

No. 17
April 2008

MERCATUS ON POLICY

ENDING TRANSLUCENT GOVERNMENT: Putting Government Data Online

By Jerry Brito

MERCATUS CENTER
GEORGE MASON UNIVERSITY

The federal government makes an overwhelming amount of data publicly available each year. Laws ranging from the Administrative Procedure Act to the Paperwork Reduction Act require these disclosures in the name of transparency and accountability.¹ However, the data is often only nominally available to the public. First, much government data is not available online or even in electronic format. Second, the data that can be found online is often not available in an easily accessible or searchable format. If the government made information public online in standard open formats, the millions of Americans online could help ensure the transparency and accountability that is the reason for making information public in the first place.

THE HISTORY OF PUBLICLY AVAILABLE GOVERNMENT INFORMATION

LAWS ENCOURAGING GOVERNMENT transparency and accountability have been a feature of the American system of government since the founding of the Republic. The Constitution, for example, requires that each house of Congress “keep a Journal of its Proceedings, and from time to time publish the same, excepting such Parts as may in their Judgment require Secrecy.”² Today, the Congressional Record satisfies this requirement.

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 has moved again to shed light on its own activities.³ In 2006, Senators Barack Obama and Tom Coburn introduced legislation requiring the full disclosure of all organizations receiving federal funds through an online database to

be operated by the Office of Management and Budget (OMB).⁴ The result was the Federal Funding Accountability and Transparency Act of 2006.⁵ Additionally, House Democrats, led by Speaker Nancy Pelosi, pledged that after the 2006 congressional elections they would enact legislation to “restore accountability, honesty, and openness at all levels of government.”⁶ The result was the Honest Leadership and Open Government Act of 2007, which requires that information about earmarks be published on a public, searchable website 48 hours before a vote can be taken on the bill containing the earmarks.⁷

There are no other official means of searching the databases, something that presents a major barrier to widespread dissemination of nominally publicly available information.

PUBLIC GOVERNMENT DATA IS OFTEN NOT ONLINE

UNFORTUNATELY, MANY OF the 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 and vice president, and members of Congress.⁸ The Act further requires that all filings be available to the public, subject to certain limited exceptions.⁹ You might imagine, then, that every representative’s or senator’s information would be just a web search away, but you would be wrong.

The House and Senate maintain searchable electronic databases of its members’ filings.¹⁰ However, to access these databases, citizens must go to Washington, D.C., and visit Capitol Hill records offices during business hours.¹¹ There are no other official means of searching the databases, something that presents a major barrier to widespread dissemination of nominally publicly available information.

NOT EVEN GOOGLE CAN HELP: THE DIFFICULTIES OF USING ONLINE PUBLIC GOVERNMENT DATA

EVEN WHEN PUBLIC information is available online, it is often not available in an easily accessible form. If data is difficult to search for and find, the effect might be the same as if it were

not online. Additionally, for users to exploit the full potential of the Internet—subscribing to data streams and mixing and matching 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.¹² In theory, users of the FCC website can see active rulemakings, search for and read FCC documents and comments filed by interested parties, and file their own comments. In practice, the site seems to be an exercise in obscurantism.

To obtain a listing of documents in a given docket, you must know the docket’s number and search using that number.¹³ There is no way of searching within dockets for specific keywords. This is fine for an industry lobbyist with expertise and resources, but a mom in California wanting to look up comments filed in an inquiry into children’s television would have a difficult time finding the information. Moreover, even if one could run a keyword search, many documents are posted as image files.¹⁴ Computers cannot easily parse these files, so a keyword search would not return any information from these files. Not even Google can help: the FCC blocks the search engine from its document repository.¹⁵

THE PROMISE OF STRUCTURED DATA

MOST GOVERNMENT SITES do not offer access to their data in a structured format. What does this mean? 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 allow the automation and aggregation of data.¹⁶ For example, the *New York Times* offers an RSS feed for its homepage. A user can subscribe to these feeds with an application called a “feed reader.” Any time something is added to the home page of the newspaper, it is simultaneously published in that newspaper’s RSS feed. When a subscriber turns on his feed reader, it checks all the subscribed feeds for new items and then displays them. So, with one simple feed reader application, a user can keep track of dozens or hundreds of feeds without having to regularly visit the websites of the publishers.

