Transparency Camp
April 30, 2011
Joshua Tauberer

Principles and A Brief Legal History of Open Government Data

An Introduction

If you watch some streaming video from the U.S. House of Representatives, you are first greeted with a note not unlike the scary FBI warnings that used to start off all VHS tapes. "The use of duplications of broadcast coverage of the Committee on Transportation is governed by the rules of the House. Use for political or commercial purposes is expressly prohibited." Political speech is the most sacred of all speech in the United States. Could it be true that we are actually prohibited from using certain government records in political speech? I don't think this message was ever intended for citizens --- I think it was originally intended to prohibit Congressmen themselves and the media who are registered in the press galleries from using the videos in unbecoming ways. But the reasoning has been lost to history and citizens are stuck wondering not only whether they could get in trouble for using the video, but also whether this is an appropriate rule for the House to make.

What I'm going to talk about are the development of principles of open government data, and in particular the small amount of legal history around open government data. And, especially when I get into the bit about the history of the open data movement, I'm really focusing on a very small part of the broader open government movement.

The Constitutional Option

Last year I was asked to give a talk on designing government transparency at the Center for Information Technology Policy at Princeton. Let's say we could start from scratch, what would a technologically-enabled open government look like? Would it involve mandatory open records? Live streaming of all public meetings? Maybe the right to information would be a constitutional right:

"(1) Every citizen has the right of access to information held by the State. (2) Every person has the right to the correction or deletion of untrue or misleading information that affects the person."

In fact, this is no hypothetical right. I took this from the new constitution of Kenya[1] which was ratified last summer. But Kenya wasn't the first to make information a right. According to the website Right2Info.org[2], the right to government held information is protected by the constitutions of around 40 countries. And although the United States is conspicuously missing from this list, access to government information isn't a new idea.

Turn back the clock to early 18th century Qing-dynasty China. Check this out. According to the French writer Jean Baptiste du Halde, who wrote comprehensively on China at the time:

> "Every three years they make a general review of all the Mandarins [officers] of the Empire, and examine the good or bad qualities that they have for government. Every superior Mandarin examines what has been the conduct of the inferior since the last informations have been given in, or since they have been in office, and he gives Notes to everyone containing praises or reprimands...They reward a Mandarin by raising him to a higher degree, and they punish him by placing him in a lower, or by depriving him of his office."

---

1 http://www.nation.co.ke/blob/view/-/913208/data/157983/-/l8do0kz/-/published+draft.pdf
and for more: http://j.mp/98AtiE

2 http://right2info.org/constitutional-protections-of-the-right-to#

The reviews would then be passed up the chain of command, each officer adding his notes onto those of his subordinates. At the highest level, where an account of all of the officers of the empire was put together, the punishments and rewards would be set and instructions would be distributed back down the chain of command, all the way down to the common people.

> "[T]he Mandarins are obliged to put at the head of their orders the number of degrees that they are to be raised or depressed: For instance: I, the Mandarin of this city raised three degrees, or depressed three degrees, do order and appoint, etcetera. By this means the people are instructed in the reward or punishment that the Mandarin deserved."[3]

This could honestly be the first public disclosure practice in history.

A daily gazette published in the empire's capital was distributed throughout the provinces of the empire. The gazette recorded punishments of officers, "expenses disbursed for the subsistence of the soldiers, the necessities of the people [probably care for the old and poor], the public works, and the benefactions of the prince," "laws and new customs that have been established", and other news and propaganda for the emperor. (pages 69-71) I think it is remarkable that in a truly absolute monarchy that such attention and resources would be given to disclosure of government spending and other government decisions. Now, the gazette appears to have been intended to be read by government officers and not the common people, but literacy wasn't always as it is today. The best way to reach the people might have been through the literate officers.

It was practices like these that inspired European thinkers and politicians a quarter way around the world. In Europe, freedom of the press was taking shape, but not government transparency as far as I know. It had been a common practice for governments of the time to grant monopolies. You might remember the East India Company's tea trade monopoly as a precursor to the Boston Tea Party. Although some monopolies like that one continued on throughout the 18th century, one in particular had ended around the time du Halde was writing about China. That was the English monopoly over printing given to the Stationers' Company, which according to Wikipedia ended in 1694. The end of the Stationers' Company monopoly over printing was the beginning of freedom of the press in Europe.

