

THE WIKI STATE: *Government Innovation Using New Online Tools*

GARY CHAPMAN

IN DECEMBER 2007, THE U.S. FEDERAL GOVERNMENT launched a new Web site that is a harbinger of things to come. The USASpending Web site, at <http://www.usaspending.gov>, provides a searchable database of federal contracts and awards, beginning with fiscal year 2007. The database can be searched using a variety of criteria, including the name of the contractor, the amount of the award, the location of the contractor or awardee, etc. Online access to this information was required by the Federal Funding Accountability and Transparency Act of 2006, which was co-sponsored by Senators Tom Coburn (R-OK) and Barack Obama (D-IL), and signed into law by President George W. Bush on September 26, 2006. The Transparency Act requires the federal Office of Management and Budget (OMB) to offer the searchable Web-accessible database to the public at no charge.

When this task landed in the lap of the OMB in late 2006, a nonprofit “watchdog” group called OMB Watch, which was founded in 1983, was already doing something similar to what was required in the Transparency Act. OMB Watch ran a Web site called FedSpending.org (<http://www.fedspending.org>), which helped citizens analyze and track federal spending. OMB Watch executive director Gary Bass offered to help OMB associate director Robert Shea with the technology. Shea was at first reluctant because of the antagonistic relationship between OMB and its watchdog shadow. But, as he told *The Washington Post* last December, “OMB Watch had already proven it could be done, so why do it from scratch?”¹

For his part, Bass said, “Normally, we come to bury Caesar, not praise him. But they are doing something that’s very cool, that’s very innovative in government.”²

Out of this collaboration came a new gold standard for getting government information online. The data available through the USASpending Web site is not

only searchable on the site itself, but, more importantly, all of the data can be downloaded. The site also features published, open information about the application programming interfaces, or APIs, that software developers need to build their own applications with the federal data. Thus, OMB Watch's site, fedspending.gov, uses the same data and APIs, but offers more features, such as graphs and maps of the spending data. It is also an accountability check on the OMB site.

WEB 2.0 AND OPEN GOVERNMENT

Coincidentally, in December 2007, the same month that OMB's new spending Web site launched, a group of about 30 technologists and open government activists met in Sebastopol, California, at the meeting place of O'Reilly Media, Inc., a well-known publisher of technical manuals and other technology-related books. Tim O'Reilly hosted a conversation about how to promote the idea of open government through technology. One result was a consensus document on "Open Government Data Principles,"³ which describes the following eight such principles:

1. *Complete: All public data is made available. Public data is data that is not subject to valid privacy, security, or privilege limitations.*
2. *Primary: Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.*
3. *Timely: Data is made available as quickly as necessary to preserve the value of the data.*
4. *Accessible: Data is available to the widest range of users for the widest range of purposes.*
5. *Machine processable: Data is reasonably structured to allow automated processing.*
6. *Non-discriminatory: Data is available to anyone, with no requirement of registration.*
7. *Non-proprietary: Data is available in a format over which no entity has exclusive control.*
8. *License-free: Data is not subject to any copyright, patent, trademark, or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.*

The combination of these principles, in the context of new online tools and techniques, opens up the potential for a new era of transparency, accountability, and freedom of information using government sources of data. The most interesting innovations are built on open data standards and published APIs, which together help to satisfy principle number 5, that data be "machine processable."

It was also Tim O'Reilly who was widely credited with describing what is meant by the term "Web 2.0," again at one of O'Reilly's by-now rather famous

conversations in Sebastopol. In his widely read 2005 essay “What is Web 2.0?”, O’Reilly laid out the fundamentals of a paradigm shift in the way we think about the World Wide Web.

