

The Flawed Economics of Closed Government Data

Brian Timoney, MapBrief™, March 26, 2013

<http://mapbrief.com/2013/03/26/the-flawed-economics-of-closed-government-data/>

How much should citizens pay a county for a digital copy of property records and aerial photos? Sciotto County, Ohio says \$2000. Actually it hired Woolpert to figure it out for them, and they said \$2000. Which sounds a bit spendy, especially given that *cities like Philadelphia and Denver give away the same type of information. For free.* Such a discrepancy in pricing would indicate someone is doing it wrong. But Sciotto County went to the trouble of arguing their point to the Ohio Supreme Court where six of the seven judges agreed with them.

Others have pointed out the wrongness of the decision. But what if the majority's conclusion flowed from premises about the relationships between data, software, and distribution formats that simply don't apply in 2013? To what degree should a citizen be economically punished for the technical inefficiency of its government?

What if It's More Cost-Effective To Give Data Away Than To Charge For It?

The costs of distributing digital information over the web have plunged dramatically over the past few years: the infrastructure to, say, distribute aerial imagery of a city, would have cost thousands annually—now it's probably a few dollars per month. But while distribution costs have plunged, employee costs haven't: wouldn't efficient government dictate that that data dissemination be as automated as possible, without the costly frictions of employees processing paperwork and money? To which the gatekeepers of closed data gravely utter the phrase "cost recovery".

Are You Recovering Costs Or Merely Acting Out an Accounting Fiction?

The conversion of paper records to digital data was a large expense: this was an era when databases and GIS programs required their own specialized hardware.. So naturally jurisdictions wanted to get some of that expense back by charging the public for derivative products—printed maps, as well as data files on floppies, CD-ROMs, etc. And with money changing hands comes paperwork and the legal licensing and disclaimers: in a word, overhead. The city of Denver stepped back and evaluated their data sales program, and concluded they weren't making real money because:

- a large number of data sales were to city contractors, who turned around and billed the city for their outlay;
- the license for the data was so restrictive that 3rd party usage options were severely limited;
- besides processing paperwork, employees spent significant time walking purchasers through the process of FTP download;

In short, if someone wants to play the “cost recovery” card as an economic rationale for selling government data let’s open up the books and get a full accounting of employee time spent processing paperwork, hand-holding over the phone, and determining if a contractor is simply adding another cost to the bill leaving the *net gain* from the sale at less than zero.

License-Free Publishing Without Vendor Baggage

Another important piece of the value puzzle is having effective, low-cost publishing platforms. A project with impressive traction is the open source publishing platform CKAN. Already used around the world by governments of all levels, the US recently announced that CKAN will be powering the next version of Data.gov, including Geo.Data.gov. Such a critical mass of usage means that the cycles of iterative innovation will be difficult for any commercial vendor to keep pace with. The City of Chicago is taking an even simpler tack by posting datasets on Github.

Github as A Potent Accelerant For Tools and Best Practices

Github? Isn’t that a website where dorks post programming code? Yes and no. Yes, code is posted, but that isn’t why some consider it a more potent social force than Facebook or LinkedIn. Clay Shirky lays out some far-reaching implications of the Git architecture that go way beyond being simply a code repository. Sure, large open source projects such as CKAN are freely accessible, but just as important are all manner of “glue” tools to build bridges for information out of proprietary systems. Here in Colorado we have the OpenColorado data catalog based on CKAN. But available too is a suite of tools is available on Github to automate the transfer of data out of the ESRI eco-system. It’s not just about lowering the sticker price of data publishing, but also lowering the level-of-effort so that it’s more cost-effective to push your information to the commons than any one-off custom in-house solution. With such an efficient platform for best-practices dissemination, small jurisdictions such as Scioto County, Ohio are no longer dependent on the technical savvy of their in-house personnel (or the mixed motives of Woolpert).

Do You Believe In Economic Multiplier Effects?

Of course you do, especially if you own a smart phone. Because in 2000, when the US turned off the “selective availability” that intentionally degraded GPS accuracy (yes, Al Gore was involved!) you and I did not have GPS-enabled phones. Think of all of the businesses and services that use GPS that didn’t exist in 2000? One consultant’s report estimates the *direct* economic impact of GPS in the US to be a cool \$67 billion per year.

The last refuge of the zealous closed-data bureaucrat reluctant to release datasets to the public is “what would they use it for anyway?” Parcels and foreclosure lists have obvious economic benefits, but fire hydrants? But given a critical mass of reliably updated datasets, who can predict what economically important insights will be gleaned in the near future, especially with the advent of the sensor web.

Consider the City of Denver’s experience: the 75% reduction in phone calls for assistance in purchasing and downloading data was predictable. But what has surprised the Open Data backers has been the tangible increase in *interdepartmental efficiency*. You know the drill: mid-level person in Department A emails mid-level person in Department B looking for data.

Department B person asks her boss; boss tells her to ask Department A what they want the data for, etc., etc., *ad nauseum*. Now communication is limited to a single email response, a hyperlink, and “have a nice day.” New York City, an early pioneer in the Open Data Movement, has a team dedicated to analyzing seemingly disparate datasets to tally everything from how many trees were lost to Hurricane Sandy (9,662) to which restaurants were clogging the sewers with illegally dumped cooking grease. It turns out that unclogging a city’s data flows actually unclogs its sewers: who knew?

With three of the United States’ five largest cities (New York, Chicago, and Philadelphia) making clear, credible commitments to Open Data, such everyday “victories” will become more commonplace. The Governor of New York recently launched a state-wide open data catalog along with an Executive Order to state agencies to publish data further nudging a “critical mass” of information that opens up opportunities for using data in creative, unforeseeable ways.

Given these large-scale commitments, and the technology platforms highlighted above, it’s increasingly obvious that in an environment of restricted government spending few jurisdictions will be able to continue down the economically irrational path of keeping information collected at the taxpayers’ expense walled off from the public.