Open Privacy: An open government approach to data privacy

Introduction

Search and Internet services giant Google recently settled a privacy dispute with the Federal Trade Commission (FTC) with a penalty of $7 million and an instruction to beef up their privacy policies related to Google's street view service. At issue in that case was not any invasion of privacy carried out by the company's photo-snapping cars, but the actions of a few of their drivers who took advantage of security flaws in household wireless networks to steal passwords, emails, and troves of other personal information from private computers. At the same time Google's biggest competitors have taken aim at the company's handling of personal data through their email, search and other Internet tools. Microsoft's *"Don't get Scroogled"* marketing campaign highlights the 'reading of email' Google does to serve ads to email, and Google's other data practices.

From another angle, the Supreme Court of the United States considered a case this year that asked the extent to which drug companies had a legitimate interest in using data about doctors' behavior for marketing. Do these companies have a First Amendment right to do as they please with the data doctors are forced to give them? The third-party doctrine adopted first in telecommunications privacy policy offers some guidance: once data is surrendered, the third party may do as he pleases with it. A countervailing interest lies with our understanding of health care privacy under HIPAA, information exchanged between a doctor and patient should not breach that dyad except under special circumstances, and marketing does not seem to be among them.

Among these are other controversies that involved government use of data. The IRS in April claimed it could read citizen emails without a search warrant (The Hill, 2013). Traditionally government agencies would need a warrant to intercept mail between citizens, but insufficiencies in the Electronic Communications Privacy Act leaves some grey area over when and by what means (warrant or subpoena) the government may access email held by third parties (like Google and Microsoft). Unmanned Aerial Vehicles, also known as drones, present a similar problem for citizen privacy. The technology has potential to improve the safety and effectiveness of law enforcement opeartions, but if unchecked it could also lead to constant surveillance of innocent civilians. GPS technology has similar implications for law enforcement and civilians. If law enforcement do not need a warrant to track a citizen's car, trail a citizen with a drone, or read emails, what else might the be able to do?

These kind of situations are among the more puzzling privacy issues policymakers face with information privacy, these, added atop some other classic examples (*q.v.* Mayer Schönburger, *Delete* 2010) further exacerbate the anxiety policy makers should experience when thinking about the legislative lift required to

'solve the problem.' But Congress need not exert itself so strenuously. A single undertaking to fix everything would be over-acting. Instead, Congress should enact a law instructing the Federal Government to create an infrastructure of open standards, couple them with a broad directive to government agencies to create information privacy policies around those standards that encourage voluntary compliance from the private sector.

Do not Break Anything

Data surround us. Researchers Bounie and Gile (2012) estimated the volume of data produced globally in 2008 was nearly 15 exabytes, enough to fill two billion DVDs. Some of these data are being collected by private companies who often store them and send them to other third-party processors or sell them to advertisers to build web services that work quite efficiently. Google's Gmail service, for example, uses data about who its users email and how frequently to try and guess who you mean when you start typing just a few letters.

These kind of predictive services are amazingly efficient. Before Google rolled out their 'Instant' search service in 2010 the typical user took about 9 seconds to type a search. Since then they estimated "more than 3.5 billion seconds" were saved daily simply by reducing that average to fewer than five (Google, 2010). This product built off of Google's existing search algorithms, but also the data collected about what people *mean* when they search. Without the collection of troves of data about what word it's users most frequently completed when they type a given combination of letters, they could not have built such a well functioning tool.

For a show of how effective data-informed services are, a quick, unscientific comparison of Google and their user-data-free search competitor DuckDuckGo was telling. Searches for "Boston Marathon," and "Boston Marathon 2013" the morning after the notorious 2013 bombing on DuckDuckGo showed only generic results about the race.While a search for *Boston Marathon bombing* did, few of the results were from major or even Boston-based new sources (Utah television broadcaster). Google's results, conversely, floated links to the New York Times, CNN, and Fox News above blogs and smaller, less reputable sources. In fact, after the 14th result, none of DuckDuckGo's results even mentioned the bombing. Google were able to deliver three news results at the top of the heap, and an additional eight results about the bombing on their first page. DuckDuckGo's results were inferior. While refusal to use user data cannot account for all of this discrepancy, the difference is nonetheless stark.

