Crooked Timber Open Data Seminar

Edited by Henry Farrell

June 07, 2012

Contents

Introduction

In May 2012, Tom Slee wrote two posts on the Open Data movement which got a lot of interesting argument going. To push the contradictions further, we've invited a number of people with differing perspectives to write short pieces on the theme of when and how, if ever, open data makes for better politics. Contributors are:

- Henry Farrell is an associate professor of political science and international relations at George Washington University. He blogs at Crooked Timber and The Monkey Cage
- Steven Berlin Johnson is the author of several books, including most prominently *Emergence*, *Where Good Ideas Come From*, and the forthcoming *Future Perfect: The Case for Progress in a Networked Age.*
- Tom Lee is the Director of Sunlight Labs at the Sunlight Foundation. He writes and works on open access issues.
- Beth Noveck is a professor at New York Law School, and former Deputy Chief Technology Officer at the White House). She is the author of *Wiki Politics* and is currently on leave, teaching at New York University and the MIT Media Lab.
- Clay Shirky is a professor teaching new media at New York University. He is the author of *Here Comes Everybody and Cognitive Surplus.*
- Tom Slee is a computer programmer, and author of *No-One Makes You Shop at Walmart.* He blogs at Whimsley.
- Victoria Stodden is an assistant professor of statistics at Columbia, and an advocate for data preservation.
- Aaron Swartz is the co-author of the RSS specification, a co-founder of Reddit, and general issue public intellectual.
- Matthew Yglesias writes *Slate*'s Moneybox column. He is the author of *The Rent Is Too Damn High* and *Heads in the Sand: How the Republicans Screw Up Foreign Policy and Foreign Policy Screws Up the Democrats.*

Tom Slee - Seeing Like a Geek

```
Yes, as through this world I've wandered
I've seen many men, I guess;
Some will rob you with a six gun,
And some with a GIS.
```

In the state of Tamil Nadu, near the town of Marakkanam, right next to a reserved forest, lies a contested plot of land. Records say these three acres belong to a member of the Mudaliar caste, but lower-caste Dalits living nearby claim the plot should be part of the reserved forest, which is not privately owned. The Dalits claim that the Mudaliars have pulled a fast one, using their influence in the local bureaucracy to fix the land records, and that older records will bear out the Dalit claim. Complicating the case, officials say that boundaries between land parcels in the area are often difficult to ascertain.[1]

According to Bhuvaneswari Raman, the Dalit claim was sideswiped by a Tamil Nadu government program to standardize, digitize and centralize land records. The program, promoted by the World Bank as a pro-poor, pro-transparency initiative, was undertaken to capitalize on the boom in nearby Chennai. The absence of clear land titles made extensive land purchases time consuming and expensive, and this was a bottleneck to large-scale development projects. As part of the program, the Tamil Nadu government declared that the digitized records would be the only evidence admissable in court for land claims, so the older records and less precise data that formed the basis of the Dalit claims lost any legal

[1]The case is discussed by Bhuvaneswari Raman, *The Rhetoric and Reality of Transparency*, Journal of Community Informatics, 8(2), 2012, available here.The case is discussed by Bhuvaneswari Raman, *The Rhetoric and Reality of Transparency*, Journal of Community Informatics, 8(2), 2012, available here.

footing they had, and their claim was sunk.[2]

A new generation of land developers grew up alongside the digitized records: firms with the skills and information to make efficient use of this new resource. These developers lobbied effectively for records and spatial data to be made open, and then used their advantages to displace smaller firms who, as Raman writes, "relied on their knowledge of local histories and relationships to assemble land for development". The effects went far beyond the three-acre plot near Marakkanan: newly visible master plans became used as "the reference point to label legal and illegal spaces and as a justification for evicting the poor from their economic and residential spaces." The "pro-poor" initiative turned out to be anything but. Tamil Nadu was not alone in running an open data project that made life harder for the poor; neighbouring Karnataka's "Bhoomi" (or 'land') e-governance program has had similar effects: a 2007 publication concluded that "the digitization of land records led to increased corruption, much more bribes and substantially increased time taken for land transactions. At another level, it facilitated very large players in the land markets to capture vast quantities of land at a time when Bangalore experiences a boom in the land market."[3]

The open data doppelgänger

Making data "open" has two effects:

1. By cutting the price of the data to zero, for everyone and for any purpose, it undermines the power of those who previously controlled access to it.

2. Just as cheap fish increases the demand for chips, so free data increases the demand for, and raises the value of, complementary resources and skills.

Effect 1 has many benefits, both real and potential. While all point out that open data is just one part of a complete breakfast, the essays by Victoria Stodden, Tom Lee and Matthew Yglesias in this seminar highlight the possibilities for improved accountability in

[2]Bhuvaneswari Raman, Solomon Benjamin and others have done extensive research around the impact of Karnataka state's "Bhoomi" (or 'land') e-governance program (web site) to digitize 20-million land records. See references listed in Raman above and particularly here (PDF).

[3]The quotations are from Raman, footnote 1.

government, those by Clay Shirky and Steven Berlin Johnson focus on the possibilities for improved services, and Beth Noveck emphasizes the possibilities for enhanced participation.[4]

But there is an inevitable flip side to open data, which is the rise of new markets in its complements (effect 2). The point of this post is to draw attention to this open government data doppelgänger—the shadow of commercial interests that follow civic hackers wherever they go; the new markets that spring up inevitably from the ruins of the old—and to its dangers. I am suspicious of this doppelgänger: more so than most open data proponents, who tend to use the language of entrepreneurship and innovation when discussing companies who work with open data, and who contrast the new firms with the aging business models they seek to replace, and they often present commercial use as a complement to civic use.[5]

The problem is, it's not *just* that new markets and new businesses replace old ones. The markets undermined by open data are generally traditional in structure, characterised by decreasing returns, with market power that is distributed and limited in scope. Before digitization, the property developers of Tamil Nadu had particular knowledge about land ownership patterns in a specific area and each used that knowledge to build their own little empire. In contrast, constant fixed costs and zero marginal costs are "the baseline case" for information goods,[6] so markets in open data environments are likely to consist of a few, big firms, each with significant market power. It's no surprise that the new generation of property developers in Tamil Nadu were larger than those they displaced.

The dynamic is familiar from other "open" movements and from previous price changes forced by digitization. A range of institutions have been overthrown (with much rhetoric about the stifling effect of "gatekeepers" and the democratizing nature of the Internet), only to be replaced by fewer, bigger institutions.

- The digitization of books undermined publishers and booksellers, and gave us a great big Amazonian bookseller/publisher.

- The digitization of video pulled the market from under the feet of Blockbuster and

[4]See Joshua Tauberer's self-published *Open Government Data*, available here, for a history by a proponent. See the home pages of Code for America, Sunlight Foundation, and mySociety for typical self-descriptions of the movement.

[5]A provocative example is *Government Data and the Invisible Hand*, by David G. Robinson, Harlan Yu, William P. Zeller, and Edward W. Felten, available here and the subject of some debate here.

[6]Hal R. Varian, Joseph Farrell, and Carl Shapiro, *The Economics of Information Technology: an Introduction*, Cambridge University Press 2004, p.3.

from independent video stores, and now we have Netflix.

- The mass sharing of digital music toppled major music labels, and saw the global rise of iTunes as the whole world's music store.

All this is, of course, very general, but the downsides of open data are real and need to be addressed. Describing them as paradoxical "unintended consequences" (see Tauberer p. 14) suggests they are anomalous edge cases, which misses the ubiquity of the problem.

Effective use: empowering the empowered

A small chorus of voices has been calling attention to the dangers of the open data's free-market doppelgänger, particularly in countries where the gap between rich and poor is large. Bhuvaneswari Raman, Solomon Benjamin and others' work (above) around land record digitization in India are one set of voices. Another is Michael Gurstein, a leading light in the field of "community informatics" who has been constructively raising concerns about how open data may "empower the empowered" for some time. The skills and resources needed to make "effective use" are complements to data.[7] As just one case, Gurstein quotes from a recent study of who uses the British mySociety TheyWorkForYou.com open government initiative:[8]

> "people above the age of 54 tend to be over-represented, while dangers younger than 45 are under-represented in comparison to the Internet population. In terms of demographics there is a strong male bias and a strong overrepresentation of people with a university degree that also translates into strong participation from high income groups... One in five users (21%) of the site has not been politically active within the last year"

Gurstein comments that:

> this attempt to enhance democratic participation has ended up providing an additional opportunity for those who already, because of their income, education, and overall conventional characteristics of higher status (age, gender etc.)

