



**Your on the
Street Reporter**



Uyless Black

**A Monty Python Internet:
The Convergence of Big Data and Metadata**

A Monty Python Internet: The Convergence of Big Data and Metadata

February 8, 2015

Hello from Your on the Street Reporter. To begin this report, we revisit earlier articles to review two terms:

- **Big data:** The use of huge data bases containing billions of records about any subject. For these reports: data about citizens' personal and professional lives.
- **Metadata:** in which these files do not specifically identify a person. For example, the purchase of a pair of shoes does not reveal that you or I bought the shoes; only when and where (and for what amount) they were purchased.

Another technology combines these two technologies, one that I have coined bigmeta. Granted, it is a contrived word, but it will be useful as a short-hand descriptor for these two ideas.

Bigmeta software is now being used to make extraordinary intrusions into the privacy of Internet users. While reading about the technology, keep in mind that this system does not need our names, street addresses, zip codes, or an Internet identifier (such as an IP address or an email name) that could be associated with our names:

Researchers at the Massachusetts Institute of Technology...analyzed anonymous credit-card transactions by 1.1 million people. Using a new analytic formula, they needed only four bits of secondary information--- metadata such as location and timing---to identify the unique individual purchasing patterns of 90% of people involved, even when the data were scrubbed of any names, account numbers or other obvious identifiers.¹

One might think: "What is the harm of someone knowing I prefer buying brand x shoes instead of brand y? This information is for marketing purposes, a trivial aspect of life." At the risk of raising *the sky is falling* alarm, this technology can be extended beyond selling shoes, wrist watches, and such. Data analysts are applying it to make further inroads into a citizen's private life.

As well, there is discussion about its application to the collection of data from government monitoring programs (such as the NSA bulk data surveillance program), which already correlates cell phone calls with data bases, such as addresses and credit cards.

Bigmeta is astonishing in its inferential power:

After isolating a purchasing pattern, researchers said, an analysis could find the name of the person in question by matching their activity against other publically available information such as profiles on LinkedIn and

¹ Robert Lee Hotz, "Even Nameless Data Can Reveal Identity, Study Finds," *The Wall Street Journal*, January 30, 2015, A3.

Facebook, Twitter messages that contain time and location information, and social-media “check-in” apps, such as Foursquare.²

What is the limit and where is the line drawn on the latitude to invade one’s privacy? Uber claims it is able to calculate the possibility (likelihood) of its weekend riders visiting local prostitutes. With bigmeta technology, the Johns of the world risk having themselves and their Janes exposed. But it goes beyond a social embarrassment.

America’s courts have stated that the Constitution guarantees the privacy of mail.³ But Internet email appeared almost out of nowhere. Few people knew that it would become a successor to conventional first-class mail. However, internet correspondence does not have a government stamp on its envelope. Thus, the way these systems are evolving could lead to extraordinary losses of privacy.

(As a side bar, this writer has discussed this subject several times in previous articles. While doing on-going research on the subject, I have been surprised by the seemingly lack of concern among many citizens about the loss of their privacy.)

With the world going online, with the likelihood that electronic correspondence will replace pen and paper; that email will replace conventional mail, we citizens have a choice to make. Albert Schweitzer frames the issue well: “No one should compel himself to show to others more of his inner life than he feels it natural to show.”⁴

But do we actually have a choice? Have we crossed the Rubicon? I cannot do business if I do not have access to the Internet. Nor can millions of others.

Even more, we have no control over the cameras that cover our movements on highways and streets. We have no way to know that our seemingly few moments of solitude---not in the remote reaches of Glacier Park---but in our back yards, are not being recorded by Google’s maps or our neighbors’ drone.

On the upside, perhaps the U.S. government can use the technology to determine the possibility of terrorists showing up at a marathon race. That is problematic, as it is unlikely a terrorist will sign on to social networks or leave behind credit card trails.

Reins on Uncle Sam?

Nonetheless, bigmeta systems are relatively new. With the amazing hit rates of 90 percent cited earlier and with the technology surely to improve, the current low hit rates of the NSA bulk data surveillance programs should not lead to these programs being terminated, Quite the opposite.

As I have stated before, I believe the American citizens will accept the capture of their personal data, as long as they have assurance it is not being used unless it is subject to the same restrictions as the old wire-tap procedures; as long as they have assurance no one is reading their mail without a court order. The disturbing aspect of U.S. government surveillance programs is that they get out of hand, and at times, are used illegally.

However, eventually, citizens’ patience will be exhausted, and someone is going to be doing jail time. NSA is not the problem. NSA is part of the solution. The problem is commercial.

² Hotz, A3.

³ That action—handling someone else’s mail, reading the contents and then destroying it—directly violates 18 U.S.C. § 1703. It is a federal crime punishable by a fine and up to a year in prison.

⁴ Albert Schweitzer, “Memoirs of Childhood and Youth,” in Leonard Roy Frank, *Quotationary* (New York: Random House, 2001), 743.

Reins on Internet Providers?

The bigmeta technology can be used on any large database that stores information about humans' activities. So, how can a citizen protect him/herself from this sophisticated form of surveillance? In this situation, metadata itself would have to be encrypted and decrypted end-to-end: from you to me. That is to say, not from a Google's server to another Google server but from my machine to your machine. Even then, as this article has articulated, the combined use of this analytic software in consonance with big data and metadata (bigmeta) yields a treasure trove of information for Internet data miners.

What can be done? We can throw away our credit cards, close our accounts with the social databases, and never logon to the Internet. Is that what we want in our future? Stated in a less lofty way: to be fodder for advertisers' swill?

Monty Python's Bigmeta Circus

Given the fact that electronic correspondence will eventually be *the* major way in which we communicate with one another, we humans should begin challenging the current mentality of: "Your data is my data. Your personal correspondence is mine to use. I can open your email envelope anytime I wish. Your privacy is irrelevant to my mining your data for selling you a car."

Think about that last statement. It is so absurd that it could be in a Monty Python skit. The way things are going now, targeting a Ford coupe instead of a Toyota SUV to an Internet user is becoming more important than the right to privacy; a right that is declared by America's judicial system to be implicit in the Constitution. Is that what we want our society to become?

If our current acquiescence continues, we will end-up consigning ourselves to giving-over a vital part of our lives: private information about ourselves.

What can be done? On occasion, I wish for those times of the 1960s and 70s when sustained mass movements of the citizenry moved those in power to make changes. As military man, I did not support the Vietnam war protesters, but I admired their persistence and tenacity.

For now, contact your representatives in Congress and express your concern. Perhaps in between their gerrymandering, spending our taxes on trying to stay in office, and building super PACS, they will focus on one of most important social issues of their time.

I wish I had a firmer recommendation, but all of us have been over-whelmed by the effect the Internet has had---and is having---on our lives. But not the advertisers and other data miners. They have found their mother lode: Us.

By the way, it is only a matter of time before those advertisements that come up on your screen to block what you logged-on to see cannot be "clicked-off." The good news is that they will be targeted to your browsing and buying habits, because the characteristics of your behavior is available in bigmeta. Easy shopping exchanged for privacy. Which do you prefer?