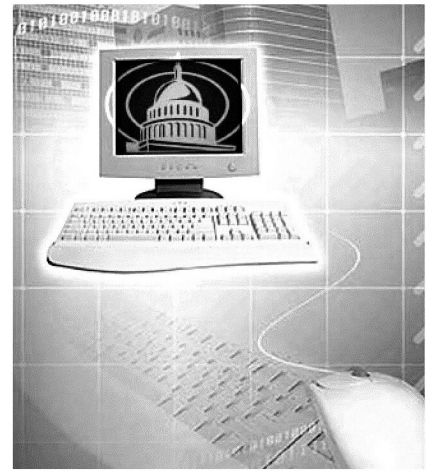


Improving Government Transparency Online



**Citizen-engaging
technology can make
government data
available online,
easy to access,
and understandable.**

by Jerry Brito

Democracy is founded on the principle that the moral authority of government derives from the consent of the governed. That consent is not particularly meaningful, however, unless it is informed. When government makes decisions in secret, opportunity for corruption increases and government's accountability to the people decreases. This is why we strive for transparency in government. When official meetings are open to citizens and the press, when government finances are open to public scrutiny, and when laws and the procedures for making them are open to discussion, the actions of government enjoy greater legitimacy.

Recent years have seen a renewed effort to increase government transparency in the United States. In the wake of the Jack Abramoff, Duke Cunningham, and William Jefferson scandals, Congress moved once again to shed light on its own activities. Senators Barack Obama and Tom Coburn introduced legislation requiring the full disclosure of all organizations receiving federal funds through an online database operated by the Office of Management and Budget. The result was the Federal Funding Accountability and Transparency Act of 2006. House Democrats, led by Speaker Nancy Pelosi, also pledged that after the 2006 congressional elections they would enact legislation to "restore accountability, honesty, and openness at all levels of government." The result was the Honest Leadership and Open Government Act of 2007, which requires

that information about earmarks be published on a public, searchable Web site forty-eight hours before a vote can be taken on the bill containing the earmarks.

Disclosure Requirements

Unfortunately, many statutory requirements for disclosure do not take Internet technology into account. For example, the 1978 Ethics in Government Act requires the disclosure of financial information—including the source, type, and amount of income—by many federal employees, elected officials, and candidates for office, including the president, vice president, and members of Congress. The act further requires that all filings be available to the public. One might imagine, then, that every representative or senator's information would be just a Web search away, but one would be wrong.

Members of the House of Representatives must file their disclosures with the clerk of the House, while senators must do the same with the secretary of the Senate. Each of these offices maintains a searchable electronic database of the filings. However, to access these databases, citizens must go to Washington, DC, and visit those Capitol Hill offices during business hours. No other means are available for searching the databases, greatly impeding widespread dissemination of nominally publicly available information.

Even when public information is available online, it is often not available in an easily accessible form. If data are difficult to search for and find, the effect can be the same as if it were not online. Also, to allow users to exploit the full potential of the Internet—to subscribe to data streams, to mix and match data sources—data must be presented in a structured machine-readable format.

For example, the Federal Communications Commission (FCC) is an independent government agency with an active regulatory agenda that it manages via its online docket system. In theory, users of the FCC Web site are able to see active rulemakings, search for and read FCC documents and public interest comments filed by interested parties, and file their own comments. In practice, the site seems to be an exercise in obscurantism.

The dockets containing proposed rules and other official FCC documents, as well as public comments, are

available on the Web site through a search form, but each docket contains neither an index of open proceedings nor indexes of documents. To obtain a list of documents in a given docket, the seeker must know the docket's number and search using that number. The resulting list is presented in chronological order with no way to sort by author, document length, or any other field.

There is also no way of searching within dockets for specific keywords. Even if there were a function that allowed one to search within documents, the results would be incomplete because many documents are posted as image files that are not easily parsed by computers and would not be returned in a search. This situation applies both to comments submitted by the public and FCC documents. This is the case even though public comments are usually created in word processing applications, such as Microsoft Word, which produce machine-readable electronic documents.

Really Simple Syndication

The most common form of subscribable structured data is an RSS feed. RSS stands for “really simple syndication” and usually refers to a family of data formats that allows the automation and aggregation of data. For example, the *New York Times* offers an RSS feed for its homepage, as does the *Washington Post*. A user can subscribe to these feeds with a desktop application called a “feed reader” or a Web-based reader such as Google Reader. Anytime something is added to the homepage of the newspaper, it is simultaneously published in the newspaper's RSS feed. When subscribers turn on their feed reader, it checks all the subscribed feeds for new items, which are then displayed. So, with one simple feed reader application, a user can keep track of dozens or hundreds of feeds without having to regularly visit the Web sites of the publisher, in this case the newspapers.

Imagine being able to subscribe to feeds from government agencies and departments. You could subscribe to the FCC's RSS feed and then never visit the site again just to check whether new regulations have been proposed. If a new regulation (or other document) is added, your reader automatically alerts you—but it could be even more useful.