It was in the Kingdom of Sweden, which then also encompassed Finland, that priest, farmer, doctor, and Enlightenment thinker Anders Chydenius created the first freedom of information law.[4] Sweden was a parliamentary government with a weak king, but it had not yet outgrown the practice of government-granted monopolies. Chydenius's home province in a remote part of the Kingdom, in today's Finland, had been on the losing end of government deals that restricted trade. Either for the sake of free trade --- or else for the sake of his own province --- Chydenius demanded freedom of sailing at a local government meeting. His brief role in local politics was followed by a well timed shift in political power in the country which opened the door to his service in the national parliament in 1765. There he continued to defend economic freedom, and a new subject for him, the ability of the public to participate in national debate.

By the end of his time in parliament in 1766 he had created the freedom of the press act, combining both freedom of the press and access to government information. It was enacted without objection. (Ironically, he was forced out of parliament shortly after.) The act actually only remained in effect until King Gustav III's 1772 coup, but it returned in various forms, and today a Freedom of the

---

3 Jean Baptiste du Halde. 1736. The general history of China containing Geographical, Historical, Chronological, Political and Physical Description of the Empire of China, Chinese-Tartary, Corea, and Thibet. Second volume. John Watts: London. page 64-65.

4 Juha Mannine. 2006. Anders Chydenius and the Origins of World's First Freedom of Information Act in The World's First Freedom of Information Act: Anders Chydenius' Legacy Today, ed. Juha Mustonen, Anders Chydenius Foundation.

Press Act is one of the four documents that comprise the Swedish constitution.

Chydenius's act guaranteed access to two types of government information: documents and records of votes:

> 6. "[A]ll exchanges of correspondence, species facti, documents, protocols, judgments and awards . . . when requested, shall immediately be issued to anyone who applies for them."
>
> 7. "[I]n order to prevent the several kinds of hazardous consequences that may follow from imprudent votes, likewise graciously decided that [judges] shall no longer be protected behind an anonymity that is no less injurious than unnecessary; for which reason when anyone, whether he is a party to the case or not, announces his wish to print older or more recent voting records in cases where votes have occurred, they shall, as soon as a judgment or verdict has been given in the matter, immediately be released for a fee, when for each votum the full name of each voting member should also be clearly set out . . . and that on pain of the loss of office for whosoever refuses to do so or to any degree obstructs it."[5]

The Act was broad and strong, and the tradition of freedom of the press that began there in 1766 is the very same reason why today Wikileaks is based in Sweden.

Progress since 1766 has been relatively slow. Twenty-one years later we adopted our constitution. It did not have a general provision for access to government information. It did call for each house of Congress to maintain a journal.

> "Each House shall keep a Journal of its Proceedings, and from time to time publish the same, excepting such Parts as may in their Judgment require Secrecy; and the Yeas and Nays of the Members of either House on any question shall, at the Desire of one fifth of those Present, be entered on the Journal." (U.S. Constitution)

Since the creation of the Constitution, it seems like concerns like access to government information have become marginalized and elite. Somehow issues of government *process*, the way the government operates, has lost its position as a legitimate policy question. And so Congress's constitutionally proscribed journal --- today called the Congressional Record --- has also been known to contain records of events that never occurred, and this doesn't seem to bother many people.

FOIA and Going Online

It wasn't until 1966 --- two hundred years following Sweden --- that we had a Freedom of Information Act, the third such law in world history. The second country to enact a FOIA law, after Sweden, was not surprisingly Finland, which kept Chydenius's spirit going through Russian control in the 19th Century and enacted its own law as a new country in 1919.[6] Our own FOIA law was passed by the House unanimously but was signed only begrudgingly by President Johnson, who in a signing statement pushed back on public access to personal files and other sorts of documents. His worry, which I think is fair, was that some decisions could be made better if deliberations could remain private.[7] And that's still a question of ongoing debate.

