

A Taxonomy of Open Government Applications

■ indicates to advance the slide, it's a note to self

■ Some time in the last fifty years or so something important changed in the national debate. From 1951 to 1971, five amendments to the U.S. Constitution were ratified all having to do with intricate aspects of the process of governance, including setting presidential term limits, refining the electoral college, and lowering the voting age. The last major reorganization of Congressional procedure was forty years ago, in 1970. Since then, questions of government process have been marginalized. Should states elect representatives proportionally or by district plurality? How many seats should Congress have? These issues are now considered elite.

And yet, and yet, the apparent marginalization of government process belies what I think Americans really feel, which is that government processes are really important to us. For one, we still really hate corruption. ■ A survey presented in the book *Stealth Democracy* (2002) by Hibbings and Theiss-Morse found something interesting: individuals who responded that they would disobey the law if they thought it was wrong were not those that actually thought the law was wrong, but rather were those who thought government *process* was wrong. In other words, Americans can believe in the rule of law even if they disagree with the law, but only to the extent they think the law was made *in a fair way*. Process is still really important, we just don't talk about it!

And since we don't talk about it, citizens don't understand how complex our government right now is, and that leads to cynicism. Some cynicism is warranted. But not all of it. Unwarranted cynicism is a problem.

■ I don't like politics. But I do like data, and solving problems with data. Seven years ago I launched my first attempt to solve a problem using government data. GovTrack.us was the first website to present to the American public a unified account of what Congress was doing, along with tools to track future legislative activity. The site includes bill status, voting records, ■ biographical information on Members of Congress, detailed outlines of congressional districts, and other information collected from official sources. GovTrack was initially just a hobby, but now I consider it more of a light-weight business.

There *are* government websites that display different aspects of all of the information, such as THOMAS.gov. ■ THOMAS.gov was created in 1995 as part of the Republican Contract with America and while it's been largely unchanged for fifteen years, it continues to be a reliable source of bill status information. ■ But there is no government *database* of legislative information that programmers like myself can transform into new applications. The difference between THOMAS.gov and the database *behind* THOMAS.gov is a bit like refrigerator poetry magnets that have been glued into a pre-written sonnet. THOMAS.gov is the prewritten sonnet. You can appreciate the sonnet, sure, but the glue has limited the potential that comes from being able to rearrange the pieces and discover new meaning --- that's what open government applications do.

May I make my own sonnet, I ask? ■ I wrote Congress's response as a haiku: "Use THOMAS.gov // that should be enough for you // we don't share data." But the open data response is a one-word poem: Innovate! ■ [joke slide]

By putting lots of information together, new possibilities emerge. ■ Novel statistics about the performance of Members of Congress become possible. In this graph, each point is a congressman. On the x-axis we have a political spectrum score, and a leadership score is graphed on the y-axis. The data

tells us something interesting: the V-shape pattern seems to indicate that congressional leaders are also the most ideologically extreme. ■ Many things are possible when you have the data. By overlaying Census geographical data with Google Maps, it became possible to reliably determine your congressional district by zooming to street level, which is crucial if you live either near a district boundary or in a metropolitan area. ■ The site has the only bill text reader that can compare versions of bills to show text added or removed during a bill's life cycle, and also will pop-up excerpts from the U.S. Code next to references to the Code in the bill text.

Now I scratch my ear like this <!-- to get data out of THOMAS. What I do to get the information isn't pretty, because the Library of Congress, which runs THOMAS, doesn't share its database, it doesn't share its unglued refrigerator poetry. I've been asking the Library to share its data for more than 10 years. In fact, in 2009 an Act of Congress encouraged the Library to open its data. After that stalled, there was a bill dedicated to this issue in 2010. Still no progress. The current House Republican leadership has promised to make this sort of data available, but we can only wait and see whether and how the promise will be fulfilled. Instead, I've programmed GovTrack to periodically go out to government websites and figure out whatever information it can extract out of what it sees. The process is called "screen scraping": programmatically loading up web pages, looking at their HTML source, and extracting information using simple pattern matching. It's not interesting programming work, and screen scrapers are easily confused. This is not a good way to do it, but it's the only way to do it right now.