[7] See *Open Data: Empowering the empowered of effective data use for everyone?*, First Monday, 16(2), 2011, available here. Also a longer blog poston the same topic and other entries at his blog.

[8] Tobias Escher, *Analysis of users and usage for UK Citizens Online Democracy*, May 2011, available here (PDF).

> have the means to communicate with and influence politicians. The additional information and an additional communications channel thus has the effect of reinforcing patterns of opportunity that are already there rather than widening the base of participation and influence. (link)

Another dissenting voice is Kentaro Toyama, an expert in the use of information technology for development. He argues that "in contexts where literacy and social capital are unevenly distributed, technology tends to amplify inequalities rather than reduce them. An email account cannot make you more connected unless you have some existing social network to build on." Again, in thinking about the effects of new technologies we must look at the complements to the technology, and how those complements shape new markets.[9]

Seeing like a geek

Shunning the free-market doppelgänger can have a positive effect on outcomes.

Development studies scholar Kevin Donovan[10] sees similarities between open data efforts and the demands of the state as described in James Scott's "Seeing Like a State".[11] Open standards and structured, machine-readable data are key parts of the open data programme.[12] For Donovan this formalization and standardization is "far more value-laden than typically considered". Open data programmes, like the state, seek to "make society legible through simplification". Standardized data, like the state, "operate[s] over a multitude of communities and attempt[s] to eliminate cultural norms through standardization". He writes:

> Eliminating illegibilty in this way reduces the public's political autonomy because it enables powerful entities to act on a greater scale. Scott argued, 'A thoroughly legible society eliminates local monopolies of information and cre-

[9] See Toyama's talk *Ten myths about technology and development*, summarized here and available on YouTube here. Also see his contributions to the Boston Review forum Can Technology End Poverty? and his blog The ICT4D Jester.

[10] Kevin Donovan's excellent short paper *Seeing Like a Slum*, appears in the *Georgetown Journal of International Affairs*, 13(1), 97-104, 2012 and is available here.

[11] *Seeing Like a State* was discussed in these parts a few years ago here and here, as well as by Brad DeLong.

[12] See the *8 Principles of Open Government Data* spelled out in 2007 here. These principles have "become the de facto starting point for evaluating openness in government records".

> ates a kind of national transparency through the uniformity of codes, identities, statistics, regulations and measures. At the same time it is likely to create new positional advantages for those at the apex who have the knowledge and access to easily decipher the new state-created format'

Open data undermines the power of those who benefit from "the idiosyncracies and complexities of communities... Local residents [who] understand the complexity of their community due to prolonged exposure." The Bhoomi land records program is an example of this: it explicitly devalues informal knowledge of particular places and histories, making it legally irrelevant; in the brave new world of open data such knowledge is trumped by the ability to make effective queries of the "open" land records.[13] The valuing of technological facility over idiosyncratic and informal knowledge is baked right in to open data efforts.

More encouragingly, Donovan looks at how some "data geeks" recognized their own myopia in the Map Kibera project. The project started as a community-mapping project to trace the massive Nairobi slum. Some questioned the need for the project as "locals [already] knew their surroundings intimately", arguing that making mapping information available would more likely benefit external parties than the residents themselves.

The problems the project seeks to address (Kibera's poverty and marginalization) were of the class Donovan calls "wicked" problems: ill-defined, tangled, and resistant to technological fixes. However, "Although it began as an example of misdiagnosing a wicked problem... as a tame one (insufficient information availability), Map Kibera has admirably grown beyond a reductionist approach"; it has expanded to include other forms of activity such as citizen reporting, and has taken steps to ensure local ownership of the project. The project has moved beyond a technological goal to a set of social goals. Its list of sponsors, interestingly, includes only non-commercial organizations.

Donovan contrasts Map Kibera's evolution with that of commercial, and more narrowly technological mapping projects, such as Google's Map Maker initiatives which have been accused of unethical "exploitation of open communities." The danger of such projects is that, by eliminating the illegibility that privileges local knowledge over outsider knowledge, they may allow "more powerful entities to see like a slum" and benefit those already in power.[14]

[13]For examples of "Data-Driven Government" proponents see this talk by New York City CTO Rachel Sterne, The Data Driven City,and Citivox.

[14]Mikel Maron, *We Need to Stop Google's Exploitation of Open Communities*, blog post, 11 April 2011.

When it comes to development programs, Donovan concludes, making data available is not enough. Instead, transparency must be linked with deliberative development. Effecting social change cannot avoid the need to actually address underlying dynamics of power.

Have you considered the benefits of an alarm system?

Combining open data with its complements is a step on the road to surveillance.

One of the most valuable complements to open data is, of course, other data (mashups!): a bus schedule is more valuable if you can combine it with a map. This combinatorial aspect of open data raises problems for government-collected data, as legal scholars Teresa Scassa and Lisa M. Campbell highlighted recently, because data protection legislation "typically requires that information collected for specific purposes should not be used for other purposes without consent."

Scassa and Campbell look how "even relatively low quality spatial data may attract the application of data protection or privacy law, particularly when it is matched or combined with other data sets".[15] Take, for example, Ottawa Police's crime mapping tool (link), which is a map of calls for police assistance provided through a collaboration between Ottawa Police and US company Public Engines. If insurance companies make decisions about rates or insurability based on the crime-mapping data, or if security companies use it to target specific areas for marketing campaigns ("Did you know there were three robberies on your street in the last two months? Would you like a visit from one of our salespeople?") then this site could be violating those conditions.

Again, it is worth thinking about what knowledge increases in value and what is displaced when local data is made digitally public in this way. Brandon, Manitoba released property tax and assessment for every single property in town (here). For residents of Brandon, and particularly for local real-estate agents, this data release will not tell them a lot that is new. But now you don't need to know anything about Brandon–even where it is–to have a good idea of the wealth level of each inhabitant. Who cares about such stuff? The people who attend the Toronto Dx3 Canada event in January, for sure: "the first and only trade show

Available here. See also the controversy over an agreement between Google and the World Bank over the use of its Map Maker program over OpenStreetMap, here and here for example.

[15]See Teresa Scassa and Lisa M. Campbell, *Data Protection, Privacy and Spatial Data*, in R. Devillers & H. Goodchild, eds. Proceedings of the 6th International Symposium on Spatial Data Quality, Taylor & Francis, 2009, pp. 211-220, available here.

dedicated to Digital Marketing, Digital Advertising and Digital Retail, is offering attendees the chance to get intimate with the City of Toronto's Open Data Initiative."

Open data advocates commonly address privacy issues by reference to personally identifiable information, but there is no clear dividing line between data that identifies individuals and data that doesn't. It is well known that the right way to think of privacy when it comes to data made available in a "release and forget" manner (which open data is by definition) is in terms of information entropy or, to be less jargony, in a twenty-questions kind of way. Each question reveals a little more about the subject; no one question tells us what we need to know, but by successive filtering we arrive at the only possible answer.[16]

The commercial potential of combining open government data with other data sets is an irresistible temptation for the open data doppelgänger, regardless of the privacy consequences. There is a need for vigilance against its vulnerability to these temptations.

Reining in the doppelgänger

When I have brought up conflicts between markets and civic open data initiatives, I have occasionally been accused of cynicism and negativity (who me?) and exhorted to "get involved" instead. Many open data activists sees themselves as being idealistic and positive yet they retain a deep cynicism of government agencies while maintaining faith in the market's ability to maintain diversity and consumer power. I find it odd to see this combination of attitudes in a movement that often describes itself in egalitarian terms.

The faith in markets sometimes goes further among open data advocates. It's not just that open data *can* create new markets, there is a substantial portion of the push for open data that is *explicitly seeking to create new markets as an alternative to providing government services.* Influential advocate Tim O'Reilly claims not to be in favour of such an agenda (see comment here), but his "Government as Platform" initiative has been readily adopted by many who are.