Delays and fees, such as documents copied at $1 per page, lead Congress to amend FOIA in

---

5 Peter Hogg. 2006. Translation from Swedish to English of His Majesty's Gracious Ordinance Relating to Freedom of Writing and of the Press (1766) in The World's First Freedom of Information Act: Anders Chydenius' Legacy Today, ed. Juha Mustonen, Anders Chydenius Foundation

6 http://www.freedominfo.org/regions/europe/finland/

7 http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB194/index.htm

1974. But that wasn't a good time for President Ford, who was in the middle of responding to the New York Time's coverage military leaks. Ford vetoed the changes to FOIA hoping to keep control over his office's information, but Congress easily overrode the veto.[8]

Now although FOIA has been a landmark law for government transparency, we're still quite a ways away from open government data. In the '90s things got electric, but there was no substantive change in thinking. The Government Printing Office (GPO) went online in 1994 with the Congressional Record, the text of bills before Congress, and the United States Code.[9] The Republican Party published its Contract with America that year with an emphasis on public accountability. Though the Contract did not mention the Internet, it was the new Republican House leadership that created THOMAS in January 1995, the first website to provide comprehensive information to the public about pending legislation before Congress. Later that year a website created in the private sector by Carl Malamud and others to disseminate to the public for free Securities and Exchange Commission corporate filings was adopted by the SEC and became the official website.[10] Then in early 1996 the Federal Election Commission opened www.fec.gov which was not only a website but, for the first time, also a repository of the raw data related to government transparency which could be analyzed independently by researchers.[11]

The events of 1994-1996 laid the groundwork for major changes in the government transparency movement. But in the first ten years of the government going online, information technology was seen only as a tool for fast and inexpensive information dissemination. What GPO began putting online in 1994 were the same documents it had been printing since 1861. In a world where legal and scholarly citations are by page and line number, it is important to have electronic forms of printed documents that are true to the original's linear form, so that makes sense. And that went on for more than a decade.

The Open Data Movement

Then technologists, rather than journalists, started thinking hard about government transparency. The Center for Responsive Politics's website OpenSecrets.org was probably the first government transparency website based on Big Data, and I learned of the site as a reporter for my college newspaper. After graduating in 2004, I launched GovTrack.us, a website that tracks the activities of the U.S. Congress. It was one of the first websites world-wide to offer comprehensive legislative tracking for free and directed at everyday citizens. Most of the information on the site could be found elsewhere, but in so many difference places and in formats that were hardly useful to the American public. GovTrack "screen-scrapes" THOMAS and other government websites, normalizes the information, and creates a large, open database of Congressional information. GovTrack has created RSS feeds for the activities of Congress, interactive maps of Congressional districts, and change-tracking for the text of bills. It's goal has been to apply the latest technology to tracking the activities of Congress.

Today GovTrack reaches about half a million people each month directly, and well over a million if you count visitors to third-parties websites and mobile apps that are based on GovTrack in one way or another. When I opened up the source data that powers GovTrack in XML and RDF formats, others quickly began using it. The two biggest reusers of the data are OpenCongress.org, which has a fantastic blog that covers the activities of Congress in a way that's clear for folks outside the beltway, and MAPLight.org, which puts a new spin on the connections between money and politics.

8 http://www.gwu.edu/~nsarchiv/NSAEBB/NSAEBB142/index.htm

9 http://www.gpo.gov/congressional/testimony/jcp-may94.htm

10 The website was created by the Internet Multicasting Service (lead by Carl Malamud; now Public.Resource.Org), New York University, and a National Science Foundation Grant. http://public.resource.org/sec.gov/index.html

11 FEC 1996 Annual Report http://www.fec.gov/info/arfrm.htm

Another interesting one is IBM ManyBills, which is a visualization tool like their ManyEyes but for congressional legislation, and there's a mobile app called Congress built by the Sunlight Foundation. At least two dozen websites have popped up relying on GovTrack data all trying to give the public a new way to get a grasp of their government --- I'm sure there are many I'm not even aware of.