The *New York Times*, for example, offers a feed just for its automotive section. Subscribe and you are notified only when new articles about cars are published, and you never have to wade through all the other published content. The *Times* also offers feeds for its food section, its section on European news, and dozens more.

Jerry Brito is a senior research fellow at the Mercatus Center at George Mason University. This article is adapted from “Hack, Mash, and Peer: Crowdsourcing Government Transparency,” forthcoming from the Columbia Science and Technology Law Review in 2008. He can be reached at gbrito@gmu.edu.

There is no reason why the FCC could not similarly publish a feed for each of its bureaus. Once users are aware of a regulation they would like to track, why not allow them to “subscribe” to the regulation? A government Web site could allow them to subscribe to regulatory dockets and be notified of all official actions and public interest comments filed in a particular docket, such as wireless spectrum regulations or cable regulations.

The *New York Times* also offers a series of “Times Topics” Web pages and companion RSS feeds. These range from persons (Rupert Murdoch, Hillary Clinton) to countries (Sudan, Colombia) to organizations, general subjects, and issues (New York Yankees, U.S. Supreme Court, cancer). Subscriber feed readers display articles published relating to a particular keyword anywhere in the pages of the *Times*. Imagine if such keyword subscriptions were available from regulatory agencies. The Environmental Protection Agency, for example, could offer topic subscriptions such as “pesticides,” “Superfund,” or “Vermont,” making it easier for citizens to engage in the topics that matter to them.

Government Information Online

The government cannot predict every possibly useful topic, but readily available technology today allows for RSS subscriptions to keyword searches. Google News, for example, allows users to make a regular Web search and then subscribe to the results. Each time a new item using your search term appears anywhere on the Web, you are alerted.

Making government information available online would not only benefit individual users of government Web sites, it would also make it simpler for third parties to aggregate government data. By aggregating data, Web sites can present government information in innovative and useful ways. For example, federal spending data gathered from a government Web site could be presented by a third party as an interactive map that shows the locations of funding recipients. Such applications make data exponentially more valuable. Government need not develop such innovative tools itself; as long as the data are available online in a structured format, private parties can make good use of the information.

Structured Data

“Structured data,” as we have noted, is a term of art, meaning that information is presented in a format that allows computers to easily parse and manipulate it. Al-

though a static Web page that lists a series of news stories or proposed regulations is not structured, it may have a companion XML file containing the same information. A structured XML file would allow a user to sort the data by ascending or descending date, alphabetically by headline or author, by number of words, and in many other ways that a static Web page does not afford.

In 2007, a group of interested citizens collaboratively produced a report detailing how the House of Representatives could use Internet technology to better serve its constituents. In it they explained,

❖❖❖ The notion of structured data is not new to the federal government. The Census Bureau, for instance, has for years not only provided a Web interface for census statistics—that is, a page where users can find simple data such as population numbers—but also the complete set of numeric data files to be downloaded and imported into database and statistics programs. The benefit of a download of the data is that with the complete data set computers can help people delve more deeply into the data and put it in new forms, such as charts and maps, that would be too time consuming to create by hand. Another example is the Securities and Exchange Commission’s practice of making investment filings available to the public in XML format through its EDGAR program. Likewise, the Federal Election Commission makes campaign contribution information available in a downloadable structured data format, allowing the public to absorb the information in a variety of ways. ❖❖❖

Mashups

When the government makes data available in a structured format, it opens the doors to innovative and enlightening remixes of information known as mashups. Mashups are tools that can potentially be used by journalists, bloggers, and citizens—the Internet’s intelligent crowds—to better scrutinize government’s activities.

Originating in the music world, the term mashup now also applies to applications that mix together disparate sets of data to create new and unique information. For example, the popular free-classified-ad site Craigslist.com is an almost definitive source for rental housing listings in urban areas. However, the site lists ads in the order that users add them to the site. This means that—using the Washington metropolitan area as an example—one listing could be for an apartment in the Adams Morgan neighborhood of the District and the very next ad for a house in Arlington, Virginia. This frustrated software engineer Paul Rademacher when he was looking for a place to live in Silicon Valley in 2004, so he

built HousingMaps.com, a mashup of the listings from Craigslist.com and Google Maps. This mashup allows users to bring up a map of the area in which they are interested (say five square blocks in a particular neighborhood) and pushpin icons will appear representing the properties available for rent in that area. Clicking on a pushpin brings up a bubble with the rental listing data, including rooms, price, location, photos, and a link to the actual listing.

What is amazing about a service like Housing Maps.com is that it is a new and unique information source that is richer and more useful than either Craig's List or Google Maps standing alone. What makes it possible is that Google chose to make its maps application interface open for anyone to use, and Craig's List chose to make its data freely available in an open and structured format. Their decisions to support openness and useful data formats allowed for an innovation that neither company could have predicted would emerge.

