

Open Government Data Standards and Setting Expectations

Josh Tauberer, GovTrack.us

Transparency Camp, Feb. 28, 2009

Open Data Standards

- Open Gov't W.G. “8 Principles”
- OMB Policies for Federal Public Websites
- ALA’s Key Principles of Gov’t Info.
- Sunlight Foundation’s Principles
- Open Knowledge Definition
- USACM Rec. on Open Government
- “Open” for Scholarly Work

Open Government W.G.

- 30 peeps gathered in Sebastopol, Calif. in November 2007
- Found “8 Principles” of open government data.
- www.opengovdata.org
- Not what *should* be open but what *makes* it open.

Open Government Data

1) Data Must Be Complete

2) Data Must Be Primary

Data are published as collected at the source, with the finest possible level of granularity, not in aggregate or modified forms.

3) Data Must Be Timely

Open Government Data

4) Data Must Be Accessible

Data must be made available on the Internet so as to accommodate the widest practical range of users and uses.

Disabled users

Multiple software platforms

Industry standard formats

Can be retrieved with automation

Open Government Data

5) Data Must Be Machine Processable

Tabular, normalized records with documentation.

6) Access Must Be Non-Discriminatory

including allowing anonymous access

Open Government Data

7) Data Formats Must Be Non-Proprietary

Proprietary formats add unnecessary restrictions over who can use the data, how it can be used and shared, and whether the data will be usable in the future.

8) Data Must Be License-Free

(cf. the purpose of the 8 Principles)

OMB Policies / webcontent.gov

- Policies December 2004; best practices continually updated
- http://www.usa.gov/webcontent/usability/accessibility/access_to_data.shtml

OMB Policies / webcontent.gov

- OMB Policies for Federal Public Websites **require agencies to...**
 - “Provide all data in an open, industry standard format permitting users to aggregate, disaggregate, or otherwise manipulate and analyze the data to meet their needs.” (#5D)
 - “establish and maintain communications with members of the public and with State and local governments to ensure your agency creates information dissemination products meeting their respective needs” (#4A)

OMB Policies / webcontent.gov

- **Webcontent.gov best practices:**
 - “New uses of your agency’s data may become a valuable public resource that would be out of the scope of your own website, such as helping to keep the public informed about the work of your agency and supporting civic education and participation.”
 - “Providing a uniform method to access raw data can also be the first step in internal development, accomplishing both goals at once. When a uniform method to access data is available, developers and web–services can focus on data presentation.”

OMB Policies / webcontent.gov

- “One benchmark for determining whether data is made sufficiently available is whether the public has all of the data needed to replicate any searching, sorting, and display functionality provided on the agency's own website.”

(see Princeton CITP's “Invisible Hand” paper.)

ALA's Key Principles of Govt Info

- What *should* the government be doing.
- 11 principles.
- Common ones:
 - Promote wide use, integrity, privacy, license-free
- <http://www.ala.org/ala/aboutala/offices/wo/woissues/governmentinfo/keyprins.cfm>

ALA's Key Principles of Govt Info

- 1) **Access to government information is a public right** that must not be restricted by administrative barriers, geography, ability to pay, or format.
- 2) The government has a **responsibility to collect, maintain, and disseminate** information to the public.
- 9) Government has an **obligation to preserve public information** from all eras of the country's history, regardless of form or format.

ALA's Key Principles of Govt Info

- 4) **Depository library programs** must be preserved to provide equitable, no-fee access to government information for the public.
- 5) The cost of collecting, collating, storing, disseminating, and providing for permanent public access to government information should be **supported by appropriation of public funds**.
- 6) ... Private sector involvement does not relieve the government of its information responsibilities.

Sunlight Foundation: Principles for Transparency in Govt

- February 2009
- http://issuu.com/johnwonderlich/docs/johnwonderlich_legislative2
- Covers both data issues and related expectations for making government more transparent through technology.