The combination of "Web 2.0" as a buzz-word and grassroots digital campaigning in 2004 probably had a lot to do with what happened next. The Howard Dean presidential campaign got a publiciy boost from developers coming together to specialize the open source CMS Drupal for political campaigning, making CivicSpace. That sent a message, even if no one quite recognized it at the time, that developers have a role to play in the world of civics. CivicSpace crystalized some vague notion of civic hacking.

Toward the end of 2005 I got an email from Micah Sifry, who ran the Personal Democracy Forum, the conference at the intersection of technology and politics. He wrote, "I'm doing some consulting for a new organization that is seeking to open up the system to more scrutiny, and it would be great if we could do this in a way that maximizes the network effects." Little did I know that what happened was that Ellen Miller, who had run the Center for Responsive Politics, had met Mike Klein, a lawyer and entrepreneur who had made a killing by discovering the value of commercial real estate data. And between Ellen, Mike, and Micah, and some others, the Sunlight Foundation was born.

Sunlight was primarily a grant-making institution until late 2006 when a telemarketing manager in Pittsburg posted a note on DailyKos recruiting help for a citizen journalism project. The goal of this Congressional Committees Project was to assign one person, a regular citizen, to each committee and subcommittee in Congress. That person would follow the committee closely and report back what the committee was doing to the group in a non-partisan way. The project was all set to begin, just waiting for Congress to come back from winter recess, when a staffer for the new Democratic Speaker of the House (Nancy Pelosi) contacted the leader of this project and asked to have the group collect some feedback about how Congress could be more transparent in the coming years. And that was the end of the committees project because the telemarketer went off to lead what became the Open House Project, which became a cornerstone of the work of the Sunlight Foundation. And that was John Wonderlich, who is now Sunlight's policy director.

Later that year, 2007, Carl Malamud --- who had liberated the SEC data in the '90s --- lead a workshop that wrote the 8 Principles of Open Government Data. The 8 Principles gave us consensus on general principles, and I think it's principles that make a movement. That workshop was a real who's-who of this movement and I was in shock to be meeting folks like Larry Lessig and Tim O'Reilly, heros of geekdom. Before I went out to the workshop a friend of mine asked me to find out why O'Reilly's books all had animal covers, and Tim didn't give me a satisfactory answer.... but he later became an early investor in POPVOX so I'll forgive him. Anyway, I'll come back to what those eight principles were later.

It wasn't until 2009, the year of the first two transparency camps and Clay Johnson's announcement that we had become a movement, that open government data started to finally make strides actually inside government. That year the GPO recognized that legal documents could be useful in other electronic forms besides PDFs, that information technology can make it easier to search, sort, share, discuss, and understand government publications. Along side its publication in plain-text and PDF of the Federal Register --- which is the executive branch's publication of notices and new rules --- GPO added a database of the Federal Register in XML format, a data format that makes it possible for

innovators in the private sector to create new tools around the same information.[12] Then we get a similar story from the private sector. The Archivist of the United States explained what happened next on his blog:

> "In August 2009, Andrew Carpenter, Bob Burbach, and Dave Augustine banded together outside of their work at WestEd Interactive in San Francisco to enter the [Sunlight Labs Apps for America 2] contest using data available on data.gov. Understanding the wealth of important information published every day in the Federal Register, they used the raw data to develop GovPulse.us, which won second place in the contest. In March 2010, the Office of the Federal Register approached the trio to repurpose, refine, and expand on the GovPulse.us application to bring the Federal Register to a wider audience. Federal Register 2.0 is the product of this innovative partnership and was developed using the principles of open government."[13]

Again, a private sector innovation based on government data was good enough to be adopted as an official government resource. USASpending.gov was created in this manner as well, starting as a project of the nonprofit OMBWatch.