In a recent paper, Jo Bates highlights the way in which open government data programs

[16] I wrote a summary of the jargon and issues here. A recent paper by Paul Ohm has prompted much thought: *Broken Promises of Privacy: Responding to the Surprising Failures of Anonymization*,University of Colorado Law Legal Studies Research Paper No. 09-12, available here (PDF). A response to Ohm by Ann Cavoukian and Khaled El Emam (PDF) emphasizes that the concept of "personally identifiable information" retains much value, but in conjunction with other limits on the use (and users) of data: they are concerned primarily with health researchers' ability to access rich data sources.

can be used as a form of privatization and deregulation: a deliberate attempt to create new markets in "Public Sector Information (PSI) reuse" instead of providing government services. Here is a summarizing quotation that I've used before:

> the current 'transparency agenda' [of the UK government, supported by prominent Open data advocates] should be recognised as an initiative that also aims to enable the marketisation of public services, and this is something that is not readily apparent to the general observer. Further, whilst democratic ends are claimed in the desire to enable 'the public' to hold 'the state' to account via these measures, there is an issue in utilising a dichotomy between the state and a notion of 'the public' which does not differentiate between citizens and commercial interests... The construction... encourages those attracted to civic engagement into an embrace of solidarity with profit seeking interests, distanced from the ever suspect notion of the state.[17]

Here is the kind of activity that now comes under "open data" initiatives (again from Jo Bates, here):

> [T]here has been significant lobbying by the financial industry to get better access to UK weather data so that it is able to compete in this [weather risk management] market. Groups such as the Lighthill Risk Network, of which Lloyds of London are a member, have lobbied government for better weather data so that they can develop risk based weather products. Similarly, the insurance industry has requested real time information on the pretext that they might respond more quickly to extreme weather events such as flooding. My own research and the recent announcement suggest that these demands have been met enthusiastically by well placed policy makers in national government who are keen to develop a UK weather derivatives market.

Weather risk management might seem like an odd duck, but Bates reports that "This weather risk management market far outweighs the USA's commercial weather products market which in 2000 was estimated at approximately $500 million a year", touching over $45 billion in 2005-06.

[17] Jo Bates, *"This is what modern deregulation looks like": co-optation and contestation in the shaping of the UK's Open Government Data initiative*, Journal of Community Informatics, 8(2), 2012, available here.

Welcoming corporate involvement in open data activities will lead to new Amazons and Apples, while undermining the community activism that is the movement's strong point. Whatever we think of Amazon and Apple from a consumer point of view, it is difficult to see how their rise has positive political outcomes.

A final example: one of the leading companies in the open data space is Palantir Technologies, highlighted by the civic-minded Code for America ("The success of Palantir or Socrata in offering innovative, web 2.0-style services for government shows the way forward for new government-focused enterprise companies." - here), a sponsor of O'Reilly's gov 2.0 summit (link) and adopter of "Government as a Platform" terminology, and an early partner of USAid's Food Security Open Data Challenge. And what do we know about Palantir? It is hooked closely in to US intelligence agencies, with early funding by the CIA through its http://en.wikipedia.org/wiki/In-Q-Tel venture capital arm and Peter Thiel's Founders Fund, both organizations known for their profound commitment to openness and equality. It is deeply involved in anti-terrorism programmes. Peter Thiel is Palantir's Chairman of the Board: perhaps he will be pursuing open data projects for the secretive Bilderberg Group, on whose Steering Committee he sits?

Are there ways to rein in open data's free-marker doppelgänger? The parallels between the economics of information goods and the economics of cultural goods can give us some ideas for dealing with the new oligopolies that threaten to grow around open data.

One lesson of cultural economics is that creative works for which there is significant demand in a small market can be swamped by near-zero-marginal cost exports from large markets. It is more profitable for TV stations in smaller markets to broadcast cheap American shows than it is to broadcast more expensive home-grown material, *even in cases where the latter would draw a bigger audience*, because cultural producers seek to cover their costs in their home market and are typically sell at discounted rates elsewhere.

To maintain cultural diversity in the face of winner-take-all markets, governments in smaller countries have designed a toolbox of interventions. The contents include production subsidies, broadcast quotas, spending rules, national ownership, and competition policy. In general, such measures have received support from those with a left-leaning outlook.[18]

Unfortunately the Open Data Movement demands that data be provided without borders and in a uniform way: machine processable, available to anyone, and license-free.[19] It

[18]See Peter S. Grant and Chris Wood, *Blockbusters and Trade Wars*, Douglas and McIntyre 2004.

[19]See the *8 Principles of Open Government Data* spelled out in 2007 here. These principles have "become

mandates non-discriminatory licensing, focuses on standards-based formats, and generally insists that data be accessible to rich and poor alike, like justice and the Ritz. It insists that any measures governments would like to take to favour—for example—non-commercial users or local users, be taken off the table. It strikes me as bizarre that this logic has gained such a significant hold among left-leaning digital enthusiasts that it has become orthodoxy.

I am not convinced that a coherent case can be made for "open data" as a public good, independent of the social changes that must accompany it, until the movement confronts its doppelgänger. This will require putting far more emphasis on experimentation in standards, licensing, and selective provision of data at the municipal and higher levels of government to ensure that what is a potentially valuable public resource is not plundered by those with the digital skills and resources to make most use of it.

---

the de facto starting point for evaluating openness in government records".See the *8 Principles of Open Government Data* spelled out in 2007 here. These principles have "become the de facto starting point for evaluating openness in government records".

Victoria Stodden - Signal and Noise: What Really Matters Is How Data Are Used In Decision Making

It's hard to argue with increased government transparency and accountability. Who wouldn't welcome a bulwark against opportunist backroom deals and increased incentives for rule-makers to think carefully about policy decisions? However, the link between these goals and open data isn't obvious and depends on what is being made available, and how it is being made available. I argue that what's actually useful is the reasoning process that underlies decision making, of which the data are just one part. A very real "open data" movement is occurring right now in the computational sciences largely independently of open government data. Scientists are just as affected as anyone else, or perhaps more so, by digital technology: they are making use of new computational tools and answering questions that weren't possible before our societal obsession for data collection. Not all computational scientists use the term "open data" to describe this movement, and rightly so. The complexity of scientific research in the digital age is rendering the traditional communication mechanism – the published scientific article – woefully inadequate. In order to make verifiable claims, scientists are finding they need to communicate more information to explain their work that can be included in the traditional published article, in particular the precise computational steps that were taken to generate the results. Efforts to do this are emerging at the grassroots level in myriad fields, from biostatistics to geophysics to applied math to medical

imaging.[20] [21]

Accessing the data is a key part of verifiability of scientific results, but it isn't enough. Scientists have the burden of convincing a skeptical audience they have done everything possible to root error out of their research, and so they need to explain (and justify) the steps from data collection to final results. This is nothing new, as reproducibility has been part of the scientific method since the 1660's, but now the reproducible research movement advocates the sharing the computer code as well as the data that generated the results in the published article. `http://www.RunMyCode.org` is a new effort to facilitate broad scientific understanding that I am involved in, for example.

An interesting question is whether any of the structure developed for scientific knowledge dissemination can be carried over to the political case and further aims of understanding how decisions came about i.e. transparency and accountability. There appears to be at least some superficial concordance between the scientific and governance goals: communicating in complete detail how outcomes were reached, and the deep importance of broad community buy-in. If we follow standards for scientific communication, provenance and justification become paramount. In the political context this might mean providing explanations for how new rules were arrived at, including the evidence used in reasoning. For this perhaps idealistic goal open data is clearly a crucial element.

There is a nice link between the movement toward reproducible computational science and the Obama administration's push for evidence based policy when that evidence is scientific in nature. For these types of government decisions, the reproducible scientific research movement can supply the data and reasoning behind the supporting scientific findings, such as those from climate science or public health, and fill in this part of the chain of reasoning. Evidence based policy evaluation by governments is another area where there seems to be a tighter link between the methods used in communicating scientific findings and the explanation behind government policy funding decisions - convincing skeptics of your evidence through transparency in the reasoning process.[22] Here, scientific practice could provide a useful framework to help guide principles of government communication.[23]

[20] See e.g. "Reproducible Research: Addressing the Need for Data and Code Sharing in Computational Science" `http://www.stanford.edu/vcs/papers/RoundtableDeclaration2010.pdf`

[21] Scientists for Reproducible Research - Google Group: `http://groups.google.com/group/reproducible-research`

[22] See `http://www.whitehouse.gov/omb/blog/09/06/08/BuildingRigorousEvidencetoDrivePolicy`

[23] The Obama administration is inserting itself directly (and usefully) in the Reproducible Research

Scientific publication is, in some sense, "linearizing a nonlinear process." Interest lies in understanding the steps necessary to replicate the results, not in following all the steps that actually took place during the discovery process. How the political decision making process gets "linearized," if it even does, is an open question. We'd like to know every influence on the outcome, but not the inconsequential ones. The fact that we don't have enough government data to implement this vision today doesn't bother me, as long as such data exist and they are making their way into the public sphere with a rapidity eclipsing that of Mickey Mouse entering the public domain, and a narrative can be constructed by decision makers that accurately captures how data were used in the political process. Data release seems to be steadily occurring, but a greater emphasis on communicating the reasoning made from the data could move the discussion toward transparency in government decision making.

movement in science. See `http://1.usa.gov/vNNFNT`.

