Shared Values, Clashing Goals: Journalism and Open Government

The difference between aggregating public data and investigative journalism.

By Sarah Cohen

Journalists could be considered among the first open government advocates. In the 1960s, a small band of newspaper editors allied with a member of Congress to create the Freedom of Information Act (FOIA), the first law that codified the public's right to government information. In the 1978 case, NLRB v. Robbins Tire & Rubber Co., the U.S. Supreme Court affirmed the purpose of that law:

" ...to ensure an informed citizenry, vital to the functioning of a democracy, needed to check against corruption and to hold the governors accountable to the governed."

Reporters had historically gone undercover to learn about the workings of important institutions. However this law, combined with the widespread use of new copy machines, changed both the nature and ethics of investigative and beat reporting, ushering in a documents and data-based journalism that was less anecdotal and less ethically hazardous.

Modern journalists now rely extensively on open government information. Data journalists depend on federal satellite images to create fascinating real-time visuals after disasters ranging from the Japan earthquake to the Iowa tornadoes [1, 2]. Feeds from local governments provide fodder for crime maps and election results [3, 4]. Political coverage relies on easy access to campaign finance records and lobbying activity information.

Given this relationship, it would seem natural for the open government movement to align with journalism and media. Their goals are often the same: provide the public with information critical to a successful democracy. Put another way, using government data to keep citizens informed about matters of public importance is a core job of journalism and a core tenet of open government. Citizens,

> **Sometimes data released under open government plans becomes an end in itself rather than a route to answering questions.**

journalists, and entrepreneurs have all benefited from open government when information is easily obtainable in convenient and accessible forms, as long as that information was what governors wanted the governed to see.

Despite this apparent alignment of goals, a culture clash between open government advocates and journalists is beginning to emerge. It helps to understand how reporters find and report their stories—a process that seems misunderstood by many critics who watch the opinion mongering on cable television surrounding big national stories, or who see Twitter and Facebook updates from participants in breaking news events. Most reporters work on local or state stories, not the big national and international events that dominate daily news coverage. They view their job as reaching beyond the talking points of public information officers and acting as a check on powerful institutions. Some cover local crime, local and state government, and federal agencies in their daily beats. Others specialize in investigative reporting or enterprise features.

Investigative reporting, in particular, depends on piecing together disparate bits of information gathered from participants, whistleblowers, documents obtained in both official and unofficial ways, and databases that hold the records created in the course of daily activity ranging from emails to inspection reports. Paige St. John of the *Sarasota Herald-Tribune* won the Pulitzer Prize for Investigative Reporting this year for her series on weaknesses in the Florida property insurance system. To piece together her reports, she listened to dozens of analyst calls and hearings she recorded from webcasts, traveled to Bermuda and Monte Carlo to track billions of dollars sent overseas, cultivated anonymous sources deep inside the industry, learned the intricacies of the arcane insurance industry, and analyzed databases obtained through public records laws in Florida. In the end, she learned that Florida residents were paying among the highest rates for insurance in the nation, but the system was rigged against them.

St. John had a simple question that led her on a yearlong quest: Why are

insurance prices so high? This type of hard question can stem from observation on a beat or a whistleblower but is rarely a product of public information campaigns or Twitter. The question is also at the heart of what journalists do in both new and old media.

In the realm of investigative journalism, specifically, five issues have emerged that make some journalists worry about open government initiatives and open government advocates worry about journalists.

THE DIFFERING CIVIC ROLES

In the past half-century, journalists have avoided advocacy and activism in most mainstream outlets. At the *Washington Post*, where I worked, we were reminded each spring to avoid the seasonal protest marches on the Capitol—attending, except under an assignment to cover the event, was a firing offense for virtually all newsroom employees. In 2004 we were barred from attending the MoveOn.org's Vote for Change tour featuring Bruce Springsteen, R.E.M., and Pearl Jam because tickets amounted to partisan campaign contributions.

Advocacy journalism also avoids collaborating with government. *Mother Jones*, one of the oldest non-profit progressive news outlets in the U.S., is likely to choose stories critical of the FBI or law enforcement; Watchdog.org, funded by the conservative Franklin Center for Government and Public Integrity, focuses its work more on government spending and waste. But neither works hand in hand with government agencies—instead, these organizations choose which elements of government to watch.