Herein lies the tension between privacy and making things that work. Most of Google's products would not work as well without the data they use in their products. They certainly would not have the capital to make new products without selling something, and click-through data are easier to sell than other models like subscriptions and premium services. (Imagine trying to convince

someone to pay for each search made on Google.) On the one hand, individuals could search DuckDuckGo quietly without worry about third parties connecting search terms back to them, but they might get an inferior product.

Search is an easy example of the efficiencies data can provide in technologies, but others face similar issues. Drones equipped with sensors, GPS radios, cameras and more could be used by a farmer trying to optimize their crop, a peeping Tom (or worse) to stalk and prey on innocent people, or by a nefarious national security program to track individuals without their consent with little change in the underlying technology. This list goes on and on, from GPS trackers in cellphones and automobiles, to the sensors and services powering the devices that make up the 'Internet of things:' the collection, storage, aggregation and processing of data are key to a new technology's most optimal operation, but also it's most dangerous.

Any public policy made on information privacy must be done with special attention given to this delicate balance between clever innovation and mitigating its dangerous uses. While solutions like Mozilla Firefox's default third-party cookie blocking and the Do Not Track standard built into other Internet browsers are compelling, they only solve part of the problem. These solutions give users more control over what is shared, but leave them without much guidance as to how to discover what happens to the data after they choose to share it.

Past regulation attempts have failed to respond effectively to the problems thrown by these new technologies. This is in part because they are outdated. ECPA was enacted some 27 years ago, and the only law we have about online privacy applies only to children (the Children's Online Privacy Protection Act, [COPPA]). There is a sense of urgency to correct this: the longer Congress waits to act, the further technology will progress, the larger the databases will grow, and the more out of hand the issues around data use will become. The European Union sought to update its 1995 directive on electronic privacy in 2013 using the 'grand undertaking' approach they employed in the past, but the United States has yet to enact any new legislation, mostly leaving any action to agencies like the Federal Trade Commission (FTC) to audit based on existing laws. Over time, these agencies have crafted a plurality of privacy recommendations and policies mostly derived from some understanding of the Fair Information Practices Principles (FIPPs), which lends credibility to Helen Nissenbaum's ideas about privacy being context specific.

Indeed these various problems each seem like discrete areas where privacy should be protected distinctly from the others. Nissenbaum argued that privacy policies should be crafted in line with the social norms attached to different kinds of information flows. There is economic incentive for society to individuals being forced into the credit system (fewer bad loans being issued) and to the individuals themselves (potentially better rates and access to credit when it is needed) and individuals are compelled into the system. In this case our right to privacy is consistent with the expectations we have of other parties to handle information about our creditworthiness. Nissenbaum uses the case of health care. We expect

that information shared with our doctors in clinics stays between us and the other medical professionals our doctors need to consult to return a diagnosis and care plan. We regulate when the use of our data breaches our expectations (Nissenbaum 2010).

Nissenbaum's ideas are compelling, to be sure, and she's correct that privacy seems to be context aware. But the problem is not that these data are or could be used outside of a given context of social norms, because indeed they will be, it is that new technologies, once deployed into a society, create situations for which there are no new social norms or challenge society our norms to the technology's utility.

For Nissenbaum the solution to privacy policy problems is a surgical, sociological approach to understanding the situation and making policy around the appropriate flow of data for that context. Fred Cate takes a more legalistic approach. Cate dissected the history of the "fair information practice principles," (FIPPs) to ultimately ask "which FIPPs should we follow?" What Cate found was that the interesting and useful idea of a universal declaration of principles to guide all data flows has failed spectacularly. What we have accumulated since the 1970s is a collection of enumerated lists all of which are different versions of the same idea and only result in long, complicated policies nobody is really sure how to enforce.

