

Joshua Tauberer  
<http://razor.occams.info>  
[tauberer@govtrack.us](mailto:tauberer@govtrack.us)

Office of Science and Technology Policy  
Attn: Open Government Recommendations  
Executive Office of the President  
725 17th Street  
ATTN: Jim Wickliffe  
Washington, DC 20502

May 22, 2009

To whom it may concern,

I am writing in response to the call for public comments on the creation of an Open Government Directive following the President's Transparency and Open Government memorandum.

**By way of introduction**, I am the creator of [www.GovTrack.us](http://www.GovTrack.us), a website that helps the public track the status of pending legislation in Congress, a significant innovation on the Library of Congress's website THOMAS which still uses technology from the 1990s. GovTrack is a website I developed in my spare time while a student, and it is nonpartisan, noncommercial, and open source. Stemming from my work on GovTrack, I have been deeply involved in policy issues around the use of technology by the Congress in the interests of transparency. In 2007 I co-wrote The Open House Project report<sup>1</sup>, which was done at the encouragement of Speaker Nancy Pelosi under the guidance of Sunlight Foundation. Since then, I have worked with Congressional staff on directing the Library to issue a bulk-data-download of their legislative information (via the latest omnibus appropriations bill<sup>2</sup>) and on having the Senate adopt XML as a public data format for roll call vote records (a change made just a few weeks ago<sup>3</sup>). These are major components to advancing the use of technology to keep the public better informed about the activities of government.

I would like to take this opportunity to address several of the questions you raised from the perspective of **bulk raw data**. One of the primary concerns of an open government is to provide the public with government records. But because of the vast volume of records, a record is not meaningfully public unless citizens can use technology to sort, search, and transform records. As a result, the concern of government must not be merely to provide records, but to provide records in a way that supports technology's role in civic engagement and government oversight. This usually means providing the public with a "database".

There is ample evidence that **bulk raw data has a lot to give to society** for these purposes. The U.S. Congress makes its voting records available as a database to the public, and this database has been used by newspapers (The New York Times and The Washington Post) as well as start-ups (my own GovTrack.us, for instance) to aid the public in understanding what decisions were being made in the Congress. Websites like [www.iLive.at](http://www.iLive.at), a winner in the 2008 Apps for Democracy contest run by now CIO Vivek Kundra, then of the Washington, D.C. municipal government, can transform city operational data into useful tools, in this case providing a peek into neighborhood life for someone that might be moving to the area. It is not hard to find examples of where government data is put to good use.<sup>4</sup>

---

1 <http://www.theopenhouseproject.com>

2 <http://www.wired.com/threatlevel/2009/03/federal-bill-wo/>

<http://razor.occams.info/blog/2009/03/11/bulk-data-downloads-approved-in-the-omnibus-spending-bill-success/>

3 [http://www.cjr.org/campaign\\_desk/senate\\_goes\\_xml.php](http://www.cjr.org/campaign_desk/senate_goes_xml.php)

4 <http://www.miller-mccune.com/media/deep-throat-meets-data-mining-875>

<http://gcn.com/articles/2009/05/04/data-democratized.aspx>

I strongly recommend adopting a set of **baseline technological principles** that guide how the government shares data with the public. Issues to consider include how to make data available for the widest possible range of uses, how to ensure data will be available archivally, how to choose data formats appropriate for different use cases, and avoiding technology that gives commercial entities “veto power” to over free access to government records.

A number of groups have published recommendations along these lines.<sup>5</sup> The most notable is the *Eight Principles of Open Government Data* by an Open Government Working Group convened in 2007 (which included myself). I’ve summarized many of the recommendations already elsewhere<sup>6</sup> and came up with sixteen of the most important principles behind open government data. Of course, it would be impossible to satisfy every principle except by means of a magic wand, but they are goals to work toward incrementally. Briefly, they are:

- Information is not meaningfully public if it is not available on the Internet.
- The public is in the best position to determine what information technologies will be best suited for the applications the public intends to create for itself. Public input is therefore crucial to disseminating information in such a way that it has value.
- Data should be primary (i.e. original, disaggregate, and high-resolution) and made available in a timely fashion.
- Data should be made available in a non-discriminatory way and not subject to licensing restrictions.
- Data formats should promote analysis and reuse of the data. Data should use industry-standard machine-processable formats without executable content, and with digital signatures for provenance. Data should be made available in non-proprietary formats (possibly in addition to proprietary formats).
- The use of interagency standards is a plus, though standardization should not stand in the way of making data available.