Sunlight Foundation: Principles for Transparency in Govt

- Transparency is a **responsibility** of government.
- “public” means “**freely accessible online**”
- Data quality & presentation:
 - Prompt, complete, accurate, and accessible
 - Searchable, manipulable, parse-able
 - Permanent and preservable

Sunlight Foundation: Principles for Transparency in Govt

- For the executive branch:
 - Create public database of **unique identifiers across jurisdictions** for ethics information.
 - Priority for info. related to **influence, corruption, and oversight**.
 - A high-level, centralized authority to coordinate policy is needed.
 - Citizen participation.

Sunlight Foundation: Principles for Transparency in Govt

- For the legislative branch:
 - Updating lobbying disclosure.
 - All required filings by lawmakers --- reports and documents --- should be done electronically and made public.
 - Legislative documents online before consideration.

Open Knowledge Definition

- What defines open knowledge? (2006)
- 11 factors
- The Open Knowledge Foundation
- opendefinition.org
- A little less restrictive than Open Gov Data.

Open Knowledge Definition

1) Access

Available at marginal reproduction cost. Convenient format.

2) Redistribution

Permits a license if it allows royalty-free redistribution.

3) Reuse

Permits a license if it allows modifications and derivatives to be shared.

Open Knowledge Definition

4) Absence of Technological Restriction

No DRM preventing access, redistribution, and reuse.

Use open data formats.

5/6) License may require proper attribution.

7/8) No discrimination against users or **uses**.

USACM Rec. on Open Govt

- Posted Feb 2009, after a year of discussion.
 - ACM=Association for Computing Machinery
 - These are the thoughts of the computer scientists and computing professionals.
- Like the 8 Principles, this is *what to do* with information that's already considered public.
- Common aspects:
 - Accessible, Complete
- <http://www.acm.org/public-policy/open-government>

USACM Rec. on Open Govt

- Data published by the government should be in **formats and approaches that promote analysis and reuse** of that data.
- Data republished ... should **preserve the machine-readability** of [existing] data.
- Citizens should be able to download complete datasets... [to use] standard methods such as queries via an API...

USACM Rec. on Open Govt

- Government bodies publishing data online should always seek to publish using data formats that **do not include executable content**.
- Published content should be **digitally signed** or include attestation of publication/creation date, **authenticity, and integrity**.

“Open” in Scholarly Work

- Budapest Open Access Initiative
 - Open Society Institute, Dec. 2001
- Science Commons Protocol for Implementing Open Access Data
 - “a method for ensuring that scientific databases can be legally integrated with one another” / Dec. 2007

Why it matters

- Government data is a **resource** that can make our world better:
 - NOAA weather data: Our daily forecasts! Ag. sector
 - SEC data: Fairly trading in public companies.
 - NASA's photos of Earth: inspiring!
 - Geospacial data: maps and directions!
 - Census/epidemiology data: we're healthier!

Setting Expectations

- Some recent things...
- Open House Project recommends **coordinating web standards** (May 2007)
- Obama calls for CTO during campaign (2008)
- Carl Malamud keeps setting data free.
 - SEC, judicial decisions, CFR

Implementation

- *Require* open access.
 - Enumerate data to be made open, or
 - Add a presumption of openness.
 - Public redress through legal system.
- *Review* open access (hat tip to J.W.)
 - Independent regular reviews of what data is open.
 - Review adds pressure to maintain standards.
 - Requires fewer policy changes.

ADDENDA

- Thoughts from the participants in the conference session:
 - Did an OMB circular in 1996 say agencies should not do things that compete with the private sector? Did USPTO face transparency hurdles w.r.t. this restriction?
 - Is there a requirement somewhere that agencies should use existing standards before creating their own?

ADDENDA

- One participant says: well-formed Unicode markup is all we need!
- Not everything is disability-accessible, like videos. Is it ok to trade accessibility for other achievements?
- Mandating data standards is too much because they can only be achieved in incremental steps.
- Should uses of government data have a (viral?) license that requires uses of the data be similarly open? Maybe a GPL for government.

ADDENDA

- Category Theory: Something for translating data between formats.