The same year we saw the Senate start publishing votes in XML (following six years behind the House). The two chambers of Congress operate entirely separately so the two chambers of course use completely different schema and IDs for Members of Congress, and I don't know of any plan to unify them. The Government Printing Office released bulk XML downloads for the Federal Register, as I mentioned, and the Code of Federal Regulations. These make up a substantial part of U.S. law. The House has published bulk data for the U.S. Code for some time, so it's really just the judiciary now that is the sore thumb in creating open access to federal law. That year the House began publishing its spending data electronically, and we had the launch of data.gov and the Open Government Directive[14] which called for innovation in transparency, participation, and collaboration. In fact to my surprise, the Directive addressed nearly all of the 8 Principles of Open Government Data, and essentially added two of its own: being pro-active about data release and creating accountability by designating an official responsible for data quality.[15]

Modern Legal History

Data.gov and the Directive have been generally criticized on two fronts. First, the apparent success of data.gov has largely ridden on the inclusion of data sets that had already been available to the public. My favorite data set included early on in data.gov was Federal Aviation Administration flight on-time statistics, which has been released in some form since 2003[16]. (A fellow coder Josh Sulkin and I built FlyOnTime.us, which used historical on-time and weather data to predict future delays. As it happens, another group had the same idea and built flightcaster.com.) The second line of criticism is that whatever new and supposedly "high value" data that was released following the Directive has been simply not very interesting for government transparency. The highest rated dataset on data.gov now is "Active Mines and Mineral Plants in the US" from the U.S. Geological Survey. Environmental and weather data comprise a large part of the data catalog. These are certainly important data sets for their connection to public safety. If journalists get a deeper perspective on mine safety and

12 http://www.washingtonpost.com/wp-dyn/content/article/2009/10/04/AR2009100402533.html
13 David Ferriero, "Coming Soon: Federal Register 2.0", National Archives Blog, July 26, 2010. http://blogs.archives.gov/aotus/?p=1317
14 http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf
15 http://razor.occams.info/blog/2009/12/09/open-government-directive-evaluation-on-principles/
16 http://www.transtats.bts.gov/releaseinfo.asp?tb=236&display=release

if that saves lives, then it would be hard to name an even more important data set. But the datasets don't fulfill the promise of transparency. For that, we're looking for open access to administrative records, records that tell us how decisions were made and that help the public stay informed about agency activity. It's really no surprise that data.gov has excelled in the sort of data sets it has since the Environmental Protection Agency and especially the National Oceanic and Atmospheric Administration in the Department of Commerce, the original source of most weather reports, have been leading the public dissemination of raw government data since well before there was an open government data movement. Still, the benefit of data.gov may be less in the catalog itself and more in the standards it sets for federal agencies and the cultural change it symbolizes. Harlan Yu points to what he says is underappreciated infrastructure:

> "There's a Data.gov manual that formally documents and teaches this process. Each agency has a lead Data.gov point-of-contact, who's responsible for identifying publishable datasets and for ensuring that when data is published, it meets information quality guidelines. Each dataset needs to be published with a well-defined set of common metadata fields, so that it can be organized and searched. Moreover, thanks to Data.gov, all the data is funneled through at least five stages of intermediate review—including national security and privacy reviews—before final approval and publication. That process isn't quick, but it does help ensure that key goals are satisfied."[17]

Health and Human Services stands out as one of the few federal agencies that hadn't had a prior commitment to open data that has strongly embraced the Directive, now having released to the public data sets including FDA drug labeling, farmers' market locations, Medicare and Medicaid aggregate statistics, and national health care spending estimates, among others.[18] <<<challenges?>>>

The Directive also called for a 10% reduction on FOIA backlogs per year, and on that point the Administration may have set itself up to fail by not realizing its agencies had already cut their 2009 backlog by more than half compared to the backlog at the end of the year before. An analysis by OMBWatch of 25 key agencies showed that the backlog was substantially reduced even though there weren't any fewer requests than the year before. But the Directive was only issued at the end of that year. At the end of 2010, the first year of the Directive, the backlog remained roughly the same.[19] So on this point, it seems technically the goal was not met.