Steven Berlin Johnson - Searching for John Snows

Sometime in the early 1840s, a British doctor and statistician named William Farr took control of the Weekly Returns Of Births And Deaths, a publication of the Registrar General's office where Farr worked. Variants of the Weekly Returns had been published by the state for at least two centuries before Farr took over, but for most of that time the Returns recorded only the name of the newly born or newly dead, and the parish where they resided. But Farr was what we would now call an Open Data advocate, and over time he greatly expanded the information disseminated through the Weekly Returns. By the mid 1850s, the Returns tracked age, cause of death, occupation–even the elevation of the dead's primary residence. (Farr believed that people living in higher altitudes had healthier lives.) Inspired by a debate with one of his contemporaries, the Soho doctor John Snow, Farr even added information on the deceased's regular source of drinking water.

I knew nothing about William Farr, or indeed the Weekly Returns, until I sat down to research my book The Ghost Map, which tells the story of John Snow's brilliant solution to the riddle of cholera, as it emerged in the middle of devastating outbreak in the summer of 1854. Snow is rightly famous for creating a map of the outbreak that helped convince authorities of his waterborne theory of cholera's origins. But it turned out that Snow was greatly assisted by the data that Farr had accumulated in the Weekly Returns. Indeed, it is an open question whether Snow would have been able to make his case to the authorities—and thus likely save hundreds of thousands of lives around the world—without the additional information he drew from Farr's dataset.

Farr's rationale for releasing that data is very much in sync with the argument for open data today. No, it was not a solution in and of itself, and its successes were unpredictable

(and often indirect, as in the case of Snow and cholera.) But Farr recognized—as I think many of us have come to understand in a contemporary context—that a much larger network of minds existed outside of government, outside the public health authorities, minds that might perceive patterns in the data that escaped the eyes of the authorities. John Snow happened to have one of those minds: a classic 19th-century amateur intellectual, pursuing the great mystery of cholera as a hobby while he kept his day job as a local doctor and anesthesiologist. Most of the official public health establishment had ignored his ideas about cholera, but Farr's data helped Snow change their minds in the end.

When I look at most of the Open Government initiatives today, I can't help but see them as a kind of search probe for all the John Snows out there across the country, unaffiliated with government, but willing and able to solve some small piece of the puzzle. This is good news for at least three reasons. The first is simple enough: we will have better ideas inside of government, and a sharper understanding of the problems that confront us, if more people are focused on those problems, even if that focus comes in their spare time, on someone else's payroll. This is a core principle in Henry and Cosma's notion of cognitive democracy; with the right tools, when we expand the density and diversity of minds engaged in solving problems, we get better solutions.

The second benefit is almost as direct. By creating platforms that encourage the John Snows of the world, we make more John Snows. In other words, we expand the ranks of the semi-pros and the hobbyists, the people who spend some part of their live trying to improve their government, beyond just voting every couple of years. The line that divides the politicians/bureaucrats from ordinary citizens becomes more porous. And as the class of part-time participants widens, it attracts more people who have other, equally useful, talents to share. The end result is more engagement, more civic participation, and an increased awareness of the services that states provide, and the challenges they face.

Open data has an additional benefit that is worth mentioning, given "the future of news" debates of recent years. I've argued elsewhere that there is a great deal to be optimistic about in terms of long-term journalistic trends, even if part of that story is the slow demise of Newspapers As We Know Them. But I think skeptics like Paul Starr are right to be worried about the fate of the investigative journalist, the city hall reporter that unearths corruption (or, on occasion, showcases civic achievements.) Open data can subtly help us avoid this bleak scenario—not by paying for investigative journalism directly, but rather by making it cheaper. When public data is actually public, the investigative side of being an investigative

journalist gets a lot easier, or at least it gets more easily crowdsourced by a large group of amateurs and hobbyists who want to help out. Yes, information abundance meant that the newspapers lost their local advertising monopolies to Craigslist and Groupon, but it also means that the crucial data they used to have to unearth by hanging around City Hall for months is now available to anyone with a Web browser or an API key. We may well have fewer investigative journalists on the payroll of newspapers, but if we play our Open Data cards right, we might well end up with more investigations.

Matthew Yglesias - Open Data Journalism

In the practical community of professional journalists writing about political events, the term "open data" is hardly ever in circulation. And yet, to those who are doing the best work it's an invaluable tool. David Simon succeeded in turning the idea that information age journalists need to learn to "do more with less" into a national joke, but the underlying concept makes perfect sense. The very same information technology revolution that's undermined the business models of traditional newspapers has done an enormous amount to increase the productivity of working journalists. Open data is an enormous part of that.

Especially for those of us who want to do informed commentary on economic issues, the FRED database and associated tools that the Federal Reserve Bank of St Louis has compiled is invaluable. Its companion set ALFRED that let's you compare different iterations of the same data series as agencies revise their estimates is, if anything, even more amazing. For example, it took me about fifteen minutes to throw together a chart comparing current GDP estimates for the critical Q3 2007–Q3 2009 period to those available to policymakers in 2009. Debates about the adequacy of the policy response to the recession should be informed by the reality that the economic shrinkage began a full quarter earlier than was contemporaneously known and that the decline during the winter of 2008-2009 was much more severe than people realized.

This basic National Income and Product Account data has always in some sense been available, but the internet and the determination of the FRED team have made it much more available than ever before. And it makes a difference, as FRED outputs are a regular feature on my blog, on Joe Weisenthal's policy writing at Business Insider, on Ezra Klein's Wonkblog, on Paul Krugman's blog for the New York Times, and wherever else on the

internet serious economic policy discussion is taking place.

In debates on the value of open data, some put what I think is undue weight on a distinction between commercial and civic activity that the case of journalism tends to undermine. The New York, clearly, is a commercial enterprise that's also primarily controlled by a founding family that sees it as serving some civic functions. Krugman, personally, is paid for his work but it beggars belief to imagine that he's driven by purely pecuniary motivations. And journalists of all kinds are dependent, on one level or another, on non-compensated contributions from quoted sources, experts used for background, or freely available data sources. An civic-minded person might want to write for or be quoted in a commercial publication precisely because the engine of commerce is a powerful motive to widely disseminate information.

The fundamental issue is that as the marginal cost of transmitting information falls ever closer to zero, two things happen simultaneously. One is that it becomes increasingly difficult to internalize the value of information-production because the facts (or "facts") once unleashed into the world tend to spread beyond the control of the producer. The second is that for that very same reason, information becomes more socially valuable. Governments are ideally situation to serve as producers of these goods. In the U.S. debate this is widely acknowledged in the special case of basic scientific research, where there's a strong bien pensant consensus that subsidies are socially and economically valuable. But the issue has nothing in particular to do with science. As the marginal cost of information distribution falls, market systems increasingly fail to produce it at an optimal level. Governments should step in wherever it seems feasible to do so. The push for "open data" is best viewed as, like scientific research, a particular case of this general principle.

Whether this will actually lead to better politics in the end has more than a little to do with the question of to what extent political decisions are actually driven by information. I'm somewhat skeptical on this score that they are. But even if they aren't, all you need to believe is that some important decisions of some sort are driven by information to conclude that more production and more open dissemination of data of all kinds is of enormous potential value to society.

Clay Shirky - Cooperation and Corruption

tl;dr The Open Data movement is good at improving service, but bad at rooting out corruption

Tom Slee has done us a favor by kicking off a conversation about the values, goals, and coherence of the Open Data movement. I share his sense that the movement has been a disappointment to date. However, as my principles differ from his, my sense of disappointment, and of what to do about it, differ as well. Before I get to that, I want to position myself relative to Slee's three summary assertions about the Open Data movement. (The points are Slee's; the reactions mine.)

1. It's not a movement in a political or cultural sense of the word.

   > I think Slee has this one wrong. In particular, two of his rationales – the Open Data movement has no political goals, and what goals it does have are too variable to cohere – seem to me to be willful attempts to deny use of a word he likes to a movement he doesn't. Slee commits a sin he accuses the Open Data people of, namely over-focusing on technology and under-focusing on the political aspects of the work. The people improving bus schedules and the people uncovering graft may differ in their aims, but they share core values: they want to reconfigure the relationship between government and citizens concerning what the government knows and citizens don't. This is an inherently political goal.