Imagine being able to subscribe to feeds from official government websites. If you were subscribed to the FCC’s RSS feed and the FCC added a new regulation, your reader would alert you automatically.¹⁷ But the RSS feed could be even more useful. Just as the *New York Times* publishes a feed for its automotive section so those readers interested only in cars don’t have to wade through the rest of the paper, so the FCC could publish a feed for each of its bureaus. People interested in just wireless spectrum regulations or cable regulations could subscribe to those feeds only and not worry about indecency enforcement or homeland security issues.

Once users are aware of regulations they would like to track, why should they not be able to subscribe to those regulations? A government website should allow users to subscribe to regulatory dockets and be notified of all official actions and public interest comments filed in a particular docket.

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, Supreme Court, cancer). If you were to subscribe to the RSS feed for one of these keywords, your feed reader would display all articles published anywhere in the pages of the *Times* that relate to that keyword. Imagine if such keyword subscriptions were available from regulatory agencies. The EPA, 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.

Finally, even if the government cannot predict every possibly useful topic, readily available technology today allows for RSS subscriptions to keyword searches. Google News, for example, allows users to make a regular web searches and then to subscribe to the results. Each time a new item using the search terms appears anywhere on the web, Google alerts the subscribers.

CONCLUSION

THE FIRST BUILDING block of a foundation on which Internet technologies can help improve transparency is the idea that government should make its data public to the greatest extent feasible. While technically government makes data available to the public, practically speaking the data is out of reach. Government needs to make data publicly available in a meaningful sense, which in this day and age means it needs to put that data online.

There is no excuse for not making data available online. Almost all data today is created electronically using word processors and other computer applications. Because documents enter the world digitally, the initial step of online publication (i.e., digital formatting) is complete. The next steps—including designing and implementing useful websites to host the data—should be a minimal cost since most agencies already have online presences. The rest of the world knows that the electronic dissemination of data presents efficiencies and savings over paper. Government should know it too.

The second building block needed for a solid foundation of government data is the idea that information should not just be made available online, but that online resources must also be useful. This means putting data online in structured, open, and searchable formats.

For the good of the country, citizens should encourage the federal government to release public information online in a structured, open, and searchable manner. If the government would make data available online in useful and flexible formats, citizens would be able to use modern Internet tools to perform their duty to hold government accountable for its action. In turn, the government would begin to fulfill its responsibility of acting as openly and transparently as possible.

A government website should allow users to subscribe to regulatory dockets and be notified of all official actions and public interest comments filed in a particular docket.

FOOTNOTES

1. Administrative Procedure Act §§ 3-4, 5 U.S.C. §§ 552-53 (2006); Paperwork Reduction Act, 44 U.S.C. §§ 3501-3520 (2000 & Supp. IV 2004).
2. U.S. Constitution, Article I, § 5, cl. 3.
3. See generally Susan Schmidt & James V. Grimaldi, “Abramoff Pleads Guilty to 3 Counts,” *Washington Post*, Jan. 4, 2006, A1; Charles R. Babcock & Jonathan Weisman, “Congressman Admits Taking Bribes, Resigns,” *Washington Post*, Nov. 29, 2005, A1; Jerry Markon & Allan Lengel, “Lawmaker Indicted on Corruption Charges,” *Washington Post*, June 5, 2007, A1.
4. Press Release, Sen. Barack Obama (D-Ill.), U.S. Senate, “Obama, Coburn Introduce Bill Requiring Public Disclosure of All Recipients of Federal Funding” (Apr. 7, 2006), http://obama.senate.gov/press/060407-coburn_introduc/.
5. Federal Funding Accountability and Transparency Act of 2006, Pub. L. No. 109-282, 120 Stat. 1186.
6. Office of the House Democratic Leader Nancy Pelosi, 109th Congress, *A New Direction for America* 21 (2006), <http://www.speaker.gov/pdf/thebook.pdf>.
7. Honest Leadership and Open Government Act of 2007, Pub. L. No. 110-28, § 521, 121 Stat. 735, 760-64.
8. Ethics in Government Act of 1978 §§ 101-02, 5 U.S.C. app. 4 §§ 101-02 (2006).
9. 5 U.S.C. app. 4 § 105 (2006).
10. The Open House Project, *Sunlight Found, Congressional Information & the Internet: A Collaborative Examination of the House of Representatives and Internet Technology* 45 (2007) [hereinafter Open House Report], http://www.theopenhouseproject.com/report/open-houseproject_may8_07.pdf.
11. *Ibid.*; see also Rob Bluey, “Why Aren’t These Documents Available Online?” The Open House Project, Mar. 7, 2007, <http://www.theopenhouseproject.com/2007/03/07/why-arent-these-documents->