At bottom, Google requires a competency that Netscape never needed: database management. Google isn’t just a collection of software tools, it’s a specialized database. Without the data, the tools are useless; without the software, the data is unmanageable. Software licensing and control over APIs—the lever of power in the previous era—is irrelevant because the software never need be distributed but only performed, and also because without the ability to collect and manage the data, the software is of little use. In fact, the value of the software is proportional to the scale and dynamism of the data it helps to manage.⁴

What this means is that “data is the next ‘Intel inside’,” as O’Reilly put it—the value of new online platforms like Facebook, YouTube, MySpace, or even Google, is proportional to how much data they manage, and that data is often supplied by users of the software. The popular classified ads and messaging platform Craigslist is a simple software tool that is valuable precisely because so many people use it. Unstructured data is not useful—users need to find what they want to know or experience, which is why techniques of search and filter have become so central to our lives and to businesses. But tools that manage data are irrelevant if they don’t access lots of data—the more data is accessible, the more users are attracted to it, for a multiplicity of reasons and purposes.

The next step was obvious: making the same data accessible through different software tools. One of the earliest and most dramatic examples was a Web site called housingmaps.com, created in 2005 by Paul Rademacher, which displayed apartment and house rental ads from Craigslist on Google Maps. This allowed people to use a map tool on the Web to see a geographic display of a local rental market. Google didn’t do this, and neither did Craigslist—it was Rademacher who figured out how to “mash” the data together with the tool. (Google was impressed enough to hire him.)⁵ This has led to an explosion of online “mashups,” a new term in the online world, which refers to a successful blending of a source of data with a tool or software application that provides new ways of seeing or managing the data. (An entire Web site, GoogleMapsMania, is dedicated to tracking innovative mashups using Google Maps.)

Other examples include a site that shows YouTube music videos of the top 10 iTunes songs, or a three-dimensional globe that displays Yahoo News headlines on their story’s geographic location. After Hurricane Katrina, Google Maps mashups were prolific, including a map that showed daily FEMA-measured water levels in New Orleans neighborhoods. During and after the devastating southern California fires in late 2007, local officials enlisted technical volunteers to develop map mashups displaying fire zones, shelters, evacuation routes, firefighter assembly points, etc. Volunteers used Amazon.com’s “Mechanical Turk” service to search for missing computer scientist Jim Gray, who mysteriously disappeared in his boat off the coast of northern California in January 2007.⁶

The building blocks of mashups and other innovative ways of using data online are open data standards and published, open APIs. Data stored in an open

format can be accessed by any program that understands the format, or that can be programmed to do so, and open formats are usually non-proprietary, which means that they are not attached to royalty payments or other intellectual property restrictions. Plain text is the most open and reliable format, and text can be “coded” into other open formats for online use, such as Hypertext Markup Language (HTML) and Extensible Markup Language (XML). Free and unlimited data exchange is largely dependent on the use of open data standards, and the Internet has made this much more commonplace than it was before the era of networked computing.

Application Programming Interfaces are simply a formal way of describing how to interact with a software program using software code. Microsoft, for example, makes available its Windows APIs so that programmers can write programs that run on Windows. Google makes available its Google Maps APIs so that programmers can create mashups that make geo-coded data visible on Google Maps. OMB has published the APIs for the federal spending database—available on the USASpending Web site—so that programmers can determine the structure of the data in the spending database. Using these APIs, programmers can write new programs analyzing federal spending.

Combining open data with open APIs provides the key to sharing and using data in innovative new ways. And this is a new vista for government openness, civic engagement, freedom of information, and public accountability.

WHY THIS IS TIMELY NOW

There has been a growing trend in the public sector to seriously consider migrating from proprietary data standards to open data standards for a variety of reasons.

Government information archivists, for example, have become increasingly concerned about archiving information in a proprietary format that may not be supported far into the future, meaning that the program used to create the information will be long gone, forgotten, unsupported, or otherwise unavailable. This has already been demonstrated by the past use of word processing software that is no longer used or by incompatibilities between versions of Microsoft Word.