This culture is at odds with some open government ideals, often expressed as transparency along with participatory and collaborative government. Beth Noveck, who created President Obama's Open Government Initiative, wrote in her Huffington Post blog in April that she doubted the importance of making basic records about the workings of government public. (Read an interview with Dr. Noveck on p. 16.) She instead considered the effort a strategy for changing how government works by "using public sector information to create more innovative

Using government data to keep citizens informed about matters of public importance is a core job of journalism and a core tenet of open government.

institutions and effective democracy" [5]. She went on to further separate the view of the "innovator" from the "reformer"—her distinction between those who want to help solve problems and create value from public data, and those (often journalists) who focus on accountability.

Journalists I know have few objections to private actors helping create a better government by working hand in hand with agencies to further their missions—they just view it as someone else's job. To them, the job of a journalist is to hold powerful institutions up to public scrutiny—a very different mission.

STORIES OR STUDIES?

Journalists want stories, not studies. But sometimes data released under open government plans becomes an end in itself rather than a route to answering questions. Officials first identify the information they hope to get into the public's hands and then work with developers who can visualize the statistics in an engaging way. Those developers will use the application programming interfaces (APIs) and direct feeds that the agencies have eagerly created in the hopes they will leverage the data. I've worked on a variety of stories that relied on public records, ranging from waste and duplication in the farm subsidy system to failures of the D.C. child welfare system. None of these records were easy to obtain and only a handful have since been released under open government initiatives.

One story, for example, examined a little-known tax break that inadvertently encouraged landlords to allow their properties to deteriorate until tenants were forced to leave. The records included housing complaints (which were once released in part under D.C.'s vaunted open government program, but since have been removed) and a list of the beneficiaries so poorly recorded that we had to manually review each paper file in the agency's records (which are now available in the D.C. data repository).

Consider last year's *Wall Street Journal* series, "Secrets of the System" [6]. In partnership with the nonprofit Center for Public Integrity, the paper paid the fees required to obtain a large database of Medicare claims that has been used for years by researchers at academic institutions and consultancies to find patterns in costs across regions. But instead of looking for statistical trends, the *Journal* looked for signs of fraud. One finding detailed a Florida doctor who billed the federal government $1.2 million one year for physical therapy treatments, more than 24 times the typical doctor's income from the program. Another physician with a telltale pattern of treatment has had his privileges revoked.

At the same time that the *Wall Street Journal* was pursuing potential stories on Medicare fraud, the U.S. Department of Health and Human Services was rolling out its storehouse of datasets under its open government initiative for a different purpose. Mostly made up of compiled data, this release of information was aimed at helping consumers choose providers as well as experts researching public health. The department also launched contests for developers to use the data. One contestant created a game for kids to learn about health risks and another built a website where users could compare health indicators at a county level.

The information needed by journalists for stories is very different than that required for improving public health and consumer awareness. More troubling for journalists is the implicit assumption that government agencies will release information citizens ought to have with no agenda other than the public good.

Figure 1. Stamen Design creates interactive maps and data visualization. Crimespotting is a map of crime data in the Bay Area.
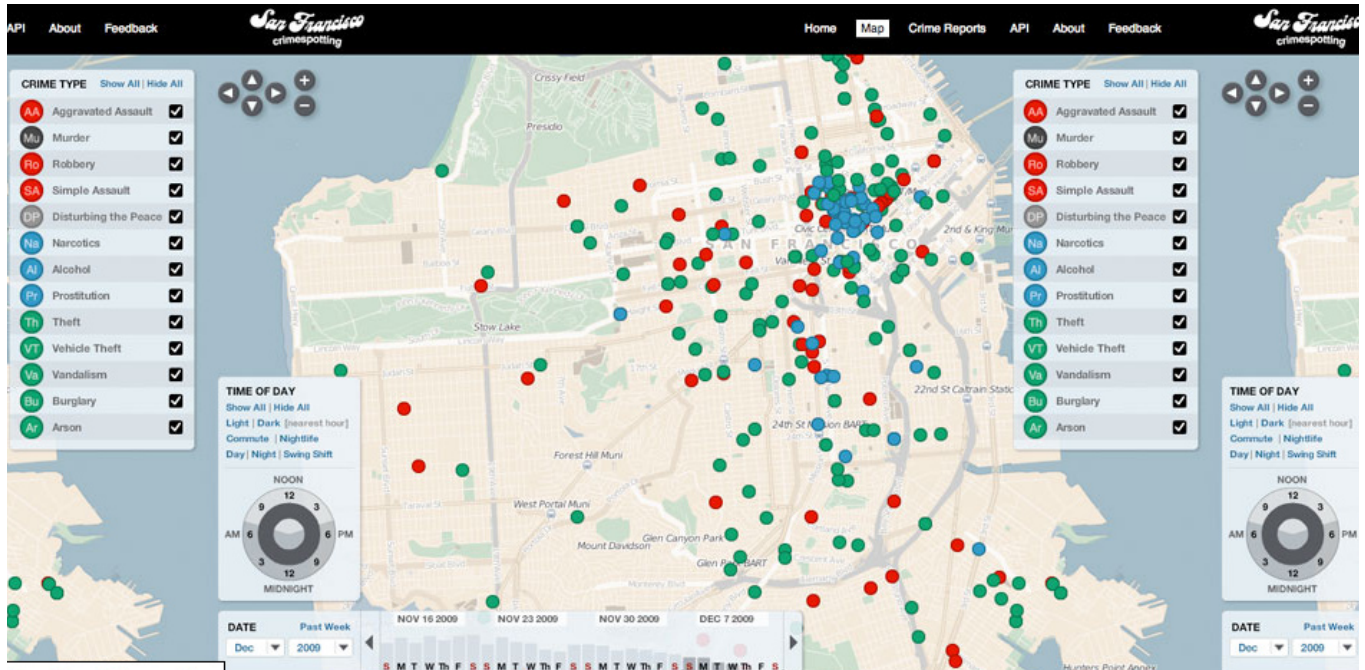
THE STREETLAMP EFFECT

