Open Data: An Introduction

About the Sunlight Foundation

The Sunlight Foundation is a nonpartisan, nonprofit based in Washington, DC. Founded in 2006, our mission is to use the power of the Internet to catalyze greater government openness and transparency. We do so by creating tools, open data, policy recommendations, journalism, and grant opportunities to dramatically expand access to vital government information to create accountability of our public officials.

Learn more at SunlightFoundation.com

Open Data = The Right to Information Online

We believe that information is power, or, to put it more finely, that disproportionate access to information is power. We are committed to improving access to government information by making it available online, redefining "public" information to mean "online" information. A key part of our approach relies on *proactive* disclosure, which means that the default for sharing information should be set to open, that information is put within reach of the public (online), and that the barriers to discovery, reuse, and consumption barely exist, if at all.

In practice, this means taking into consideration both technical formats for data publishing as well as the internal workflow and processes that contribute to whether or not government publishes data.

Key Resources and Tools

- **Sunlight Foundation's Open Data Policy Guidelines** - A series of best practice policy provisions, sample language, and use cases for opening up access to information online http://www.sunlightfoundation.com/opendataguidelines
- **Sunlight Foundation's Local Program Hub** - An evolving series of resources for transparency and open data advocacy at the state and municipal level http://www.sunlightfoundation.com/policy/local
- **Scout -** A tool for tracking state and federal legislation, regulations, and more

- **Open States** - A collection of tools for tracking state-level legislative information (including bills and information about legislators) http://openstates.org/

The 10 Principles of Open Data: *What Makes Data Truly Open?*

1. Completeness

Datasets released by the government should be as complete as possible, reflecting the entirety of what is recorded about a particular subject, including raw information and metadata. Doing so will permit users to understand the scope of information available and examine each data item at the greatest possible level of detail.

2. Primacy

Datasets released by the government should be primary source data. This includes the original information collected by the government, details on how the data was collected and the original source documents recording the collection of the data. Public dissemination will allow users to verify that information was collected properly and recorded accurately.

3. Timeliness

Datasets released by the government should be available to the public in a timely fashion with priority given to data whose utility is time sensitive. Real-time information updates would maximize the utility the public can obtain from this information.

4. Ease of Physical and Electronic Access

Datasets released by the government should as accessible as possible, with accessibility defined as the ease with which information can be obtained, whether through procedural, physical, or electronic means. Datasets should be accessible by unencumbered automated online access, in bulk, and through Application Programming Interfaces.

5. Machine readability

Information should be stored in widely used file formats that easily lend themselves to machine processing. (When other factors necessitate the use of difficult to parse formats, data should also be available in machine friendly formats.) These files should be accompanied by documentation related to the format and how to use it in relation to the data.

6. Nondiscrimination

Nondiscriminatory access to data means that any person can access the data at any time without having to identify him/herself (through registration, membership requirements, or other barriers) or provide any justification for doing so.

7. Use of Open Standards

Commonly owned (or "open") standards refers to who owns the format in which data is stored. Some formats are only accessible for use or review through proprietary processing programs. Utilizing freely available alternative formats makes the data available to a wider pool of potential users.

8. Licensing

The imposition of "Terms of Service," attribution requirements, copyright restrictions and so on acts as barriers to public use of data. Maximal openness includes clearly labeling public information as a work of the government and available without restrictions on use as part of the public domain.

9. Permanence

The capability of finding information over time is referred to as permanence. Information released by the government online should be sticky: It should be available online in archives in perpetuity. For best use by the public, information made available online should remain online, with appropriate version tracking and archiving over time.

10. Usage Costs

One of the greatest barriers to access to ostensibly publicly available information is the cost imposed on the public for access—even when the cost is de minimus. Most government information is collected for governmental purposes, and the existence of user fees has little to no effect on whether the government gathers the data in the first place. Imposing fees for access skews the pool of who is willing (or able) to access information. It also may preclude transformative uses of the data that in turn generates business growth and tax revenues.

Read and share the 10 principles online at
http://sunlightfoundation.com/policy/documents/ten-open-data-principles/

Open Data Hot Spots in the U.S.

U.S. Policies that have passed:

(Chronologically)

- [Washington, D.C.](#) (2006)
- [U.S. (Federal)](#) (2009, 2013)
- [Portland, OR](#) (2009)
- [Memphis, TN](#) (2009)
- [San Francisco, CA](#) (2009, 2010, 2013)
- [Lexington-Fayette County, KY](#) (2011)
- [Austin, TX](#) (2011)
- [Raleigh, NC](#) (2011)
- [New York City, NY](#) (2011)
- [Providence, RI](#) (2011)
- [Philadelphia, PA](#) (2012)
- [Montgomery County, MD](#) (2012)
- [Chicago, IL](#) (2012)
- [Madison, WI](#) (2012)
- [UT (State)](#) (2013)
- [NY (State)](#) (2013)
- [HI (State)](#) (2013)
- [Tulsa, OK](#) (2013)
- [NH (State) (2013)](#)

U.S. Policies currently in the works:

To be signed:

- [Illinois](#)
- [Nebraska](#) (study)
- [South Bend, IN](#)

Being heard:

- [Cambridge, MA](#)
- [Honolulu, HI](#)

Being drafted/early stages:

- Alabama (state)
- California (state)
- Florida (state)
- Maryland (state)
- North Carolina (state)
- Washington, D.C. revamp
- Arlington, VA
- [Oakland, CA](#)
- Houston, TX
- Kansas City, MO
- Louisville, KY
- Richmond, VA