

A Government Data Haiku

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

■ Okay you all know refrigerator poetry right? Magnetic words and you arrange them on your refrigerator to say something deep, right? So one day I'm out in a store buying a pack of refrigerator poetry. And I get home, and I open the pack, and the words are already glued together. It's already been put into a poem and you can't rearrange it. Didn't make any sense. Now the poem was great. It touched me. It was this poem on the slide. It's deep. But gluing the pieces together locked away the joy I could get out of the 40,319 other poems I could make with these eight pieces.

Refrigerator poetry is a lot like data. Data is building blocks. It has limitless potential for new applications, totally useless when you get it all mucked up and glued together into what a government agency wants it to say. ■ If the Public Printer only spoke in the form of a haiku, he might have said in 2009, "Use FedSys website // It should be enough for you // We don't share data." And what I mean is, there is so much more to do with government documents than making them into PDFs. Our government creates an enormous amount of data, raw data, and so much of it is all glued up.

■ Not all of it, though. And some of the most interesting reuses of that government data solve problems in our lives that have nothing to do with government. Google Maps and the whole 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. And the business world relies on corporate disclosure data from the Securities and Exchange Commission to keep investors informed. ■ And of course there is also all of the research based on reams of data that the Census Bureau has always published. All of these applications are based on what I call "data as civic capital." These are government spreadsheets, databases, and other digital records that solve real problems in our lives when they get into the hands of entrepreneurs, journalists, and researchers in the private sector.

We don't expect our government to build our driving directions apps. We don't want to turn to a government TV station to find out the weather. And we would be up in arms if the only people who could do population studies were Census workers.

Yet when it comes to the wonkiest of government data, like GPO docs and federal spending databases, somehow we're much more complacent. As if knowing what the government is doing, and being able to teach that to others, were not as important as saving ten minutes by optimizing your travel using Google Maps.

■ ■ 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, including the Census Bureau, the Library of Congress, GPO, and the House and Senate.

■ ■ Using GPO documents I've built a tool to track changes to the text of bills on their way to a vote ■ and a widget you can use to embed a bill on your website. ■ Others have built websites where people can leave comments on particular paragraphs of bills. ■ The Sunlight Foundation built a tool to show the most frequent words in the Congressional Record. And when GPO finally began publishing raw data for documents including the Federal Register, developers jumped right on it and one of private sector projects was actually adopted as the new official website for the Federal Register.

You can't build this stuff with PDFs. And GPO, and federal agencies, are starting to realize now that the public has a lot to gain from developers, statisticians, designers, and journalists having deeper access to government information.

■ GovTrack is the sort of open government application that I'd put in the category of "democratizing primary legal materials" because it gives wider access to basic building blocks of law, such as the text of bills that might become law. I think Stephen is going to talk more about these sorts of applications.

■ The third sort of open government app in this typology I'll leave for Tom to talk about.

■ 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. ■ And the basic idea is that Congress is overloaded. 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. POPVOX creates a channel for effective and efficient communication from constituents to their Members of Congress. ■ We do that in part by creating public accountability for what Members of Congress are receiving through the public bill report. It's also a tool we've found that congressional staff themselves find useful for understanding what the public is saying. And we make sure that the letters our users are writing are as effective as they can be by delivering them in an electronic format that is easiest for congressional staff to process. In our short time in operation we've delivered hundreds of thousands of messages to Congress, and many of those messages have gotten replies that make our users feel like they are a part of the government process --- which is how American government should be.

■ 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 and POPVOX might have contributed to the greater public consciousness of the legislative process. These are episodes of the Daily Show and Late Night with Jimmy Fallon when they've dug into legislative matters in surprising detail for a late night now. And it encourages me --- it says the American public wants the details and wants to be involved, but they want their government information in the right format. Most open government data applications, including mine, are pretty simply taking government data that consumers don't want and transforming it by rearranging those refrigerator poetry pieces into something consumers do want and can use.

■ 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, what you can get by visiting an agency website, or the materials in a depository library. ■ One of the roles librarians play is of course directing information consumers to the right government information producers to make that connection. ■ 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 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. ■