

*The Advocate*, Stamford: <http://www.stamfordadvocate.com/default/article/When-it-comes-to-data-gathering-the-individual-4635476.php>

*Greenwich Time*, Greenwich: <http://www.greenwichtime.com/default/article/When-it-comes-to-data-gathering-the-individual-4635476.php>

*News-Times*, Danbury: <http://www.newstimes.com/default/article/When-it-comes-to-data-gathering-the-individual-4635476.php>

*Connecticut Post*, Bridgeport: <http://www.ctpost.com/opinion/article/When-it-comes-to-data-gathering-the-individual-4635476.php>

## When it comes to data-gathering, the individual hardly matters

**Published on the Internet, June 28 and in print June 29, 2013**

I am not surprised the federal government has been spying on me. Corporations have been doing it for years.

"Big data," it seems, is about to become the catch phrase of 2013, as the federal government and corporations tap into larger and more complex data sources (and as they use more powerful tools and analytical methods to go deeper into the data than ever before).

The public has expressed shock, and has naturally focused on privacy concerns.

Individual demographic data have been available for years, along with information on customer purchases, credit scores, subscriptions and donations. But, today information sources also include social networks, blogs and mobile marketing data.

Beyond the real and significant issue of data theft, many are concerned about the ability of an organization to evaluate and judge their personal lives. They seemingly feel victim to an unstoppable "peeping Tom," perpetually peering through the window.

The reality is different. Does anyone really believe Pepsi or Coca-Cola, marketing to millions worldwide, have an interest in evaluating each single individual's information? For that matter, do we really think that Uncle Sam cares about our Facebook posts?

The individual is there in the data, but remains essentially anonymous. For the government, the anonymity apparently only vanishes when potentially threatening patterns of communication are uncovered.

Historically, data scientists leveraged data sources to help businesses assess customer risk and to better anticipate customer needs. A company then could anticipate what to stock, who to focus on for customer retention, and what products to offer. In the past, a data scientist may simply have evaluated the products that you purchased over a year to determine what you -- along with those like you -- are most likely to buy next month.

Now, new data, tools and methods result in increased sophistication.

Data scientists can evaluate your social network to see who is typically first in your online group of friends to purchase a new product. They can track how subsequent purchases of that product waterfall across a social network. When combining this information with traditional analysis, marketing dollars can be focused on "first movers" in social networks. As a result, expenditures - or scrutiny of potential terrorists -- can be fine-tuned to the most relevant individuals.

While this is easily seen as intrusive, it's surprisingly impersonal.

To conduct this research, unstructured data sources (social network information, for example) are reshaped into a fixed and highly mathematical structure.

Electronic files of text and descriptions are transformed into numbers. Sophisticated software churns through mountains of records in a fast, algorithmic, impersonal fashion. There is no time or interest in looking at any individual record. Rather, the strategist reviews statistics at the end of the process to assign a score (again in an automated, impersonal fashion) to each record -- a score suggesting a marketing strategy for each type of customer or a score suggesting the need for more detailed study of potential security threats.

Is it possible to isolate individual records and review the information associated with a person? Yes, just as it's possible for a stranger to peer through your window. But "possible" doesn't mean "likely."

It's understandable that people feel uncomfortable knowing others have access to knowledge about them. But they should know that there are checks in place to manage permissible uses of data -- both at corporations and in government.

Courts oversee what the federal watchdogs are watching. At the corporate level, there are systems in place to monitor who is using data, and what they are viewing. Frankly, there is no interest in looking at any single individual. For a data scientist, that would just be boring.

*Alan K. Gorenstein is a data scientist and is senior vice president and head of Decision Sciences for the Royal Bank of Scotland. He has been involved in business strategy, marketing and analytics for more than 20 years.*