



United States of America
Federal Trade Commission

The Social Impact of Open Data

Remarks of Maureen K. Ohlhausen¹
Commissioner, Federal Trade Commission

Center for Data Innovation
The Social Impact of Open Data
Washington, DC
July 23, 2014

Thank you for inviting me to kick off what I am certain will be an insightful discussion of open data and its benefits for society. As society has integrated and adopted increasingly powerful computers and pervasive communications networks, we have created massive amounts of data. Government has been a major participant in this data revolution¹, both gathering and creating large stores of data about an enormous range of topics. New techniques and tools have the potential to create great value from both recent and historical government data, and there are several steps the FTC can take to help facilitate more open government data.

Modern data analysis techniques – sometimes colloquially referred to as “big data” – are incredibly useful tools. Data today has greater volume, variety, and velocity. Put more simply, there is a lot of data, it has many different forms, and it is created rapidly. Tools that can pull useful needles of insight from this data haystack have great potential to make our lives better. Data alone is not knowledge, however. The promise is that big data techniques will help us

¹ The views expressed in this speech are solely those of Commissioner Ohlhausen and are not intended to reflect the views of the Commission or any other Commissioner.

extract knowledge from data, and this knowledge will help us better understand ourselves and the world around us.

Obviously, we are already seeing benefits from the use of big data techniques on private sector data. Amazon, e-Bay, Netflix, and many other online merchants use big data to generate customized user recommendations. Other companies use big data to predict commute times by aggregating millions of GPS signals, to identify potential causes of disease, and to detect and prevent credit card fraud. Kaiser Permanente used big data analysis to discover an increased chance of heart attack or cardiac death among Vioxx users as compared to users of a competing medication. Scientists are using massive data sets and powerful analytic tools to make progress on the most difficult problems in the health sciences and hard sciences. And many new uses are emerging, particularly because consumers are no longer simply data points to be researched. Today's consumers are themselves producers and users of big data, whether posting billions of photos on Facebook, using Bing's flight price predictors to make travel plans, or joining the self-quantification movement by wearing a FitBit Flex and using an Aria bathroom scale, like I do. As more of our everyday existence becomes measurable and recordable, the greater potential there is for big data to provide helpful insights.

Yet in our quest to mine data for meaning, we should not forget more conventional sources. In particular, government is – and has long been – a major data producer. Often, when people talk about government and data they focus on government as a *consumer* of information and how government should or should not be limited in the data it can collect and use. We have an entire section of constitutional law dedicated to that topic. But there is another aspect of government data that isn't discussed as much, except perhaps by the people in this room: Government as a producer of data. Federal, state, and local governments generate and store

massive amounts of data about themselves, about us, and about the world around us. Even before the very first U.S. census report, government has been producing large – and increasing – amounts of data. Government produces many types of data: Personal data, such as social security earnings, tax information, unemployment filings, and voter registration; societal data such as demographics, employment estimates, and economic indicators; and impersonal or scientific data, such as weather and climate measurements and geolocation data.

There is great potential in applying powerful new big data tools to the rich troves of government data. The private sector could use the wide range of government-produced data to reveal new insights into difficult problems in nearly every area of human endeavor. In fact, McKinsey & Company has estimated that opening government data to private sector access has the global potential to create more than \$3 trillion in additional value annually.²

We already have many positive examples of how the private sector uses government data. Every weather forecaster builds his or her forecast on the raw data and models provided by the National Weather Service and similar government organizations around the world. Investors pour over government financial, agricultural, and economic reports in making decisions. (Those who, like me, grew up in the 1980s may recall that the climax of the Dan Aykroyd / Eddie Murphy movie “Trading Places” turns on the release of one such report – the Department of Agriculture’s orange crop report.) The FTC’s recent Data Brokers Report noted that many data brokers rely on government data such as city demographics, geographical information, professional licenses, real property and assessors’ records, court records, and more to market

² McKinsey Global Institute, *Open Data: Unlocking Innovation and Performance with Liquid Information*, 6 (October 2013), available at http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information.

products and prevent fraud.³ And these are just a very few examples. Overall, in a recent report, the Commerce Department concluded “firms that intensively use the statistical agencies’ data ... combine it with other government and private sector data to create between \$24 billion and \$221 billion in annual revenues.”⁴

The upcoming panel will talk about the benefits of open government data, and I am sure they will share more examples. With my remaining time, I’d like to focus on what the Federal Trade Commission can do to promote useful access to government data.