Largely because of the Internet, most government agencies are trying to get their data online, and also trying to combine data from incompatible databases that were, in many cases, deployed before the Internet boom. Moving beyond the model of isolated data silos is a major priority for most public sector information officers. An example is the recent \$863 million outsourcing contract that will put IBM in charge of enterprise services—such as e-mail, Web services, and instant messaging—for 27 Texas state agencies. A big part of the motivation for this large contract was to move Texas state enterprise applications to open data standards.⁷

Public sector IT managers are also beginning to notice some advantages in using open source software and open source tools online, such as blogs, wikis, content management software, and platforms such as Google Maps. The City of

Austin, Texas, for example, is building a new Web presence using the open source content management system known as Plone,⁸ which, in addition to being free, is built on open data standards that will support a new level of transparency in city government. For this project, the City of Austin developed a new slogan, “Austin Go,” with a logo that includes the phrase “A new era of open government.”⁹

In his 2007 book *The New American Story*, former New Jersey Senator Bill Bradley wrote:

. . . [T]he entire federal budget should be on the Internet, with keyword accessibility. For example, if you searched for “breast cancer” or “housing,” you would be directed to all the places in the federal budget where money is spent for those purposes. That way, citizens could have the information with which to understand the trade-offs in taxes and spending and hold their their legislators accountable.¹⁰

An additional feature of an “interactive federal budget” would be to make all the budget data available, as well as published, non-proprietary APIs, so that volunteers, news organizations, academics, and others interested in the budget could develop programs that provide new ways of looking at the federal budget process. The Web site on federal spending mandated by the Transparency Act sets the standard that should be extended to the budget, before the money is spent. Senator Barack Obama, in his presidential campaign, has said that the federal spending site his bill created was meant to be a “Google for government.” It’s a step in that direction; now the task is to extend that kind of transparency and accessibility to all government information and thereby set an example for the rest of the world.

What we have before us is the next step in the unfolding of democracy in the United States, with an approach that should have bipartisan appeal. Given ongoing concerns about the influence of money and special interests in our political system, about the fiscal challenges of the federal government, and about freedom of information in the Internet era, improving transparency and accountability using the tools at hand—the Internet, open data standards, and open APIs—should be one of the chief goals of the next decade. As Bill Bradley wrote in *The New American Story*, “Without idealism, government itself seems like just another special interest. But without accountability, idealism becomes softheaded talk.”¹¹

The United States, the oldest and most influential democracy in the world, should take the lead in helping citizens engage with their governments, all over the world. The U.S. government can set the standard for openness and accountability, as well as civic engagement, and demonstrate a new set of tools that can be used inexpensively and effectively by other governments and non-governmental organizations around the globe. This is the next stage in the development of the Internet as perhaps the most useful technology of the modern age, as well as the next phase in the millennial-long story of democracy.

NOTES

1. "OMB Offers an Easy Way to Follow the Money," by Elizabeth Williamson, *The Washington Post*, December 13, 2007; p. A33.
2. Ibid.
3. "Open Government Data Principles," at <http://wiki.opengovdata.org/index.php/OpenDataPrinciples>, accessed February 22, 2008.
4. "What is Web 2.0?" at <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, accessed February 22, 2008.
5. "Paul Rademacher: The Main Who Opened Up the Map," *Technology Review* magazine, at <http://www.technologyreview.com/tr35/Profile.aspx?Cand=T&TRID=437>, accessed February 22, 2008.
6. "Help Find Jim Gray," at http://www.allthingsdistributed.com/2007/02/help_find_jim_gray.html, accessed February 22, 2008.
7. "IBM wins huge Texas outsourcing deal," *Computer Business Review*, November 29, 2006, at http://www.cbonline.com/article_news.asp?guid=7BBE6B56-BD28-4D7A-A0F0-B68533E025F4, accessed February 22, 2008.
8. Plone at <http://www.plone.org>.
9. Austin Go at <http://www.ci.austin.tx.us/austingo/default.htm>.
10. Bradley, Bill, *The New American Story*, New York: Random House Trade Paperback Edition, 2007, p. 84.
11. Ibid.