

Fact Sheet: Open Data

DECEMBER | 2016

Version 2

Contact: Amye Osti
E-mail: amy@34north.com
Phone: 530-587-3434



A COLLABORATION BETWEEN THE CALIFORNIA ENVIRONMENTAL PROTECTION AND NATURAL RESOURCES AGENCIES | www.MyWaterQuality.ca.gov

Understanding Open Data and How to Promote It

In September 2015, the Delta Stewardship Council's white paper [Enhancing the Vision for Managing California's Environmental Information](#) offered a broadly shared vision for the advancement of environmental data sharing. Associated with the paper's recommendations are key concepts that we hope to explain further in a series of "fact sheets." These brief documents each address a different mechanism within a broader data-sharing strategy: [Data Management Plans](#), [Web Services](#), and [Data Federation](#).

We invite you to review each of these fact sheets, recognizing the range of subjects from fairly basic in sophistication (Data Management Plans) to moderately complicated (Web Services) to highly complex and coordinated (Data Federation). Working together, these technologies can help to modernize public data distribution. But our shared understanding of the underlying concepts, and of their respective value, is vital.

What is open data?

- The **Open Definition** makes precise the meaning of "open" with respect to knowledge, promoting a robust commons in which anyone may participate, and interoperability is maximized¹. Data is open if it can be easily and freely accessed and reused by anyone for any purpose.
- **Legally**, open data is available under an open data license, or no license at all, for access, use, reuse, and redistribution.
- **Technically**, open data is easy to access online in a machine readable format.
- Open data programs should aim to **preserve provenance**, tracking proper attribution, even as data are transformed.



What are the benefits of open data?

- Transparency and accountability
- Increased participation and collaboration among stakeholders
- Decreases in data duplication and unauthorized versioning while increasing data integrity
- New knowledge through the combination of data from different entities
- Increases in innovation and improved products from the private sector

Making data useful and making smart investments

- Creating open data can be a financially intensive and time-consuming process. Consistent communication with data consumers and other data collectors is crucial to the process.
 - Before creating open data sets, providers should convene stakeholders and request feedback about data structure, access points, and usability of the data.
 - Communicate with others collecting similar data to facilitate data interoperability.
 - Shortly after releasing the data, ask for feedback – e.g., "Are the data useful?" "Are the data in a format you can use?" "Do they expand on current research and data collection?"

- Timeliness is a critical measure of openness. Data that are withheld for too long might be considered by many to be functionally closed.
 - Accordingly, in key cases where current conditions are important to share, publishing the raw data in near real time is much more valuable than waiting for quality assurance.
 - Subsequent updates to a dataset, such as revisions produced by a quality control process, will appear immediately to your data consumers if a [web service](#) is used to distribute the data immediately.
- Data quality is also critical to document. When posting the data it is important to note if the data is raw or has been subject to quality control measures.

How to make data open and available?

- Select datasets or selections of datasets for open access.
 - Define goals and functionality associated with opening the data to the public. Include any caveats associated with the data you are posting.
 - Create a [Data Management Plan](#) (How will that dataset be updated? Who will be in charge? Where will the original data live?) Include this information and other metadata for users with the data file.
 - During the data selection process, look for any formatting inconsistencies, resolve formatting issues or make note of non-standardized data in the metadata.
- Make the data readily available and in a useful machine readable format (e.g. ftp site or csv file)
 - Consider creating [Web Services](#) for data consumers, as this will increase data consistency when data are updated or edited
- Ensure that the data are readily discoverable. Catalog all data in a central location; this enables interested parties to have easy access to use all data sets.
 - Establish a permanent location for the data access point to ensure that data remain accessible over time.
- Encourage the data consumers and other data collectors to offer feedback through public meetings, comment periods, and direct communication with known consumers.

AB 1755 and open data

- AB 1755 The Open and Transparent Water Data Act requires the California Department of Water Resources, in consultation with other key agencies, to “create, operate, and maintain a statewide integrated water data platform.” This platform will integrate and share data from multiple databases to make water data more accessible, interoperable, and useful.
- AB 1755 requires state agencies to develop protocols for data sharing, documentation, quality assurance and quality control (QAQC) and public access.

What are examples of other open data initiatives?

- California Open Data Portal: <https://data.ca.gov/>
- “National Renewable Energy Laboratory API Case Study”: <http://developer.nrel.gov/api-case-study/>
- Data.gov, “The home of the U.S. Government’s open data”: <https://www.data.gov/>
- “Open Data for Resilience Initiative”: <https://www.gfdrr.org/open-data-resilience-initiative-field-guide-0>

Open Data Resources

- The World Bank Open Data Program: <http://data.worldbank.org/>
- The Open Definition: <http://opendefinition.org/>
- The Open Data Commons: <http://opendatacommons.org/>
- The Open Data Handbook: <http://opendatahandbook.org/>

Fact Sheet Prepared for the Data Management Workgroup By:

Amye Osti, David Osti, and Karly Wagner, 34 North; Cristina Grosso and Tony Hale, San Francisco Estuary Institute

¹ <http://opendefinition.org/od/2.0/en/>