

Big Data, Big Issues
Fordham University School of Law
March 2, 2012

Thank you, Joel, for that kind introduction. I'm so pleased to be part of this important discussion about the benefits and concerns surrounding collection, use and retention of Big Data.

As we heard this morning, there is no question that collecting, culling, dissecting and cataloguing vast quantities of consumer data, from such sources as social media, online behavior, geolocation data, and the like, has important beneficial uses. In the past several months we've heard reports about how health care costs can be reduced through large scale analyses made possible by big data. Other researchers have reported how sophisticated analyses of traffic patterns and congestion can be analyzed for "smart routing," which could be designed to save consumers' time. And this morning, we heard about uses of Big Data that really make a difference in people's lives, such as predicting infections in newborn babies – where having this information in real time can save lives.

Among the most intriguing uses of Big Data have been efforts to tease out, from social media services, how consumers in the aggregate feel about a product or brand. Internet entrepreneurs are forming companies to provide so-called "Sentiment Analysis" to investors, with insights gleaned from social networks about how consumers feel about certain brands.¹ To institutional investors, this information could be extremely valuable in building strong portfolios.

Sentiment analysis has other uses, including identifying public health concerns and other areas of need. A new initiative by the United Nations analyzes social media public information

bers to be appealing. I'm glad others now agree with me.

a form of Big Data that uses large amounts of anonymous data. At
. The individual source of the data feeding into the sentiment

So if this information indeed is – and remains – truly anonymous, and can be put to such creative and beneficial uses, I'm not going to lose sleep over it. Rather, I'll be intrigued to see how beneficial sentiment analysis proves to be in the coming years.

But there are certain Big Data uses and practices that are on my radar screen and get between me and a good night's sleep. As a Commissioner at the Federal Trade Commission, it is my job to protect consumer privacy—a business that the FTC has been in for quite some time. And a business that I have also put many years into—even before I started at the Commission in 2010.

I spent more than 20 years working on consumer privacy issues at the state level – leading the State Attorneys General as chair of their Privacy Working Group, and engaging in enforcement actions on behalf of the states of Vermont and North Carolina. So, privacy is not new to me.

But Big Data's impact on privacy is requiring some new and hard thinking by all of us.

Today I'd like to talk to you about four issues relating to Big Data and privacy that I have been thinking about, and that I pose as issues for you to consider today.

First, turning back to sentiment analysis and other uses of data that claim to be deidentified: it is critical that we drill down and determine whether the information amassed for such analysis is in fact truly anonymous. Researchers – including some here today – have demonstrated that it can be relatively easy to take some types of deidentified data and reassociate it with specific consumers. I am concerned when I hear other researchers claim that information is “deidentified” when it is merely stripped of a name and address. Much of this information may instead be linked to a specific smartphone or laptop. Given how closely these devices are now associated with each of us — many of us sleep more closely to our cell phones than we do our spouses! — data that is linked to specific devices through UDIDs, IP addresses, “fingerprinting” and other means are, for all intents and purposes, linked to individuals.

whether a consumer was pregnant, but also when her baby was due, so that Target could tailor its offers depending on her stage of pregnancy.

Now let's suppose that Target didn't use any health information in creating its pregnancy prediction score. Let's simply suppose that it used "innocuous" data – such as the purchase of lotions and then several months later the purchase of newborn-size diapers – to determine the kind of purchases and other customer habits that indicate a shopper is pregnant. That is, it used non-sensitive information to create a prediction about health status. The same type of innocuous data could be used to make other predictions of a sensitive nature, like sexual orientation, financial status, and the like.

As Steve Bellovin pointed out, even Target came to understand that its customers were "creeped out" by getting coupons and other offers that clearly indicated that Target "knew" they were pregnant. So Target "disguised" its knowledge by including among the coupons aimed at expectant moms some coupons for other items, making it less obvious that Target was targeting the women with pregnancy-

the DAA program.⁵ As usual, the details on how this will be implemented will be critically important. I will closely watch industry's progress in creating a robust solution to effectuating consumers' choices, including whether choices provided to consumers address the collection of

Just six weeks ago, I called on the data broker industry to develop a user-friendly, one-stop shop that will give consumers information about who the data brokers are, and provide access to information that data brokers have amassed about them. If a consumer learns that the data broker sells her information for marketing purposes, she should be able to opt-out.

And with respect to information about consumers used for substantive decisions – like credit, insurance, employment, and other benefits – consumers should have the ability to access this information and correct it. And correct such errors wherever they occur. I call on the data broker industry to develop a system where a consumer’s corrections to one data broker’s files will automatically correct the same information held by other data brokers. It is critical that all data brokers come to the table to develop this mechanism—including those in the mobile space.

We’ve spent the morning talking about the possibilities – both positive and negative – of Big Data.

While we learned quite a bit this morning, I’ll sum up the main points in just a few words. The potential benefits of Big Data are many, consumer understanding is lacking, and the potential risks are considerable.

We need to pay attention to these issues so that the promise of Big Data is realized, and the risks are kept to a minimum. Industry has a strong and justifiable need to continue to innovate. But in order for industry to thrive, it must engage in an honest discussion about its collection and use practices in order to instill consumer trust in the online and mobile marketplace.

Thank you.