Book Burro, MAPLight, and More

When a site makes its data available in open formats, it cannot conceive of the many creative ways the data will be put to use. Book Burro, for example, is a plug-in for the popular Firefox Web browser that senses when you are looking at a page for a book (at Amazon.com, for example) and then fetches and displays data about the book's availability at local libraries, as well as the prices at competing online stores. Trendio uses open application interfaces from Yahoo, Google, and Technorati to index articles emanating from more than 3,000 news sources. It tracks the relative trends of words contained in those articles. The result is an index of trends in the media.

Mashups built on open interfaces and structured data represent a great potential fount of information about the workings of government. Their varied and unexpected outcomes can potentially make government activities transparent and reveal patterns now hidden in murky mountains of unstructured data. To get a sense of the possible, we can take a look at a leading transparency mashup called MAPLight.org.

The MAP in MAPLight.org stands for "money and politics," and the site's mission is to illuminate the connection between the two. Founded by computer expert Dan Newman, the site mashes together congressional voting data and campaign finance information. The result is a searchable database that highlights the connec-

tions between campaign contributions and how members of Congress vote.

Using the MAPLight database, users can look up a particular bill and see the interest groups, as well as the component individuals and corporations that support and oppose it. They can also see how members of Congress voted on bills and their contributors.

Crowdsourcing

If government data are successfully opened to public scrutiny online, seemingly impenetrable mountains of data will be made available. Mashups can help ease the information overload by highlighting the most interesting connections among data sets, but human judgment is still necessary to determine the most relevant facts. Crowdsourcing presents the key to sifting through the data made available by official disclosures, hacks, and mashups.

Although the United States has only 1,452 daily newspapers, about 70 million blogs are in operation, and about 120,000 new blogs come online each day. The vast majority of these blogs no doubt serve to inform and entertain a small circle of friends and relatives. Nevertheless, tens of thousands aspire to engage in journalism, and some have been successful. This affords a massive pool of ready and willing citizen journalists the likes of which traditional media has never assembled. This strength in numbers can allow the new technologies of transparency to be put to fruitful use despite the quantity of data available.

In his seminal essay, "The Cathedral and the Bazaar," Eric S. Raymond contrasts the open-source method of software development—in essence peer production or crowdsourcing—to the traditional hierarchical model. In the former, a large number of developers contribute simultaneously to the formulation and testing of software code, while central organization and a small number of developers typify the latter. He explains that one of the key differences is the number of eyes sifting through code looking for problems and solutions. He proposes what he calls Linus's Law: "Given enough eyeballs, all bugs are shallow."

Given enough eyeballs, corruption and waste are similarly shallow problems. In the cathedral view of journalism, corruption is hidden from a relatively small number of practitioners (reporters) by the inaccessibility and sheer volume of government data. In the bazaar view, a vast number of eyes, aided by hacks and mashups, make

the amount of data less daunting. The number of eyeballs comes not just from bloggers aiming to do journalism (although they are likely the most dedicated), but also from average citizens contributing to interactive sites.

Interactive Web Sites

Interactive Web sites have begun to leverage what James Surowiecki calls the “wisdom of the crowds” to shed light on government data. For example, WashingtonWatch.com gathers data on bills pending before Congress and mashes them with Congressional Budget Office estimates on the cost of each bill to present the average cost of bills per family or individual. Aside from presenting this information, the site allows users to contribute by registering their support for or opposition to bills and by posting comments about bills. More important, the site is also a wiki for pending legislation. Each bill’s page contains a detailed summary of the bill, its status, and points in favor and against, all of which can be edited or added to by anyone. Congresspedia.com is a similar community-written wiki, which also includes biographical pages for members of Congress.

Conclusion

To hold government accountable for its actions, citizens must know those actions. To that end, they must insist that government act openly and transparently to the greatest extent possible. In the twenty-first century, this entails making government data available online and easy to access. If government data are made available online in useful and flexible formats, citizens will be able to utilize modern Internet tools to shed light on government activities. Such tools include mashups, which highlight hidden connections between different data sets, and crowdsourcing, which makes light work of sifting through mountains of data by focusing thousands of eyes on a particular set of data.

Today, however, the state of government’s online offerings is paltry. Some nominally publicly available information is not online at all, and the data that are online are often not in useful formats. Government should release public information online in a structured, open, and searchable manner.



PMA'S NEW PARTNERSHIP WITH



The federal employees' choice for premier professional liability benefits. More info at www.fedsprotection.com or 866-955-FEDS.

PMA PROFESSIONAL MANAGERS ASSOCIATION

The Voice of Leaders in Public Service since 1980

BECOME A MEMBER AND...

- Share advocacy within Government agencies
- Provide input on Government-wide issues through coalitions
- Obtain a Member Benefits Card with instant savings (www.mymembercard.com/pma)
- Sign up for your PMA membership at www.ProManager.org

Contact us at 202-874-0126 | www.ProManager.org | info@ProManager.org