Finally for 2009, open standards laws were passed in Vancouver and Portland:

> "Open and Accessible Data - the City of Vancouver will freely share with citizens, businesses and other jurisdictions the greatest amount of data possible while respecting privacy and security concerns; Open Standards - the City of Vancouver will move as quickly as possible to adopt prevailing open standards for data, documents, maps, and other formats of media;"[20]

and

> "[T]he Council of the City of Portland . . . [d]irects the Bureau of Technology Services to . . . [d]evelop a strategy to adopt prevailing open standards for data, documents, maps, and other formats of media; "[21]

---

17 http://www.freedom-to-tinker.com/blog/harlanyu/what-we-lose-if-we-lose-datagov
18 http://health.data.gov
19 http://www.ombwatch.org/files/info/fy2010foiaanalysis.pdf
20 http://vancouver.ca/ctyclerk/cclerk/20090519/documents/motionb2.pdf
21 http://efiles.ci.portland.or.us/webdrawer/rec/3675248/view/

There was one new open data law was passed in 2010: it was in San Francisco. The municipal administrative code on Open Data Policy was the first to adopt language from the 8 Principles of Open Government Data and it called for technical requirements to be created that for the purpose of "making data available to the greatest number of users and for the greatest number of applications" and with "non-proprietary technical standards" and a "generic license" such as a Creative Commons license.[22] Unfortunately, a quick perusal of DataSF.org, a sort of data.gov, shows that the city's GIS data are all still hidden by an innocuous but non-generic click-through license.[23]

A variety of other bills were introduced but not passed that year, making 2010 the year of posturing. A similar bill like the San Francisco bill was introduced in the New York City Council but was not passed.[24] At the federal level we saw introduced in Congress H.R. 4983: Transparency in Government Act of 2010, sponsored by Illinois congressman Mike Quigley. The bill would have increased financial disclosures of Members of Congress and staff, put earmark requests online, created an online House committee meeting schedule, required 72-hour availability of bill text before consideration, made CRS reports public, required shared identifiers between a few federal contracting datasets, and much more. This bill covered it all, but as far as I remember never got a hearing. That bill and H.R. 4858: The Public Online Information Act of 2010, or POIA, sponsored by New York congressman Steve Israel, were brainchilds of the Sunlight Foundation. POIA was a simpler bill with only a few points. First, records in the executive branch should be put online. Second, the bill would have created an advisory committee for goverment-wide guidelines on Internet-enabled transparency.

At the end of the year then-Illinois congressman Bill Foster, who lost reelection, introduced H.R. 6289: Legislative Data Transparency and Public Access Act of 2010. This bill should have been a no-brainer: It would have required the Library of Congress to publish its database that powers the THOMAS website, and then I could get out of the business of screen-scraping THOMAS. Of course this bill, like the others, went nowhere. Also in 2010 Quigley and Congressman Darrel Issa of California formed the Congressional Transparency Caucus, which has subsequently done nothing for transparency.

This year, the new rules of the House of Representatives brought in by the new Republican leadership addresses public access to committee records and data formats, though I am not aware if the rules have had any practical consequences yet. An open data bill incorporating all of the 8 Principles of Open Government Data was introduced in New Hampshire[25] by an open-source coder turned state representative, and POIA was reintroduced this time with a companion bill by Montanna Senator John Tester. A 72-hours bill was introduced in the House (H.Res. 230[26]). And GPO last month added a bulk data download for Public Papers of the Presidents of the United States (2009).

And that's my brief legal history of open government data. Progress is slow, but it is sure.

Principles of Open Government Data

Over the years I've kept notes on all of the different principles for open data people have suggested. I've found some sixteen principles or best practices that have been proposed for what "open government data" means. Starting with the 8 Principles, public government data should be complete and primary, basically meaning not aggregated; it should be timely enough that it is still relevant to any ongoing policy debates; it should be accessible in both a disability and digital sense; and perhaps most importantly it should be machine-processable so that tools can be built to search, sort, and transform

22 http://www.sfbos.org/ftp/uploadedfiles/bdsupvrs/bosagendas/materials/bag110910_101155.pdf
23 http://gispub02.sfgov.org/website/sfshare/index2.asp
24 http://nycctechcomm.wordpress.com/opengov/; http://www.scribd.com/fullscreen/34124837
25 http://www.gencourt.state.nh.us/legislation/2011/HB0310.html (HB 310)
26 http://sunlightfoundation.com/blog/2011/04/18/new-72-hour-rule-bill-introduced/

the data into other form. Government data should be available on a non-discriminatory and license-free basis and in a data format that is not encumbered by patents.