■ GovTrack reaches about half a million people each month directly, and well over a million if you count visitors to websites and mobile apps built by others on top of GovTrack's legislative database. When I opened up the source *data* that powers GovTrack, others started to see the potential for building other tools that shed light on government processes in new ways. The three biggest reusers of the data are OpenCongress.org by the Participatory Politics Foundation, MAPLight.org, which puts a new spin on the connections between money and politics, and mobile apps created by the Sunlight Foundation. ■ There are a lot of visualizations of legislative data now. Here on the top left is a visualization of Senate voting patterns by Don Smith, using a statistical technique called multidimensional scaling. And at clearcongressproject.com, Thomas Gibes riffed on my ideology/leadership analysis for his master's thesis and has been blogging about his results, on the lower right.

■ So I like to think that when a bill number --- like H.R. 3200 --- is said on air during a late night TV show that GovTrack might have contributed to the greater public consciousness of the legislative process. And that goes back to reducing unwarranted cynicism through civic education.

■ GovTrack is the sort of open government application that I'd put in the category of "democratizing primary legal materials." Carl Malamud has been doing this for a long time and is now leading an effort to fill in the spots where primary legal materials are not freely available to the public at all. These include a vast amount of judicial opinions and other court documents, some state law, and a great deal of building codes and other law managed in the private sector but incorporated into law by reference. Carl's project, which goes under the moniker Law.Gov (but the website is law.resource.org), points to many practical implications of broader access to the law, such as deeper research in universities, reducing costs for the government and small businesses, and greater access to justice. According to the schedule Carl is speaking tomorrow at 9am. The project called RECAP (recapthelaw.org) out of the Princeton University Center for Information Technology Policy highlights the question of why anyone should have to pay to access court documents.

Whether it be bill text or documents in court dockets, there is a lot of innovation the private sector wants to create with primary legal documents. A Knight Foundation News Challenge winner named The State Decoded, run by Waldo Jaquith, listed some possibilities for how to display primary legal documents: "embedded definitions, cross-referencing links, ..., commenting, tagging, decent

design, and humane typography.” Waldo previously created RichmondSunlight.org, which is like GovTrack but for the Virginia state legislature.

■ So by no means is open government exclusively or even mostly about corruption. In fact I think some of the most interesting reuses of government data solve problems in our lives that have nothing to do with government. Google Maps and the rest of the online mapping and in-car navigation industry might not have ever come to be without open data from the Census Bureau on the nation’s roads, the U.S. Geological Survey’s satellite imagery and terrain data, and GPS signals from government satellites. Weather reports in the media are predominantly derivative of data emanating from the National Weather Service. The business world relies on XML corporate disclosures from the Securities and Exchange Commission to keep investors informed. Early applications in the modern open data movement were crime maps using local police data (for instance Adrian Holovaty’s [chicagocrime.com](#) in 2005) and public transportation tools such as mobile apps for train schedules. ■ That’s all what I call “data as civic capital.”

■ The third sort of open government application that I want to highlight is the website that sheds light on a particular government problem. Keep in mind how these examples differ from the democratizing documents category. ■ One application that I love is Party Time, from the Sunlight Foundation. Party Time documents the “political partying circuit,” the continuous stream of congressional fund-raising events taking place in D.C. restaurants and clubs. Party Time does a remarkable job of educating the public about conflicts of interest, in part by posting scans of event invitations, like this one. In this invitation to an event for Congressman Mike McIntyre, to attend an individual must contribute at least \$250 to McIntyre’s campaign. For those who contribute more, special honors are given, which might translate into special access. This isn’t government data per se. Party invitations aren’t government records. But I find this to be a particularly compelling example of this sort of open government application. ■ There are plenty of similar applications that use real government data, for instance two other Sunlight projects named Elena’s Inbox and Sarah’s Inbox which are Gmail-lookalike interfaces into the publicly released archives of the emails of Supreme Court Justice Elena Kagan and former governor Sarah Palin.