A National Open Privacy Platform

The perplexing problem of privacy policy is that none of the frameworks seem to offer a wholly compelling way forward as technologies develop and capture more and different kinds of data put to new and creative uses. But perhaps the answer is not hitching our wagon to one solution, accepting it's problems and leaving the rest behind. Perhaps a hybrid solution is necessary, one that does not require a phenomenally, impossibly comprehensive Act of Congress, but a smaller one that builds a privacy platform that puts protecting citizen data first.

Tim O'Rilley (2010) champions a concept called "Government As A Platform" (GAAP) wherein government exits alongside other successful technological platforms. The government is the infrastructural provider for a society based on openness and interoperability. Good policy builds a simple system to tackle big problems. O'Rilley is fond of praising "the Federal-Aid Highway Act of 1956 which committed the United States to building an interstate highway system," as an exemplar of GAAP policymaking. The Highway Act set up a network atop which a whole host of previously difficult innovation could be done. In that case, the government was able to make laws regulating traffic and safety on the highways themselves, "interstate commerce,... gasoline taxes and fees on... vehicles that damage the roads,... speed limits, specifying criteria for the safety of bridges, tunnels, and vehicles that travel on the roads... as the 'platform provider.' " But once on the platform (i.e., driving on the highways), Americans

could go anywhere and establish "factories, farms, and businesses" that use the network, collectively strengthening the innovation society of post-war America.

The key lesson of the Federal-Aid Highway Act, O'Rilley argues, for future policies, is that the government works best when it "invests in infrastructure (and 'rules of the road') that will lead to a more robust private sector ecosystem." The key problem with existing privacy frameworks is that they put government in the role of competitor; government, in these models, would attempt to impress upon private citizens and business entities alike, a constraining regulation. A better solution would be for the government to lay out a platform that hold the public and private sectors to the same standards of data use that at once protect citizen data and promote continued proliferation of innovative technologies.

This solution begins with pervasive transparency from the start by compelling all federal agencies to create standard for privacy to which they will adhere. It builds off the existing Federal Privacy Act, first passed in 1974, which governs how citizen information collected and stored in electronic government databases shall be handled.

Congress should take the Code of Fair Information Practices and revise them against one of the many FIPPs laid out by various federal agencies to make clear what the government can and cannot do with data on a broad level. Congress should then instruct the various federal agencies to elucidate a clear policy for how any information collected from a citizen is stored, used, transferred, and protected and what mechanisms citizens have to gain access to data stored about them and correct that information if it is wrong. Additionally, if the agency requires long-term storage of personally identifiable information (nearly all information these days is PII) it should make clear for how long data will be stored in the database and for what purpose. The government, after all, is no stranger to data collection from citizens. Indeed, citizens are compelled to inform the government of their income, marital status, home address, and a host of other data just to pay taxes.

On top of a major technical system that would need implementing here would come a degree of documentation and accessibility. It is one thing to declare the process is operating a certain way, but government agencies should be compelled to show how their data storage practices work in as detailed a way as possible without compromising the integrity of the database. Civilians should have easy mechanisms to see an audit trail of their data and watch them move through the system. Any citizens concerned about where their data have been, who accessed them and for what purpose should be able to easily retrieve that information.

Setting this kind of regime in place will mean changing the way thousands of bits of government data are collected while at the same time structuring the vast amounts of possibly unstructured data that are sitting in government databases around the country. The first standard that will be created is a method for tagging and tracking data as they flow through the system. Paula Bruening and K. Krasnow Waterman sketched a system of affixing metadata to bits of

information that can be used to follow and govern their flow through a system. Once tagged, simple algorithms can be written to determine which data can flow to which parts of the greater system. In order to bring complete transparency to this system, the government should release the algorithms for data governance into the public domain and offer comprehensive documentation for implementing the system. This will also encourage non-governmental actors to implement the same systems since they will have no fee for use, and be interoprable with other similar systems. If implemented correctly, a citizen should need nothing more than a telephone or Internet browser to request data from the government and understand how and where they are being used.