Now this is where things start to diverge between the United States and European countries. The Open Knowledge Foundation's Open Knowledge Definition, or OKD, and the European Union Public Sector Information Directive, specifically allow for certain types of licenses. The EU PSI directive allows for licenses that cover "liability, the proper use of documents, guaranteeing non-alteration and the acknowledgment of source." But in U.S. culture where the government can't normally copyright works and where any restriction on the free flow of information comes under high scrutiny, any license restricting the use of government published information is completely unacceptable --- except where the information might impinge on private or security or other narrowly defined exceptions. The European culture is different. In countries with a crown copyright, it might seem less onerous for a government to tell you not just whether you are permitted to share any arbitrary government-copyrighted data but also *how* you are permitted to do so according to the license's terms. I'll be a moral relativist on this one and let that slide.

Government data should also be online, free, and at a permanent address. The US ACM recommends it should be presented in a format that promotes analysis and reuse, does not impose computer security risks to users, and retains provenance through digital signatures. Others have suggested it be crafted with public input and inter-agency coordination and is subject to ongoing public review, uses globally unique identifiers for records, is web searchable, and is published with Linked Open Data methods.

In the 8 Principles of Open Government Data, we phrased the principles as a definition of open government data. But it would be a mistake to take openness to be a binary value. Governments are storied institutions with long histories, deep existing infrastructure, and competing priorities and demands. It is impossible to expect that governments could make all data instantly open on every dimension. And that's fine.

The Why

So what does open data buy us? Why have we been struggling for 250 years for freedom of information? During the debate over Kenya's constitution in 2005 --- as I mentioned, it has one of the newest freedom of information provisions --- political leaders faced the problem of informing a significantly illiterate electorate about the choices they faced. To help them at the polls, oranges became the symbol of a no vote while bananas the symbol for a yes vote.[27] There is a wide gulf between government data and public value. It's up to developers and statisticians and designers and other data practitioners to help the public make sense of government data.

Let me draw a line between two types of government data. One the one hand, we can have access to government records so we can do things like read the law[28], or avoid bad drinking water[29], or get tomorrow's weather forecast[30]. These types of government records directly help the citizen have more productive lives. Health, environmental, and workplace data can literally save lives, and we just have to think of last year's mine collapse and the investigation into safety records afterward to see that.[31] And we're in a so-called information economy, so government information of all sorts can create jobs in the private sector. It's created mine, and everyone else's at the Sunlight Foundation, and at companies

27 http://www.nytimes.com/2005/10/16/international/africa/16kenya.html?_r=1

28 http://www.washingtonpost.com/wp-dyn/content/article/2009/10/04/AR2009100402533.html?hpid=sec-politics

29 http://www.nytimes.com/2009/12/08/business/energy-environment/08water.html

30 http://razor.occams.info/pubdocs/opendataciviccapital.html

31 http://investigativereportingworkshop.org/investigations/coal-truth/story/massey-had-worst-mine-fatality-record-even-april-d/

like EDGAR Online, ESRI (the geospacial data company), and Accuweather.

Then on the other hand we have access to government records for the purposes of government oversight. Records like spending data, lobbying disclosure, the personal calendars of Members of Congress, and the FOIA backlog. These records help make our government more efficient, and that's important too, but it's very different.

Beth Noveck, the professor and former deputy chief technology officer for open governmet, wrote in the Huffington Post that we still need to work on our terminology. "In retrospect, 'open government' was a bad choice. It has generated too much confusion. Many people, even in the White House, still assume that open government means transparency about government." Rather, it's broader than that, she wrote. "The aim of open government is to take advantage of the know-how and entrepreneurial spirit of those outside government institutions to work together with those inside government to solve problems."[32] In fact, that's not really even broad enough for the benefits of open government that she wrote about in the same article, but it's certainly the right spirit. The White House recently launched whitehouse.gov/goodgovernment which clearly labels their focus on that sort of open government, and leaves room for separate branding for the other type of open government. And think of it this way: open government has grown big enough at the federal level that it needs two names.