2. It's doing nothing for transparency and accountability in government.

> This is trivially wrong – it is plainly doing *something*, as Slee later notes – but his formulation in the body of the essay is more interesting: the net effect of transparency and accountability could end up being negative. I'll agree with this assertion (though for somewhat different reasons than Slee), and spend the bulk of the essay on it. (I'll also concentrate on the US case; I admire the work of Canadian participants in the Open Data movement, especially (David Eaves)[http://eaves.ca], but I don't know enough about the Harper Government to make my comments useful.)

3. It's co-opting the language of progressive change in pursuit of a small-government-focused subsidy for industry.

> This is partly true, in that the Open Data movement does not strongly distinguish between for-profit and non-profit use. Slee's use of 'co-opting' reflects his disapproval of commercial re-use; people who approve of the private sector creating new services with government data would use different language. Slee clearly regards the Commercial Service Delivery quadrant of his map as Mordor; this is where I disagree with him most strongly.
>
> I've gotten extraordinary value out of commercial services like Google Maps and Weather Underground, value I don't think the government could deliver as well. Furthermore, open access to this data limits pre-open-data monopolies of the sort enjoyed by AccuWeather or Westlaw, an improvement pursued most aggressively by Carl Malamud, our Living National Treasure of open data since before the movement had a name. I'll adopt the observation made by Tom Lee (not Slee) as my own: "I think it's flatly wrong to consider private actors' interest in public data to be uniformly problematic."

With that out of the way, I'll say that for me, the Open Data movement has been a net disappointment. In the middle of the last decade, I attended a meeting of the then-nascent movement. We gathered in a loft filled with techies and journalists and good government people, all looking for common ground. It was like a tent revival, so infections was the excitement. The job at hand, or so it then seemed, was to fit every government database with an API (that magical acronym!), whereupon bus schedules would appear on our phones and corrupt politicians would be driven from office.

We got the APIs. We got the bus schedules. The politicians, however, have yet to lose much sleep over open data.

There are several possible explanations for this. Here are some I am explicitly rejecting: I don't believe corruption in the US is rare. I don't believe it's expertly hidden; Bethany Mclean helped doom Enron by reading their financial statements carefully. I don't believe action is impossible; when ProPublica and the LA Times exposed the incompetence of the California nursing board, the Governor fired that entire board the next day.

Instead, I believe the broad failure of the Open Data movement to root out much corruption is tied to organizational failure, or rather failures. So, following Slee, I'll offer three observations of my own.

1. The institutions that are good with data tend to be bad at story telling.

   People don't consume facts. They consume stories. People who understand the importance of data are generally the people most enthusiastic about interpreting it; as a result, we systematically overestimate how general citizen interest in data actually is. (The best expression of this gap remains Tom Steinberg's Asking the wrong question about Data.gov.

   People choose proxies for understanding complex issues, not because they are lazy, but because we can't not. They look, for example, for assertions that global climate change is or is not real, rather than searching out charts of temperature charts or maps of sea level. For civil liberties activists and data journalists to have even a fraction of the effect they intend, they will have to set aside the fantasy that telling the truth is enough. They will have to get good at telling true stories, or get good at partnering with organizations that are good at telling those stories.

   Which brings me to the second failure.

2. Institutions that are good at story telling tend to be bad with data.

   News organizations are paid to tell true stories. Unfortunately, much of this story telling is uninformed by the kind of numeracy that would be required to take advantage of even the simplest open data.

   One tiny but illustrative example – newspaper articles often feature statements about income distribution like these:

> Sunnyvale "boasts an average family income of 123,647."
>
> Two companies opening warehouses "are expected to employ 1,000 workers, with an average employee's annual salary of 37,000."

You would not guess, reading these articles, that a sizable majority of families in Sunnyvale make less than 123,000 a year, or that the salaries of most workers in the new Bethel, PA warehouses will be less than 37,000.

Journalists routinely underreport degrees of financial inequality, because they routinely treat averages as if they represented something the normal participant in the system would see. (It's like that old joke: Bill Gates walks into a bar and everyone inside becomes a millionaire, on average.) Newspaper style guides, to their credit, clearly define what averages are supposed to mean; editors, to their detriment (and ours) simply do not enforce the correct use of even this basic mathematical concept.

The Open Data movement often puts forward visions of sophisticated, interactive uses of data that provides citizens with valuable insights. This does sometimes happen, as with Dollars for Docs or the visualization of gay rights state by state. But these are rare cases; as appealing as it is to imagine a press corps that exposes new truths by interpreting new data, the normal case is that they do not even correctly express existing truths with existing data.

3. Transparency is often mere translucency.

If all that were going on was a cultural misfit between people who understand data and people who understand narrative, we could improve things with a few kumbaya meetings. The third great obstacle, though, is that powerful actors do not want transparency. I agree with Slee's observation that "A government can simultaneously be the most secretive... in recent memory and be welcomed into the club of "open government." Slee talks about the decision by the Canadian government to abandon StatsCan, a decision similar to Republican attempts to reduce the effectiveness of our census. The problem is far broader, however.

As Wendy Wagner put it in her 2010 paper Administrative Law, Filter Failure, and Information Capture

> [E]very successful reform movement has its unintended consequences. What few administrative architects anticipated from the new commitment to "sunlight"

> was that a dense cloud of detailed, technical, and voluminous information would move in to obscure the benefits of transparency.

When we focus on how much data is made available, we create a world where powerful actors can live up to their nominal commitment to openness, while in practice reducing the utility of that data, by making data hard to understand or use, making it inconsistent over time, or producing high volumes of low-quality data while holding back low volumes of high quality data. In extreme cases, as with StatScan, a government can decide not to know certain things, rather than be forced to share that knowledge with citizens.

Now if this were just a technical issue, where laws needed to be written with cleaner specifications, solving these problems might be easy. But the deep problem is this – service delivery involves shared effort between public and private actors, while transparency must be oppositional to be valuable.

The distinction between service and transparency is a distinction between partnership and opposition. For me, this tension, far more than the commercial v.s non-commercial split that so exercises Slee, is the largest problem embedded in the current form and biases of the movement. I'm afraid the success of service delivery, wonderful as it is, has convinced many governments that they can make citizens happy by sharing useful data useful, while preserving secrecy in the very areas where political discipline matters most. The House Appropriations Committee has recently proposed cutting off bulk access to legislative data. If this proposal succeeds, then the Federal Government will both release more data in 2013 than in 2012 *and* the actions of our elected representatives will become even harder to oversee.

The likeliest scenario for the service/transparency coalition is that Lee's distinction between open-as-in-data.gov and open-as-in-FOIA remains unresolved, with the transparency movement being relegated to second-class citizenship in the Open Data movement. Another scenario is that the movements split – Code for America, See, Click, Fix, and all the other groups trying to make government data more useful will come to define themselves less around the access to data and more around patterns of use and re-use. This could be salutary for the transparency movement, as the necessarily oppositional character of their efforts would become clearer, though they would also become harder to pursue, in part because of that clarity.

The third possibility, though (and for me, the best argument for the Open Data movement

as a movement) is that the two halves of the movement make common cause. This would entail the service people saying to politicians "In order to take more credit for making the public's life better, you also need to be more transparent about your own behavior." This probably won't happen – it will be hard for service-oriented groups to apply this kind of pressure without having it backfire – but it would create a better bundle than we have today.

Long after the last pothole has been seen, clicked, and fixed, I think the legacy of the Open Data movement is going to be assessed on its ability to limit powerful actors. I'm afraid that that legacy will be minimal, in part because I'm afraid the transparency people have brought a knife to a gun fight. It's possible for private actors to make common cause with elected politicians and career civil servants around snow removal, but useful transparency will always require harder tactics, tactics especially including ways of using open data to tell stories that enrage the public.

Aaron Swartz - A Database of Folly

The open data movement is a hammer which has gathered the support of many nails. There are the curious taxpayers, who feel their annual checks mean they deserve a peek at the interesting facts the government has collected. There are the ambitious business owners, who see an opportunity to privatize profits from work with socialized costs. And there are the self-styled activists, who believe that if we reveal the data on what the government is really doing, we will arrest corruption by exposing it to sunlight.

The coalition is a confusing mix of these very different motivations (as Tom Slee observes), and the benefits of such a tactical alliance has come with the cost of some confusion. So let's be clear about what open data can and cannot do.