Already, open government suffers from the so-called "streetlamp effect," often described with the following anecdote: A police officer comes upon a drunk man on his hands and knees searching under a streetlamp for his lost keys. When the officer asks why he thinks he the keys are there, he answers that he actually thinks he lost them on the way but stayed under the streetlamp "because that's where the light is."

Similarly, some datasets suffer from overexposure because of the streetlamp effect. Money in campaign finance is just one example. Journalists often want to compare Congressional votes to campaign contributions to see which employees of which companies have given to a member, with the un-stated conclusion that monetary contributions influenced a Congressional vote. In practice, however, uncovering a campaign finance scandal is rarely that simple. I know of no clear cases of quid pro quo that have stemmed from an analysis of the official campaign finance or lobbying reports.

For instance, when *Washington Post* reporters discovered that Jack Abramoff was improperly paying for trips through his non-profit, they did so by tracking down emails and other documents that were created at the time—it didn't come from lobbying

or campaign finance or even IRS reports. When reporters at the *Orange County Register* discovered that Randy Cunningham had taken in funds in an apparent scheme to buy his Congressional votes, they did so by asking a simple question after a routine review of real estate deeds in their county: How did a member of Congress afford such expensive holdings?

Open government initiatives have also suffered from a particularly vexing related problem: lack of documentation. The datasets rarely describe the process used to create the datasets—what base records they were compiled from, how they were extracted or aggregated, what was left out, and how

Government should be aware that its records are now in the hands of private actors who can limit public access in a way that the government itself could not.

complete they might be. Almost none have the name of a responsible person inside government who can answer questions. In my own experience using government records, what's missing is just as important as what's there. It's difficult to use the information with any confidence when that kind of documentation and detail is absent.

ACCURATE DATA OR EASY DATA?

One aim expressed by some advocates of open government is easy and up-to-date access to open source feeds of government information, often through APIs. Journalists love this, too—it allows them to keep their crime maps up-to-date, or their campaign finance data fresh (see **Figure 1**).

But there's a risk that the government has sanitized records, or created whole new side systems only tangentially based on the originals, in order to satisfy this demand. One example is USASpending.gov, a legally mandated database of contract and grant payments maintained by the federal government. To create the database, agencies don't just upload the documents used to pay grantees or contractors. Instead, government workers have to enter their information into two separate systems. In one, they make a real-life transaction. The other serves no business purpose other than to make

that transaction public. At worst, this means that the public can't see signs of fraud or overbilling if an employee wishes to simply lie. There's no evidence that is happening, but serious errors have been documented by Sunlight Labs adding up to billions of dollars. In addition, the details of these payments are still hidden—the original documents are rarely made public, and even then only after protracted negotiations on a case-by-case basis. Contracts themselves are largely secret; grant audits are usually only available using the cumbersome FOIA.

The fake, incorrect, or simulated data is really a proxy for access to the actual administrative records of agencies—the artifacts of governing. Yes, they are easy for developers to use and to update on a regular basis, but there is no natural audit trail to make them accurate, and they are often at odds with the actual information still closely held inside government computers and file cabinets.

