



Overview

The State Data Plan calls for state agencies to improve the availability and accessibility of open data. Open data is defined in the Plan as any data that (A) is freely available in convenient and modifiable format and can be retrieved, downloaded, indexed and searched; (B) is formatted in a manner that allows for automated machine processing; (C) does not have restrictions governing use; (D) is published with the finest possible level of detail that is practicable and permitted by law; and (E) is described in enough detail so users of the data have sufficient information to understand (i) the strengths, weaknesses, analytical limitations and security requirements of the data, and (ii) how to process such data.

By proactively making public data open and accessible, state agencies can improve access to data by other state agencies and the public, can enhance the public's understanding of state government, and can promote public participation and civic engagement. The Plan recommends that agencies identify high impact, high value data assets for publication as open data by engaging with external stakeholders and states that agencies should measure and track usage of open data.

Measuring open data activity can help agencies understand how their data is being used and what impact it is having; however, it can be difficult to quantify the value of open data to Connecticut state agencies and citizens. This document will provide guidance on how state agencies can measure open data use and quality, as well as how to assess the impact and value of the open datasets they publish. It draws from the [evaluation framework](#) designed by DataSF, using three categories of metrics: 1) publishing activity, 2) quality, and 3) use and impact.

Publishing activity

Section 4-67p of the Connecticut General Statutes, requires executive branch agencies to create open data access plans that detail their plan to enhance the availability of open data. The template for the open data access plans (available [here](#)) asks agencies to list the datasets that they are already publishing as open data in addition to the datasets that they plan to publish. Agencies must list the date each dataset will be made available and the frequency with which it will be updated. The open data access plans can serve as the benchmark against which to measure open data publishing activity by agencies. Below are four possible metrics to measure publishing activity:

1. **Number of executive branch agencies with open data access plans** – Does the agency have an open data access plan?
2. **Number of datasets published by agency** – How many datasets does the agency have published on data.ct.gov?
3. **Percentage of datasets listed in open data access plan that have been published** – What percentage of the datasets listed in the open data access plan have been published on data.ct.gov?
4. **Percentage of datasets with automated updates** – What percentage of datasets published are automatically updated?

Quality

Publishing high quality data is essential to promoting open data accessibility and use. In addition to measuring publishing activity and data use, agencies must also measure the quality of the data that they make available as open data, ensuring that it is well-documented and up-to-date. The metadata standard for data published on data.ct.gov (available [here](#)) provides guidance on metadata requirements. The open data access plans for each agency should include the target publication date and the data update interval. Below are three possible metrics for measuring data quality:

1. **Percentage of datasets published on time** – What percentage of datasets are published to data.ct.gov by the date indicated in the open data access plan?
2. **Percentage of datasets that are updated at the target interval** – What percentage of datasets are updated at the interval indicated in the open data access plan?
3. **Percentage of datasets with required metadata, as indicated in metadata standard** – What percentage of datasets are published with the required metadata listed in the metadata standard?

Use and impact

Understanding how open data is used and what impact it is having is the most challenging element of open data evaluation. Some metrics of open data use can be accessed through Socrata's analytical tools. For instance, data publishers can click the "View all statistics for this dataset" link on a dataset's landing page on data.ct.gov to find information about views, downloads, API calls, embeds, and referrers. More detailed information about data use on data.ct.gov can be accessed through [Socrata's Discovery API](#).

Measuring the impact and value of open data is more difficult than measuring data use. Agencies may want to consider creating a log of use cases that show their open data being used. Some use cases may be discovered by looking at Socrata's analytics for each dataset; if a dataset is embedded or referred to by another website, the embed and referral traffic should be listed on the dataset analytics page.

External impact and internal impact within the agency or within state government should both be considered, although more tools may have to be developed to measure this impact more comprehensively. Five possible metrics for measuring open data use and impact are below:

1. **Number of dataset views** – What is the total/monthly number of page views for each dataset?
2. **Number of dataset downloads** – How many times has each dataset been downloaded?
3. **Number of API hits** – How many API hits has each dataset received?
4. **Number of products (apps, tools, websites, etc.) made with open data** – How many products have been made with open data from each agency?
5. **Measurement of estimated time saved in responding to data requests** – How much time is saved by proactively making data available as open data rather than responding to data requests reactively?