If the St. Louis Fed publishes reams of economic data, it can certainly make it easier for Mr. Yglesias to make his fantastic charts. If the MTA makes real-time subway information public, it can certainly let Mr. Ernst improve his fantastic app. And, as the talented Mr. Lee pointed out to me, his careful collection of data about members of Congress and the bills they'e passing can be an invaluable resource for professional activists.

So, if I got to choose whether the government should share the data it's collected, I'd happily vote yes. In fact, I spent several years of my life using the FOIA laws to force it to do just that. I can't claim my work had any particular impact, but as a curious taxpayer, it was a weirdly-enjoyable hobby.

But the open data movement often claims to be much more than that. They insist open data will not just help a few people with their jobs or a few kids with their hobbies but, as the Sunlight Foundation puts it, "make government transparent and accountable." And that I just don't see.

I've outlined my theory why elsewhere, but the short version is pretty simple: people hide their crimes. Imagine you learn lots of bribes are exchanged at top of the Capitol

Reflecting Pool. So you lobby Congress hard to set up bright lights and a camera to catch the perpretrators. The video would be livestreamed to the Internet so dedicated watchdogs can name and shame the bribetaking politicians. Your lobbying succeeds and, on January 1st, the lights go up and the cameras switch on.

But as an engaged citizenry tunes in, there's is nothing but disappointment. Nobody seems to be taking bribes; just a couple pieces of litter blowing by the pool.

Was Congress really squeaky-clean after all? Of course not – the bribes just moved to the other end at the pool, out of the spotlights.

When you have time to prepare, it's pretty easy to disguise the data. And this is exactly the pattern we've seen. It's always been investigative journalism, not data mining, that's revealed the big scandals about politicians. I, more than anyone, would love to believe that the next great Watergate is just lying in plain sight to be uncovered by a swashbuckling econometrician, but the sad fact is, it simply isn't so.

But it's also worth pausing to ask: what was any of this *supposed* to achieve? Imagine, for some strange reason, members of Congress didn't bother avoiding the spotlight. Every day, we saw them, in full HD video, taking money from prominent businessmen. Do we really think even this (far-fetched) instance of transparency would change much? After all, most Americans already think Congress is corrupt. Most Americans think money actually buys politicians' votes. Seeing it happen in video might be striking, and maybe make for some good segments on the evening news (or, these days, some viral YouTube videos), but would it really change anything?

After a couple weeks of chatter, and perhaps a few grandstanding legislative proposals, I suspect it'd just fade into the background. More dramatic examples are not exactly what's most missing from the reform debate – Lessig's recent book has enough to last us a couple decades. Structural reforms have failed because of the incompetence of reformers, not because there's a lack of evidence that there's a problem. (Free tip to structural reformers: get state legislators to sign on to your constitutional amendment. They're very susceptible to public pressure, there's a lot of them (so you'll have a constant narrative of progress), and they're the ones you'll ultimately need to actually pass the amendment.)

But maybe open data was supposed to improve politics in other ways. Structural reform is an ambitious goal – maybe the open data proponents wanted something much more modest. But all the more modest stories suffer from a similar excess of naïveté. Whenever geeks turn their eyes to politics, they always have the same reaction: There's so much inefficiency! And

they naturally propose the obvious ideas for reducing it – for example: If only it was easier for citizens to read bills, citizens with relevant expertise could assist Congress by sharing their hard-earned wisdom!

The fact is, Congress isn't interested in availing itself of your wisdom any more than the sausagemaker needs your help tidying the floor. Lawmaking is *The Wire*, not *Schoolhouse Rock*. It's about blood and war and power, not evidence and argument and policy. (I have one friend who was startled to learn that when members of Congress debate an issue on C-SPAN, they're speaking not to each other but to cameras in a largely-empty room.)

I don't want this to sound overly harsh. The truth is, it's really hard to do effective philanthropy. With a little work, you could mount a similar critique of the vast majority of our bumbling efforts to do good. Most ideas for helping people that seem reasonable in the abstract, turn out to fall apart upon close confrontation with reality. The real question is what happens then. There's no shame in admitting your mistakes, learning from them, and trying again. Indeed, as my old professor Carol Dweck has shown, that's the only real route to success. But most of us are too vain or too proud to take that route. We insist that the purity of our intentions reduces the need for careful scrutiny of our effects. Or we try to make ourselves feel better by grasping at any factoid that suggests we had an impact.

I have no particular interest in correcting people's pride or vanity. This movement is populated by my friends and I respect them enormously and wish them well. Throwing darts at their day jobs has only made my life worse. But this stuff matters – funders and volunteers face tough choices about which causes to pursue. It's important that they know the case for opening up data to hold government accountable simply isn't there. (And that they should invest in metaresearch, inxluding open scientific data, instead.) It's nothing personal – just trying to help everyone do their best. I dearly hope that if anyone ever has a similar critique of the causes I pursue, they will be even more blunt in pointing out my folly.

Henry Farrell - Trish, Reiner and The Politics of Open Data

Ongoing debates over open data remind me of Cory Doctorow's short story, Human Readable, which depicts attitudes to technology through a disagreement between lovers. Reiner is a classic hacker - he thinks of the world in terms of technological fixes for technological problems, and has difficulty in believing that the algorithms can be systematically skewed. Trish is a classic lefty, who thinks of the world in terms of power relations, and, specifically, in terms of how smart powerful people figure out ways to gimmick the system so that it works to their advantage. They don't understand each other very well but end up, sort of, figuring out a way to cooperate. Clearly, real life people have more complicated views than Doctorow's characters - even so, he's put his finger on an important tension. When I went to Foo Camp a few years ago, I found it incredibly intellectually exhilarating - meeting a bunch of incredibly interesting people who were both (a) smarter than me, and (b) intensely practical, interested in figuring out how to do stuff rather than study it. But I was also a bit nonplussed by how enthusiastically many of them believed that the Obama administration was going to usher in a new era of Big Data led technocracy. A lot of them (not all of them) didn't seem to have any very good idea of how politics actually worked. They were mostly Reiner, without much admixture of Trish.

Of course, there are plenty of Trishes too. Debates over open data, like many other debates over technology and politics, have calcified around a Reiner-Trish confrontation - techno-utopian naivete versus politics-led skepticism. This is *not* to say that the people on the one side of the argument are Reiners, and the others are Trishes. The smarter people on both sides of this argument have more interesting and complicated understandings. But it is to say that the Reiner-pole and the Trish-pole are the attractors - it's harder than it

should be for people to escape the gravitational pull of the one or the other position for very long. Furthermore, even when Reiners try to become Trishes, they can retain a rump utopianism. Larry Lessig, driven into shrill, unholy Trishdom by the horrors of the American political system, still quietly yearns for technocracy. His book on American corruption is an excellent and compelling indictment of our current system, but he tends to conflate political corruption with partisanship. Here, he is plausibly influenced not only by his earlier work on technology, but by a broader tradition of American progressivism, which is (as Nancy Rosenblum has documented) inherently suspicious of partisan contention. Actually existing Reinerism blends (a) a belief in the social benefits of technology, with (b) a technocratic understanding of politics. Actually existing Trishism blends (a) the belief that the social benefits of technology are limited, with (b) a power-based understanding of politics. Again - Reinerism and Trishism by no means necessarily describe the *actual beliefs* of people on the one or the other side in these arguments. Instead, they describe the *attractors structuring debate.*

This annoys me, not least because it's highly inconvenient for my own politics. Like Reiner, I'm broadly optimistic about the social benefits of technology. But like Trish, I'm not a technocrat, and believe that politics is necessarily a struggle between factions with different wants and interests. More precisely - I argue that information technology can provide extraordinary benefits - but only under the right political configuration. Cosma Shalizi and I have been trying to articulate the case for what we call cognitive democracy- political arrangements which recognize the irreducible diversity of individuals' interests and perspectives, and try to take advantage of them. Briefly, we argue that democracy, to work well, has to (a) minimize power disparities and (b)retain and harness cognitive diversity. Technocracy will not do this particularly well - what one wants is not bland consensus, but vigorous political contention, in which different factions struggle and engage with each other, likely never reaching agreement, but learning from each other, and setting out clear, alternative approaches to dealing with collective issues.