Along with creating an infrastructure, O'Rilley argues government should create standards to which private actors may voluntarily adhere. The federal privacy policies, and the algorithms and documentation that support them will become standards published by the government as a best practice. Any private sector actor collecting data will be invited to voluntarily adopt these same standards for data processing. A company like Google or Facebook could pick up the Government's data tagging standard to help it's users trace data through Google's larger infrastructure. Much like other government standards (like Energy Star), private companies may voluntarily certify themselves with these data protection protocols.

This policy would have only one enforcement mechanism: certification. The FTC would be charged with the task of auditing (as they already do) electronic data practices among companies doing business in the United States. Any entity that is Privacy certified would be exempt from these routine audits unless they are suspected of deceiving consumers. Just as anyone may drive on the highway until they break one of the "rules of the road," anyone may collect, aggregate and use data for as long as they want until they fall out of line with the open standard they have agreed to follow.

Government As A Platform can Work in Privacy, If we Only Try

One obvious flaw in this plan is the sheer volume of government agencies that would need to create privacy policies. With more than 1300 government agencies operating under the federal government, Fred Cate gets extra impetus behind his question of "which FIPPs?" The operationalization of this idea is difficult, the last thing we need is 1300 of them. But the data collection, storage, and processing standards that would emerge out of this kind of public policy would hopefully limit the variance to a manageable number. This is a challenge for implementation, but not an insurmountable one.

Electronic data privacy policy can be made one of two ways. Either by restricting good behavior and penalizing misbehavior, or by establishing a set of open, public standards the federal government will hold itself to and make available to private

organizations to voluntarily comply with. The latter allows much more latitude and flexibility for adjusting the policy to meet future needs, but more importantly, it allows the government to achieve a desirable public policy outcome without prohibiting the market from exploring, developing, and innovating technologies that may be profoundly useful to society. It gives citizens more access to the core of what government does to protect their privacy as well as how the might be treated by private service providers.

References and Sources Consulted

1. Bounie, D., & Gille, L. (2012). Info Capacity| International Production and Dissemination of Information: Results, Methodological Issues and Statistical Perspectives. *International Journal of Communication*; Vol 6 (2012). Retrieved from http://ijoc.org/ojs/index.php/ijoc/article/view/1389/744
2. Bruening, Paula J., and K. Krasnow Waterman. "Data Tagging for New Information Governance Models." 2010. In *IEEE Security & Privacy Magazine* 8.5 pp. 64-68
3. Fred Cate, "Failure of Fair Information Practice Principles", Chapter 13 in Winn, J.K. (Eds), *Consumer Protection in the Age of the Information Economy*, 2006, pp. 343-369
4. The Hill, 2013
5. Nissenbaum, Helen, *Privacy in Context*, 2010, Stanford University Press.
6. Richard Posner, The Right of Privacy, 12 Georgia Law Review 393 (1978) pp.393 - 404
7. Jonas, Jeff and Harper, Jim, "Open Government: The Privacy Imperative", in Ruma, Laurel and Steele, Julia eds., *Open Government*, 2010, O'Riley Media, Sebastopol, CA
8. Mayer Schönberger, Viktor, *Delete: The Virtue of Forgetting in the Digital Age*. 2010. Paperback. [Publisher, City]
9. O'Riley, Tim, "Government As A Platform", in Ruma Laurel and Steele, Julia eds., *Open Government*, 2010, O'Riley Media, Sebastopol, CA.
10. White House Cabinet. 2013. http://whitehouse.gov/administration/cabinet