First, to promote access to government data, the FTC should adopt and promote a realistic view of big data. Big data is a useful tool that, like all tools, has both strengths and shortcomings. Data, even big data, isn’t knowledge or wisdom. It can be misleading. There can be “signal problems,” where the data set, huge as it may be, does not represent the real world. The City of Boston’s StreetBump mobile app is an example; the crowd-sourced pothole-finding mobile phone app identified more potholes in wealthy areas of the city simply because more residents in those neighborhoods used smartphones. Big data sets can also let us mislead ourselves. As Nate Silver explains in his book “The Signal and the Noise,” our instinctual shortcut when we have too much data is to pick out the parts we like and ignore the remainder.⁵ Big data is particularly vulnerable to the “multiple comparisons problem.” Big data tools are very good at discovering statistical correlations between variables in complex data sets. And because there are many variables in the typical big data set, there are many potential

³ See FTC Report, Fed. Trade Comm’n, *Data Brokers: A Call for Transparency and Accountability*, 11 (May 2014), available at <http://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014>.

⁴ Press Release, Economics & Statistics Administration, U.S. Dep’t. of Commerce, *New Commerce Report Explores the Value of Government Data* (July 14, 2014), available at <http://www.esa.doc.gov/news/2014/07/14/new-commerce-report-explores-value-government-data>.

⁵ NATE SILVER, *THE SIGNAL AND THE NOISE: WHY SO MANY PREDICTIONS FAIL - BUT SOME DON’T* 3 (2012).

relationships for a researcher to test. If a researcher explores a big data set without a particular question or theory in mind but instead simply tries enough comparisons between variables, they will often be able to find “statistically significant” correlations that do not reveal anything useful about *causation* in the real world.

By understanding the limits of big data and emphasizing the need for human judgment in the use of such tools, the FTC can help tamp down hype over big data. The FTC can help create a healthier regulatory atmosphere by critically evaluating the claims of both the pop-science promoters of big data as a “magic bullet” solution and the naysayers who fear massive consumer harm from all-knowing algorithms. A realistic understanding of big data’s potential will help the agency to identify and focus on actual harms to consumers, if they occur.

Second, the FTC can continue to provide guidance on how to protect the privacy of individuals while promoting open data. Obviously many – perhaps even the majority – of government data sets have nothing to do with “personally identifiable information.” Open access to many scientific and economic data sets, for example, raises no privacy risks. However, opening other useful data sets may raise some privacy concerns. For example, applying big data techniques to government health data or education records could help address the most pressing societal issues we face, but people understandably worry about how such information is used and shared. The FTC can guide other government agencies on how to open access to data while mitigating privacy risks through aggregation, de-identification, use-based limitations, and other techniques. Furthermore, the FTC must continue to explore how to resolve the tension between the promise of big data and certain Fair Information Practice Principles such as notice and purpose limitation and data minimization, which, strictly applied, could hinder big data’s promise.

Third and finally, the FTC can help promote a government, industry, and consumer culture that embraces open data and innovative data analysis. By taking a humble regulatory approach to the fast-changing data industry, the FTC can help create this culture. Our most successful technological advances, such as the Internet itself, have generated massive consumer welfare and have thrived largely because market participants have enjoyed wide latitude to experiment with new technology-driven business models, allowing the market to determine which models succeed and which fail. Ultimately, our work as an agency should help strengthen competition and allow the market to better provide beneficial outcomes in response to consumer demand, rather than dictate desired outcomes to the market. Similarly, the FTC should focus its enforcement efforts on bad actors actually causing substantial consumer harm, rather than attempt to preempt speculative future harms. By doing so, we can help build and support consumer confidence in emerging technologies while allowing innovation to flourish.

The FTC recently recommended that Congress consider requiring data brokers to give consumers more transparency and control over the information such brokers have about individual consumers.⁶ I believe a proper implementation of this recommendation could help consumers embrace the big data movement. Increased government transparency about the data it holds also promises similar benefits. Allowing consumers to easily access the information the government has about them could also promote consumer interest in and acceptance of open data more generally. I urge government agencies to explore how they could achieve more transparency about the data they hold on individuals.

⁶ Fed. Trade Comm'n, *Data Brokers: A Call for Transparency and Accountability*, 49.

Conclusion

To conclude, more open government data promises to fuel insights across the spectrum of human endeavor, including in science, economics, education, and sociology. Big data tools will help us find useful insights from the massive data the government produces. By embracing a realistic, informed view of big data's potential, providing guidance on the privacy issues raised by open government data in the age of big data tools, and by promoting a culture that embraces open data and innovative data analysis, the FTC can serve an important role in ensuring that citizens benefit from the massive amounts of data produced by the government. I hope that you all will see the FTC as an ally in efforts to expand access to government-produced information and will reach out to us with any specific suggestions or concerns.

Thank you for having me today, and I look forward to your questions.