Here, open data can have three crucial benefits. First - it helps limit power disparities. Lobbyists' main advantage is often less their selective control of funding than their selective control of information. Making politically-relevant data available can (with caveats: see below) make it easier and cheaper for interests that are under-represented to make better and more compelling arguments for their perspective. Second, open data, where it is high quality, can help limit the tendency of factions to make up their own information as well as

their own interpretations of that information, hence improving democratic politics (diversity of opinion and understanding is not cognitively helpful when it is unmoored from reality). To be clear - neither of these is a cure-all. Even under the most optimistic assumptions, open data cannot correct for various economic, organizational and political disparities of power. As debates over global warming demonstrate, *no* data will be sufficient to convince a group that has dived into the deep end of crazy. Yet open data can help (and would help even more, if it were combined with structures such as, in the US case, a reborn Office of Technology Assessment.

Third, opening up specific *kinds* of data - data about what Cosma calls processes of collective cognition could help foster a general democratic experimentalism. Cosma and I don't think of the Internet as either a force for general emancipation or another instance of commercial and political cooptation. Instead, we think of it as something like an abandoned mad scientist's laboratory, in which various experiments in cognitive processing have been left to fizz and overflow together. Some of these experiments are turning into monsters, others unviable chimeras, others yet interesting hybrids. Figuring out why different experiments had different outcomes is going to be a chancy process - but nonetheless can provide highly valuable data on when processes of collective information processing and decision making work, and when they don't. The problem is, of course, that most of the really interesting data is not readily available for commercial and political reasons.

If this understanding of politics is right, open data has great promise. However, it also has clear limitations. First and most obviously, it's going to work best under different democratic arrangements than those we have in the US and other democracies, which are becoming increasingly sclerotic thanks to inequalities of power. Open data doesn't create its own politics - instead it requires sweeping politic reforms, which it can't plausibly generate itself, in order to achieve its full benefits. Second, in the actually-existing-democracies of today, the 'if you build it, they will come' attitude of some of the breezier open data proponents is badly misplaced. Open data is likely only to be taken up to the extent that it is *directly useful* to the agendas of *existing movements, factions and organizations.* Politics is about struggle between factions. This isn't likely to change, and arguably it *shouldn't* change. Hence, data will be politically relevant only when it's relevant to the political goals of particular interests, or, perhaps, to newspapers if there's a particularly juicy scandal.

Of course, this view of politics may be profoundly or subtly mistaken (it's a work in progress - we make no guarantees). But at the least, what it does is to open up the debate

a little. If we try not to conflate our views of technology with our views of politics, in the ways that Reiner and Trish do, we may be able to think more clearly about how they relate to each other. Specifically, we should be able to think better about how different forms of politics interact with different regimes of data access. There are a whole variety of possibilities, which the current set-piece battle blinds us to. It would be nice to move on from it.

Beth Noveck - Open Data: The Democratic Imperative

Open Data are the basis for government innovation. This isn't because open data make government more transparent or accountable. Like Tom Slee, I have serious doubts about whether it does either of those things. In any event, shining a light on the misdeeds of ineffective institutions isn't as imperative as redesigning how they work. Instead, open data can provide the raw material to convene informed conversations inside and outside institutions about what's broken and the empirical foundation for developing solutions together.

The ability of third parties to participate is what makes open data truly transformative. The organization that collects and maintains information is not always in the exclusive position to use it well. For example, US regulators have compiled hospital infection rates for a long time. Accessible only to government professionals, they had limited resources to make adequate use of the information. When HHS made the data publicly available by publishing the data online in a computable format, then Microsoft and Google were able to mash up that information with mapping data to create search engines that allow anyone – from the investigative journalist to the parent of the sick child – to decide which hospital to choose (or whether it is safer to stay home). When data are open – namely legally and technically accessible and capable of being machine processed – those with technical know how can create sophisticated and useful tools, visualizations, models and analysis as well as spot mistakes or mix and mash across datasets to yield insights. As Matt Parker, put it: "By making data open, you enable others to bring fresh perspectives, insights, and additional resources to your data, and that's when it can become really valuable."

Complex Democracy

Solving complex challenges requires many people with diverse skills and talents working together. In modern society, we weave our collective expertise together, enabling us to make complex products such as cars and computers that we cannot make alone. The more complex and diverse the products, the more successful – measured both in terms of wealth and well-being – the society over time.

Educating our young or curing cancer are the cars and computers of governance. They are complex social problems that require us to bring our diverse talents to bear. But our centralized institutions of government do not adequately leverage our collective knowledge to improve governance and solve problems. We can't foster complexity if we limit public participation to voting in annual elections or commenting on already written rules. There's no excuse for failing to take advantage of people's talents, abilities and desire to play a role in governing ourselves and our own communities.

Hackathons as a Model for Engagement

Open data create obvious new ways for geeky citizens to play a role in governance. All over the world, local transportation authorities are making schedules available for free and then inviting tech savvy citizens – civic coders – to create iPhone apps that tell commuters when their bus or train is coming. There's obvious value to the public as well as to institutions from having better data to inform planning, policymaking and the expenditure of resources. But what's exciting about mashathons, hackathons, data dives and datapaloozas (a Todd Park favorite term) is that these are intelligible models for taking action.

Wikipedia works because we know what tasks are required of us to write an encyclopedia entry. Only the high priesthood of government professionals knows how to write a law, craft a policy, draft procurement RFPs, or appropriate funds. Hackathons aren't the only model for participatory governance but they are one way for us to get involved that showcases how it might be possible to move away from centralized to distributed action.

Making government more participatory wouldn't have worked as well if we had only focused on releasing data-as-in-FOIA about the workings of government – politicians' tax returns, who-met-with-whom, and even spending data. By defining [PDF] High Value Data to include information: "to increase agency accountability and responsiveness; improve

public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation," the hope was to speak to more people's interests, talents and abilities. We took a lot of flak at the time from those with passion for specific kinds of data. I have written previously that the "open" in open gov was never meant to suggest data-as-in-FOIA but, rather, meant open as in open innovation and therefore always had to go beyond "civil liberties data" to include all the information that government collects as well as information that citizens might crowdsource and provide to make government smarter.

The Hard Work of Opening Data

Moving toward open innovation as a default way of working in government is not easy. It takes a religious fervor (hence the sense of movement) for those who want to open up data. It requires doing the hard and costly work of persuading data owners to shift from paper to digital and machine-readable formats and then to release that data despite political and technical challenges. But to foster engagement also requires curating the guest list for the hackathons to get subject matter experts, stakeholders, data geeks, activists, designers, computer scientists, data junkies and entrepreneurs together.

The host of a good dinner party doesn't just leave the guests to fend for themselves. He introduces people, points out what they might have in common and seeds the conversation. Transit camps have been so successful because the conversation starts itself. Everyone wants to know when their bus is coming. But give people a data set about freight routes for transshipping goods or Form 990 tax returns and some explanation might be required.

Creating a participatory innovation ecosystem is about a lot more than just publishing data sets. It requires doing the hosting, convening, persuading, and demonstrating involved in inviting diverse people to participate. The institutional players have to be prepared to collaborate with the innovators; those outside government have to know how to collaborate; civil society activists have to ensure that innovators know the problems that need solving; and research is needed to figure out what works.

Using Data to Re-Regulate

The curatorial function is about coming up with strategies for using data to develop innovative solutions to protect consumers and serve the public interest. If we merely throw data over the transom, entrepreneurs, especially large ones, are likely to be the only entities with the wherewithal to do anything with the raw information.

But when we focus on data as a means to the end of bringing people with diverse skills together to solve problems then open data can improve upon the blunt instrument of regulation enforced by litigation.

With open data (also called Smart Disclosure), the US government is experimenting with using light touch regulation combined with technical innovation (and a firm belief in behavioral economics) to create consumer decision tools. For example, the Department of Transportation enacted a rule to require airlines to make all their fees and charges transparent. Because the data is open, innovators can create new visualizations to help consumers understand the costs and make informed decisions. No Child Left Behind requires states to gather and report school performance data, which is now being used by GreatSchools.org (in cooperation with the Department of Education) to help parents choose between public schools. The tool is in use 40-50% percent of all K-12 households. The White House Open Educational Data Initiative is spurring university Presidents to provide data voluntarily to help students and parents compare college costs and college aid "so they can make more informed decisions about where to enroll."

But until we stop talking about data and start talking about complex and collaborative governance, we will fail to appreciate how open data can protect consumers, lessen the burdens on entrepreneurs and catalyze more effective institutions.

Tom Lee - Open Data: Better Politics, Winning Politics... But Still Politics

My alma mater had a celebrity professor of political science who was principally known for two things. First, for accidentally leaving his wireless mic on during mid-lecture restroom breaks. And second, for the slogan "Politics is a good thing!" which he relentlessly promoted via mediums as diverse as lectures, TV appearances and TA's t-shirts.

