viewed as being users of data.

Types of engagement: Focus Groups; Interviews; Surveys; Community Meetings; Public Forums

Know Other Users – *Look outside your area*

While knowing how to serve your community is the priority, there is no reason to "reinvent the wheel." Communities may have issues specific to their region and population, but often, the data most useful to them are similar across the board. Identifying open data projects or datasets that are successful or similar can provide opportunities for evidence that your project is worthwhile or even some inspiration for datasets. While in the search for other examples to help bolster the cause, it could also be an avenue for collaboration and resource sharing on a level that benefits both parties involved.

Areas to focus: Cities of similar sizes; Neighboring towns; Cities that lead Open Data Projects; Open Data Organizations

Know Your Data – *Understand the history*

There may be datasets already on the radar for possible publication that seem to be of no significance. In these situations, taking the mindset that each although it may not seem significant, someone out there (and there's always someone) may disagree and find a use for it. Such a situation is why the use of necessary metadata to provide context to the data is incredibly important for the continued use of open data portals and datasets. Giving metadata and a descriptive story about the datasets published will allow for a profound public interaction element.

Don't forget: Asking for complete metadata; Only publish if you can explain it; Make sure to always update

IDENTIFYING VALUABLE DATASETS FOR OPEN DATA

General Categories from WA OCIO[1]

- **Geographic reference data**: parcels, addresses (except category 3 and 4 data), where to obtain state services.
- **Public safety data**: traffic, moving violations, aggregate crime statistics, environmental hazards.
- **Fiscal data**: state salary data, expenditure data, budget data, purchasing data.
- **Health data**: quality and purity of water, food, and air, data that increases healthcare accountability, data that facilitates patient choice, aggregated incidence of diseases and medical conditions in communities (review for HIPAA compliance first).
- **Education data**: capabilities of state funded schools, achievements by the state's student population, results of state support of education.
- **Census and demographic data**: the populations of the communities of the state, trends in migration, diversity, and housing
- **Business and economic data**: aggregate changes in license counts, labor market and employment.
- **Government directory data**: who to contact for help with State services, catalogs of available state services.

The Sunlight Foundation[2]

Topics of the most popular open datasets

1		**Police and crime:** Police incidents; jail bookings; police station locations; crime statistics
2		**Transportation:** Taxi licenses; transit data; traffic counts; road infrastructure data; parking data
3		**Emergency calls:** Police, fire, and EMS responses; 911 calls; response times; incident reports
4		**Development:** Commercial developments; housing developments; property data; housing affordability
5		**Building safety:** Building permits; safety permits; certificates of occupancy
6		**Finance:** Revenue; spending; employee salaries; capital budgets; payments
7		**Elections:** Election results; polling locations; campaign finance reports
8		**Businesses and licenses:** Business licenses; liquor licenses; vendor, contract, and procurement data
9		**Inspections and service requests:** Restaurant health inspections; 311 requests; code violations
10		**Education:** Schools information; student health data; after-school programs; library locations

ODL Categories from Public Records Request Logs

Building and Construction

- Code Violations
  - Fire Code Violations
  - Environmental Assessments
- Property Records

Permitting

- Building Permits; City Project Permits

Police Department

- Police Officer Recordings (Audio/Video)
- Police Incident Reports
- Arrest Records

Transportation

- Automotive Accident Information