available-online (describing the process of accessing documents at the House Legislative Resource Center).

12. Federal Communications Commission, FCC Electronic Comment Filing System, <http://www.fcc.gov/cgb/ecfs> (last visited Oct. 10, 2007).
13. Federal Communications Commission, Electronic Comment Filing System [Enter Search Criteria], http://fjallfoss.fcc.gov/prod/ecfs/comsrch_v2.cgi (last visited Oct. 10, 2007). A note at the bottom of this page states that it was last updated on Dec. 11, 2003. *Id.*
14. One of the main document types used by the FCC is Portable Document Format (PDF). PDFs can contain digital text that is subject to search (usually created by saving as a PDF document from a word processing application) or images of text that cannot be searched (usually created by simply scanning a printed document). See Adobe Systems Inc., Adobe Reader 7.0 for Windows and Macintosh 166 (2004), available at <http://www.adobe.com/products/acrobat/pdfs/acrruserguide.pdf> ("PDF documents that are created by scanning a printed page are inherently inaccessible because the document is an image, not text that can be tagged into a logical document structure or reading order."). The Department of Justice (DOJ) has taken note of this problem as it relates to the accessibility of websites for disabled users who rely on devices that depend on machine-readable text. In a 2004 report, the DOJ stated:

A more significant problem involves agencies' use of inaccessible content on their sites. An agency may create a Web page that is easily navigated by people using a text-only browser but then include downloadable files that are inherently inaccessible. This problem occurs most frequently with two types of file content used by many components—files rendered by scanning to Adobe Acrobat's portable document format (pdf) and multimedia files.

Department of Justice, Section 508 of the Rehabilitation Act: Accessibility for People with Disabilities in the Information Age, <http://www.usdoj.gov/crt/508/report/web.htm> (last visited Oct. 10, 2007).

15. Declan McCullagh, "Feds use robots.txt files to stay invisible online. Lame.", *C-Net News.com* (Aug. 24, 2007), http://www.news.com/8301-13578_3-9765451-38.html.
16. Mark Pilgrim, "What is RSS?", *O'Reilly XML.com* (Dec. 18, 2002), <http://www.xml.com/pub/a/2002/12/18/dive-into-xml.html>.
17. There are scattered uses of RSS in government. The SEC for example provides RSS feeds for financial disclosure data that firms voluntarily submit in a structured format. See SEC XBRL RSS Feed Files, <http://www.sec.gov/info/edgar/ednews/xbrlrss.htm> (last visited Oct. 10, 2007). The Copyright Office uses RSS extensively for its publications, including its Federal Register notices. See Copyright Office Federal Register Notices, <http://www.copyright.gov/fedreg> (last visited Oct. 10, 2007).

The Mercatus Center at George Mason University is a research, education, and outreach organization that works with scholars, policy experts, and government officials to connect academic learning and real world practice.

The mission of Mercatus is to promote sound interdisciplinary research and application in the humane sciences that integrates theory and practice to produce solutions that advance in a sustainable way a free, prosperous, and civil society.

Jerry Brito is a senior research fellow with the Regulatory Studies Program at the Mercatus Center at George Mason University. His research interests include regulation, telecommunications policy, and government transparency.