Well, we all make mistakes. But only political science professors seem to make that second kind of mistake. This glib celebration of a maximally vague conception of politics always rankled, conflating, it seemed to me, everything from a heartfelt PTA meeting speech to Caesar bleeding to death on the Senate floor. I never liked that class, and "politics" still often seems unmanageably broad to me. Pondering what open data has to do with "good" or "better" politics, I find that adding a qualifier only leaves me more confused.

Still, I suspect that my professor and I would have been able to agree on at least a few concrete things. A free press; universal suffrage; a public education system; the secret ballot: these are unobjectionable, broadly agreeable foundations of democracy. Open data should perhaps be the newest addition to that list. In one sense, open data is pre-political.

Whatever parts of our political system you happen to value, unencumbered government data almost certainly plays a role in their support. Knocking on doors to get out the vote? That's made possible by Census TIGER/Line map data and voter roll information. Wielding facts and figures in the poli/econ blogosphere? The open data policies of BLS, the Fed, CBO and other institutions power these debates. Even the tedious daily point-scoring of cable news is enabled, in part, by video, audio and text material provided by various publications and outlets of the House and Senate. Whether or not you consider these mechanisms constitutive of "good politics"–in the sense of representing a productive

and positive kind of deliberation–the fact that they are possible seems like an undeniably good thing.

These examples embrace a broad conception of what "open data" means. More specific definitions exist: the Open Knowledge Foundation has a good one; at Sunlight we tend to gesture toward a set of more prescriptive and explicitly government-focused principles. But across definitions, the broad outlines are the same: open data is digital information that is unencumbered by fees, credentialing, licenses and other unnecessary limitations on its use and distribution.

The rationale behind offering such resources is straightforward. Copies of a given piece of digital information have no marginal cost; supply is infinite. By embracing that limitlessness we can lower the costs facing potential users of the information, spurring more use. Who knows what benefits might result? More original research; more technology startups; more and better public interest advocacy–these are all plausible benefits to open data. In addition to this pragmatic calculus, one can argue that citizens should be extended a right to free access to information by and about their government (this line of thought is probably best embodied by the work of Carl Malamud). Both justifications are plausible, inspiringly egalitarian, and perhaps a bit hand-wavingly utopian. In my experience, those are all useful attributes: people find them appealing. So in this sense, too–the sense of being a winning and uncontroversial issue–open data is "good politics." Certainly this has been our experience at the Sunlight Foundation, where we have successfully attracted support for open data policies from legislators and citizens representing a wide variety of ideological perspectives.

Of course, someone is going to have to pay for the collection, organization and distribution of all this data. Government might already be doing some of those things in order to fulfill its other responsibilities, but there are typically additional costs that have to be born somehow. At present our public sector data is paid for through a mix of general funds and user fees. The above account–that unlimited supply allows for potentially vast surpluses–argues for socializing those costs, and in the past Sunlight has found itself arguing to that effect.

And here, I suppose, is where open data becomes a bit more controversial than apple pie and miniature American flags. Tom Slee started this conversation on a skeptical note, proposing that open data is little more than a catchy brand name that's being used to justify new IT expenditures. Many open data advocates, myself included, rejected this idea: open data, we said, is the rare partnership where corporate interests and the public good are

well-aligned. This produced some understandable eye-rolling from Slee and others. And certainly articles with titles like "How To Cash In On Government As A Platform" have done little to quell concerns that so-called civic hackers' talk of public service is just so much first-date patter as they greedily eye the public's assets.

In my experience, this concern is unjustified: most open data advocates I know–including the author of that Techcrunch article–have a genuine interest in doing work for the public good, even if it means a pay cut. To the extent that these individuals embrace the rapacious language of Silicon Valley startup culture, it's usually to make the cause more palatable and interesting to their fellow coders. But take my anecdotal experience for what it's worth. In a world where the Department of Defense casually announces they have a couple of spare Hubble Telescopes lying around, I think it's difficult to make the case that the our government's modest open data initiatives are much of a threat to the Treasury, much less the best example of the threat posed by public/private partnerships. This is doubly true thanks to our movement's insistence on nonproprietary formats and open source code, which allow more flexibility and competition in subsequent procurements.

But none of this means that complacency about open data is justified. The current generation of open data advocates is commendably enthusiastic, but we deserve at least some criticism for our callowness. Our community has a strong incentive to insist that the idea of open data is so new that its limits can't–and shouldn't!–yet be pondered. In truth, the U.S. has had something like an open data policy since at least its first census; and that agency has been distributing its information electronically since the seventies. Our tendency to ignore this history in favor of an emphasis on novelty and exciting promises is not only making us look foolish, but is beginning to produce a sense of disappointed malaise both within the movement and among its allies.

This disappointment can be seen in some of the other contributions to this roundtable. Both Clay Shirky's and Aaron Swartz's offerings lament the open data movement's failure to rack up concrete victories against corruption. That our community fostered such unrealistic expectations is a strike against us. There is good empirical evidence that open records laws produce lower-corruption equilibria. But the dream of writing a cron job that moves misbehaving lawmakers smoothly from office to prison was never likely to come to pass (though public disclosure systems have inarguably played important roles in the downfall of figures like Bob Ney, Jack Abramoff, John Ensign and Duke Cunningham). Open data's effect on corruption will more commonly involve altering malefactors' cost:benefit calcu-

lations, consigning corrupt acts to a counterfactual that we're unlikely to ever be able to precisely measure. This is just how enforcement works: putting more cops on a beat doesn't reduce crime simply because they arrest more law-breakers. Applying digital technology to sunshine law disclosures can clearly produce more and better oversight, but it is unlikely to transform that well-established practice into a solved problem.

I fear that our community is presently setting itself up for similar disappointment in the promises we are making about open data's commercial potential. For all of the excitement about Brightscope, DarkSky[24] and a handful of others, the supply of stories about open data startups seems clearly unable to keep up with demand. One way of explaining this is to point out that although open data may be useful, it's also easily accessible to rival businesses; perhaps open data is destined to be as simultaneously useful to and taken-for-granted by businesses as city streets and water pipes. Less palatably, one can acknowledge that the government data resellers of decades past have grown up into the Elseviers and LexisNexises of today.

Unfortunately, there doesn't yet seem to be much appetite for considering these possibilities. At the moment it's more common to hear that we're still in this sector's early days; that a wave of civic startups is just over the horizon. Perhaps that's right. Certainly it would be useful if it were. But at the moment, I think there's reason for doubt.

It seems to me that both of these misjudgments stem from the same underlying error. And it is this problem of philosophy, more than anything else, that threatens to make the final judgment of the open data movement a negative one.

Like most open data advocates, I came to this field by way of software engineering. Writing code is a valuable skill, and an intellectual exercise that I would encourage anyone to explore. But like any discipline, programming colors how one views the world. Computer code is about boolean logic–algorithmic procedures that map input to output in a way that is mathematically perfect and so powerful that it can engender a kind of giddy terror. Many of us who are trained to write code know some linear algebra, but our knowledge of statistics is often surprisingly meager. This makes it all too easy to give in to the hopeful assumption that reality is a comprehensibly deterministic machine; that data are necessarily objective. Once you begin scanning early adopter-types for this expectation you can see it everywhere, from the thirst for a perfect weather app to the quantified self movement's breezy confidence

[24] http://darkskyapp.com/

that the human body is an experimental apparatus that works even when n=1.

The open data movement is no different. It is not uncommon to hear open data advocates promise that newly-released information will allow government to make better decisions. It's a dream embodied by sites like healthdata.gov and the EPA's Apps for the Environment contest. And in one sense, it's a perfectly coherent vision: information *can* lead to better decisions; so can opening deliberative processes to include more qualified participants. If you have any faith at all in democracy and rational deliberation, these ideas are inescapable.

But these ideas can also be easily overextended into the assumption that governance has computable solutions–that politics lingers not because, even after decades of thoughtful analysis, groups have competing claims that must be resolved; but rather because post-partisan technocracy is only now becoming able to offer definitive answers. This is the same wishful thinking that motivates efforts like We The People, MADISON, Americans Elect, and optimism about the net bloc's ability to translate its successful activism against SOPA/PIPA to other issues.

This tendency to deny of the inescapability of politics is a relatively quiet current in the open data movement, but it is a real one. And while I doubt that open data as a cause will live or die by the success of its commercial ambitions, the implicit promise that open data can rescue policy from politics seems destined to end in disappointment. We can smooth the flow of information through our institutions, but this alone will rarely be enough to redeem them, much less render them obsolete.