The city of Los Angeles, for example, provides feeds of its basic crime information to the public, which has been harvested for years by the hyperlocal site EveryBlock. In 2009, the *Los Angeles Times* data blog noticed an odd pattern: Crime seemed centered around the corner from City Hall. Reporters Ben Welsh and Doug Smith found the answer—the contractor used by the city assigned a default location when geocoding failed.

Inside the LAPD, officers and crime analysts know where the crimes happen—the city just doesn't provide all of its internal information to its contractor, including the coordinates. In turn, the contractor only allowed a limited view of the data, making the mistakes harder to catch.

This is really an issue of the definition of "data." To some, it means useful datasets that can be mapped, analyzed, and easily interpreted. It's anonymous and usually aggregate data—no detail is necessary for this kind of work. To others, it means "records"—those messy, ugly documents that were created in the course of doing business that tell citizens what its government is up to in all of its human and unpredictable forms. To this second group, Noveck's "reformers," the form is less important than the substance.

PRIVATIZATION OF PUBLIC RECORDS

In the 1990s, news organizations thought they had won the battle of private companies claiming monopolies over data collected by and for the public with taxpayer money. A series of lawsuits from New York to California determined that Geographic Information System (GIS) data—usually real estate parcel information—had to be considered a public good and could not be given to the contractor that created the database.

Despite these rulings, the lines between public data and the private sector are often blurred. For example, there are currently more than 100 active challenges for developers listed on challenge.gov, usually in partnership with interested corporations or nonprofits. These challenges sometimes give access to information to a select group of developers and analysts for specific purposes, not to the general public or reporters who might have different goals. Similarly, the White House occasionally holds a social media conference, encouraging (but not requiring) citizens to interact through Facebook and Twitter. Government videos on YouTube could end up the property of Google—and disappear if the social video site were to disappear.

I know of no one who would suggest that the government ignore powerful social media tools or fail to leverage expertise outside its walls. But it should also be aware that its records are now in the hands of private actors who can limit public access in a way that the government itself could not.

Closely related to this problem is what Canadian informatics expert Michael Gurstein has called the "data divide" [7]. He suggests making data available online in open forms empowers the already empowered and does little to make the ordinary citizen more informed or capable of using the information that's available. The solution, most open government experts assert, is for private groups (both commercial and not-for-profit) to make it accessible for the ordinary person. This assumes they have no self-interested motive to make certain points of view more accessible than others, or to make some data more equal than others.

CONCLUSION

Open government and media have championed many of the same values: democracy, citizen engagement and the power of information as an empowering force. But journalists worry that the movement inside government agencies has been more centered on achieving its own aims and improving collaboration and too little centered on the harder job of making themselves open to scrutiny.

Biography

Sarah Cohen joined Duke University's DeWitt Wallace Center for Media and Democracy in 2009 as the Knight Professor of the Practice, where she is leading its initiative to develop the field of computational journalism. Before Duke, she was a prize-winning reporter and editor, including more than 10 years in investigative units at the *Washington Post*. Her awards include the Pulitzer Prize for Investigative Reporting, Harvard University's Goldsmith Prize and the Robert F. Kennedy Public Service journalism award. She was elected to the board of directors of the 4,000-member Investigative Reporters and Editors association in 2010.

References

[1] McLean, A., Quealy, K. Ericson, M., Tse, A. Satellite photos of Japan, before and after the quake and tsunami. *New York Times*, March 2011; http://www.nytimes.com/interactive/2011/03/13/world/asia/satellite-photos-japan-before-and-after-tsunami.html

[2] Before and after: a bird's-eye view of Joplin. National Public Radio, May 2011; http://www.npr.org/2011/05/26/136655052/before-and-after-a-birds-eye-view-of-joplin

[3] L.A. Crime Maps. *Los Angeles Times*; http://projects.latimes.com/mapping-la/crime/

[4] Election 2010: House race ratings. *New York Times*, December 2010; http://elections.nytimes.com/2010/house

[5] Noveck, B. What's in a name? Open gov and good gov. Huffington Post, April 2011; http://www.huffingtonpost.com/beth-simone-noveck/whats-in-a-name-open-gov-_b_845735.html

[6] Secrets of the System. Series. *Wall Street Journal*; http://topics.wsj.com/subject/S/secrets-of-the-system/6281

[7] Gurstein, M. Open data: empowering the empowered or effective data use for everyone? Gurstein's Community Informatics. September 2010; http://gurstein.wordpress.com/2010/09/02/open-data-empowering-the-empowered-or-effective-data-use-for-everyone/