Both types of government data can have unintended consequences. Michael Gurstein wrote in a blog post[33]: "Newly available access to land ownership and title information in Bangalore was primarily being put to use by middle and upper income people and by corporations to gain ownership of land from the marginalized and the poor." Gurstein pointed out that not all data yields an "effective use" of the data. Clay Shirky has claimed[34] that transparency fueled the lobbying business. I don't know if that's true, it's a little before my time, but I found LobbyData.com. From $50 to $500 per month you can get profiles of lobbying firms to help you find your next client or do competitor research. Clay wrote: "If transparency lets all interest groups make use of improved information, then we would expect that the better organized interests to make better use of any new transparency. This is not to say that transparency is never good; it _is_ to say that it isn't _always_ good, and that the negative effects result from imbalances in the will to collective action, not just access to information." This is the first way transparency can be a paradox: transparency is not necessarily the end game that you want.[35]

New levels of transparency can also affect the procedures within government in unexpected ways. The more we can see of Congress, the more congressmen want to take their negotiations somewhere else more private. And by no means is that necessarily a bad thing --- some negotiations need that off-line personal touch --- but it's never an intended outcome of transparency. The New York Times famously reported in 2009[36] that a ban on some lobbyist-financed trips by Members of Congress successfully cut the number of such trips in half or more. But they continue. The times article said, "Some lawmakers and even their families continue to take trips hosted by private groups and companies that revel in their access to Washington power brokers. An examination by The New York Times of 1,150 trips shows that some of them bent or broke rules adopted in 2007 to limit corporate influence in Washington. Others exploited glaring loopholes in the guidelines." Another Times article reported that after a new 2008 ethics law imposed sticter lobbying reporting requirements the number of registered lobbyists started to dramatically decline.[37] The article's headline said it all: "Law to Curb Lobbying

---

32 http://www.huffingtonpost.com/beth-simone-noveck/whats-in-a-name-open-gov-_b_845735.html

33 http://gurstein.wordpress.com/2010/09/02/open-data-empowering-the-empowered-or-effective-data-use-for-everyone/

34 http://groups.google.com/group/openhouseproject/msg/53867cab80ed4be9

35 http://groups.google.com/group/openhouseproject/msg/53867cab80ed4be9

36 Rules for Congress Curb but Don't End Junkets, The New York Times, December 6, 2009. http://www.nytimes.com/2009/12/07/us/politics/07trips.html

37 http://www.nytimes.com/2010/01/18/us/politics/18lobby.html?ref=politics

Sends It Underground." This is the Wonderlich Transparency Paradox. This is one of those terms that is bound to make the textbooks one day. The paradox is this: "How ever far back in the process you require public scrutiny, the real negotiations . . . will continue fervently to exactly that point."[38] Or, to paraphrase, no matter how much transparency you put into the system, the players in that system are always going to find a back room to work in. This was said by John Wonderlich and I first heard it called the WTP by Marci Harris, who's now my business partner in POPVOX. See, I'm even abbreviating it to push it along on the road to becoming a household term.

In another example, antiquated transparency laws can actually hinder future transparency and innovation. State open meetings laws are preventing local officials from engaging citizens online for fear of inadvertently having what would be legally considered a meeting, that had not met other requirements such as public notice and public access.[39] At the federal level, a good example is the 1972 Federal Advisory Committee Act. The purpose of the act was to prevent federal agencies from having backroom consultations with corporate executives and other entrenched sources, but the requirements for obtaining public input have grown so complex under the act that today the law is probably preventing federal agencies from getting the best knowledge.

Transparency can be a paradox. We can't always use data to pry a closed government open. And data by itself doesn't necessarily empower citizens, and if it does it might empower the wrong ones. Data is not a great end goal, it's only a way-point on the way to education, journalistic reporting, and computer applications.

But I will end saying that maybe transparency is *sometimes* an end to itself. In The Matrix, remember that Neo had to choose between leading a fake life, ignorant of how the world really was. Or he could leave the Matrix but face a more challenging and sometimes horrifying life. He chose to face the truth.

38 http://groups.google.com/group/openhouseproject/msg/94060a876083d86a

39 http://texasmunicipallawyers.com/pdf/OpenGovt-theNet-TTAdmin-Fall09.pdf