■ The last sort of open government application is the open government business. Now, there are plenty of businesses based on government data, and many more that make use of government data. But here I mean companies whose mission is to use government data in the furtherance of an open and transparent government.

So I have a business like that. It’s called POPVOX.com. ■ Congressional offices receive more than 300 million emails a year, per day that’s around 300 to 2,000 emails for each congressional office. Now knowing this number, you could imagine that these communications clog email in-boxes. ■ Congressional offices only have a few staffers to process the incoming letters and to write bulk responses. Some offices tally the constituent opinions in the letters, but no one is really reading the letters. That’s really unfortunate. Americans put in a lot of time into writing these letters, trying to make convincing arguments, and advocacy organizations pay PR firms what can come out to \$125 per letter sent to Congress. POPVOX creates a channel for effective and efficient communication from constituents to their Members of Congress.

■ We don’t have any special access to Congress, but what my partners and I have is the knowledge of what makes an effective campaign. An effective campaign isn’t one with higher volume, it’s one that provides the right sort of information to Congressional staff, in a way that is easy for them to process, and with a public, accountable component. It’s not rocket science, and yet some of the largest advocacy organizations are getting it wrong and simply pissing off a lot of Congressional offices for dumping on them a lot of noise. That’s all it is. If the office can’t process the information,

it's just noise.

■ To give you a better idea of what POPVOX does, I'm going to walk through a few screenshots. You start by finding any bill in Congress that interests you. Congress considers some 10,000 bills in every two-year session, and there is something for everyone on POPVOX, from balancing the budget to ending horse slaughter. ■ Once you find a bill, you research what the bill is really about and then weigh in, by choosing support or oppose. It's cut and dried for a reason. Nuanced choices like "support if you amend it a certain way" is just not something congressional staff are prepared to process when they take in these communications. ■ So you choose support or oppose, and you can write your letter to Congress. We know that congressional offices find personal stories the most useful. A letter making a logical argument is nothing compared to a story like this one. Donkeygirl from Missouri's 2nd congressional district wrote, "We have owned horses for over 40 years. There was an incident many years ago when we woke up and discovered that all 8 of our hoses were missing. This is when a slaughter facility was located in Missouri. I can't put into words the fear that kept me awake at night, nightmares of my mare stuck in a trailer going to God knows what fate." We delivered that letter, like all of the others, electronically to the right Members of Congress. ■ And then because every letter on POPVOX is a public letter, you can go on to see what other constituents have said on the issue.

■ POPVOX uses the same sort of legislative information as GovTrack, and gets it in the same roundabout way from THOMAS.gov. We're using that information to make Congress's job of processing their mail easier. If only congressional offices recognized that their own legislative data is helping us make their mail processing more efficient, then maybe they would get THOMAS to share its legislative database, and we wouldn't have to scratch our ears like this <!> anymore.

There is a lot more to the site, but that is the core of it.

■ I want to wrap up by sketching the ecosystem of open government data in terms of information producers and information consumers. This [what's on the slide] is rare. The amount of information consumers get directly from government is relatively small. This is, uh, FOIA requests, sound bites, or what you can get by visiting an agency website. ■ One of the roles librarians play is of course directing information consumers to the right information producers. ■ Most information flow is different. Mediators such as journalists have always played a substantial role in the dissemination of government information. Today mediators include not just reporters but also ■ programmers, designers, statisticians, and other data practitioners. Open government data therefore isn't some revolutionary thing of the future. It's just a new technological approach to the mediated relationship between government and consumer that we've always had. But because data geeks like me tend to look toward the government for new data toys, we tend overall to not be very consumer focused [of course this doesn't apply to POPVOX which, as a business, is entirely consumer focuses]. ■ But libraries are often advocates for their members, and I think between the data practitioners and the librarians ■ we could find some harmony.

■