Sometimes these recommendations appear to be esoteric or ideological. Don’t let the language fool you, however. These recommendations **relate directly to whether and to what degree government data is going to improve society**. Each technical recommendation has a particular civic goal in mind. For instance, executable content in data makes reading the data a security risk to users. Non-proprietary formats lessen the likelihood of data format obsolescence and ensure no one commercial interest has the power to cut off individuals’ access to government records. However, you will need to read any of the detailed

---

<http://www.sunlightfoundation.com/resources>  
<http://openregs.com>

5 American Library Association. *Key Principles of Open Government*.

<http://www.ala.org/ala/aboutala/offices/wo/woissues/governmentinfo/keyprins.cfm>

Association of Computing Machinery. *Recommendation on Open Government*. February 2009. <http://www.acm.org/public-policy/open-government>

Cornell University Library. *Digital Preservation Management*. <http://www.icpsr.umich.edu/dpm/dpm-eng/introduction.html>. Accessed April 4, 2009.

Open Government Working Group convened in November 2007 in Sebastopol, California, USA. *Eight Principles of Open Government Data*. They can be found at <http://www.opengovdata.org>.

Open Knowledge Foundation. Open Knowledge Definition. 2006. <http://www.opendefinition.org>.

Robinson, David G., Yu, Harlan, Zeller, William P. and Felten, Edward W. Government Data and the Invisible Hand. 2009. *Yale Journal of Law & Technology*, Vol. 11, p. 160. <http://ssrn.com/abstract=1138083>

Sunlight Foundation. Principles for Transparency in Government. February 2009. [http://issuu.com/johnwonderlich/docs/john-wonderlich\\_legislative2](http://issuu.com/johnwonderlich/docs/john-wonderlich_legislative2)

Webcontent.gov, U.S.A. government. Provide Appropriate Access to Data.

[http://www.usa.gov/webcontent/usability/accessibility/access\\_to\\_data.shtml](http://www.usa.gov/webcontent/usability/accessibility/access_to_data.shtml)

6 <http://razor.occams.info/pubdocs/opendataciviccapital.html>

recommendations cited above for more background on the individual provisions.

Certainly **some government agencies have been leaders in this area**. Environmental data has been one of the leading areas of open public data in the United States. The National Weather Service sets an example as the largest federal agency on the web in terms of data flow in most months. What is striking about NWS is not that they provide weather information to the public, which is of course a part of their mission, but that they do so in a way that promotes innovation. I spoke several months ago with Edward Johnson, the director of strategic planning and policy for NWS, who saw the NWS's role (in terms of data) as supporting the "health of the environmental information enterprise", meaning the wider community involved in weather information including commercial providers, the media, and the academic community. This is a model to apply elsewhere, that the government has a responsibility not just to plainly disseminate information but to support a picture larger than itself. Likewise, the Environmental Protection Agency's initiative to share data goes beyond its core missions of research, regulation, and education, and their recent initiative to post their calendars in a tech-savvy way is something to be drawn on for inspiration.<sup>7</sup> The Census, FEC, and SEC have also been leaders in the use of open government data. Nevertheless, there is always room for improvement.

As government websites turn more to use of **web services provided by the private sector**, such as federal agencies posting YouTube videos, care must be taken to ensure the web services comply with privacy, accessibility, archiving, and other standards required and expected of government websites. Private sector services almost always employ user tracking cookies and a Terms of Service agreement that governs how users may interact with the site and what users can do with the media they find (e.g. are users permitted to share it?). The use of these services by government websites may, if not attended to, require users to submit to user-tracking cookies and to enter into a contractual relationship<sup>8</sup> with these companies in order to view government-produced media. This would be unacceptable if it was the U.S. Government imposing these requirements, and it should be equally unacceptable for the U.S. Government to impose the same requirements for access through an intermediary. Baseline practices regarding government data should extend to data made available to the public through private sector intermediaries.

Finally, I would like to return to a point I noted earlier, that the public is in the best position to determine what information technologies will be best suited for the applications the public intends to create for itself. Despite the federal regulatory comment process, it seems quite rare indeed that the government would ask the public about its use of technology, and rarer yet that the feedback would occur in an open forum. This is a critical mistake since it ignores the role of technologists in the public in innovating on what is made possible through technology, and that **innovation comes best from free flowing ideas and open discussion**. Open discussion is quite possible. A case in point is the Open House Project Google Group<sup>9</sup>, run by Sunlight Foundation. This group, comprising Congressional staff, technologists in the public, and some staff for government agencies, has frequent serious discussion that often leads to tangible outcomes — notably some of the successes I listed at the start. Public discussion in an open and electronic setting works, and works well. And of course the model could easily be applied to any domain in which expert opinion would be of value. Agencies should be *required* to experiment with open forums, though I understand that laws such as FACA may need to be revised first and that there may be concerns if the forums are moderated. However, concerns do not mean it is not worth trying.

Thank you for your consideration.

Joshua Tauberer

---

<sup>7</sup> <http://www.epa.gov/adminweb/opa/workschedules.html>

<sup>8</sup> I should point out that Terms of Service generally allow themselves to be updated by the company without notice, making it impractical for users to both use the service and know what they have agreed to.

<sup>9</sup> <http://groups.google.com/group/openhouseproject>