by Marc Kosciejew

The Perils of Big Data

Big Data is transforming many aspects of our lives. In "The Era of Big Data," I argue that Big Data "represents a transformative shift in processing information that, in turn, is beginning to change the way we approach and think about the world."[1] Big Data involves the great expansion of information coupled with new, cheap, and user-friendly technological devices that permit the efficient collection, management, and use of such massive amounts of data. Big Data is consequently making it "possible to explore and examine many things that previously could not have been done, let alone imagined, from preventing diseases to improving social conditions, monitoring personal habits, analyzing business trends, predicting terrorist activities, and so on; indeed, the list is seemingly endless."[2] But this opening up of new possibilities also reveals Big Data's dark side.

The Dark Side of Big Data

As Big Data begins to define our times, it is crucial that we recognize and understand not only its benefits but also its risks. Big Data presents many perils that may cause more harm than good. Handling Big Data responsibly and ethically requires knowledge and appreciation of its dark side: knowing the potential dangers will help us avoid or minimize them. This article aims to help shed light on Big Data's dark side by discussing three of its major perils—to privacy, freedom, and agency—in order to provide a more balanced analysis of this fêted phenomenon.

Imperiling Privacy

Big Data can imperil privacy. Big Data permits us to examine and measure things as never before. But many of these things are the most private and intimate details of our lives, thereby raising significant concerns about privacy infringements. Big Data has helped make tracking more common: it is now a cheaper, easier, quicker, and more useful practice, mainly benefiting companies, governments, and other corporate and public institutions (also known as data-collectors and/or data-holders).

Facilitating widespread, invasive, and seemingly ubiquitous surveillance in almost every sphere of activity—from shopping to driving to travelling to working to living at home—by collecting, generating, storing, using, and reusing data, much of it personal, Big Data's surveillance possibilities would shock George Orwell. It is indeed ironic that there are more than thirty surveillance cameras collecting huge amounts of data within two hundred yards of the London flat where Orwell wrote *1984*.

Further, the value of personal information lies not only in its primary reason for being collected, but also in its possible reuses. As Viktor Mayer-Schönberger and Kenneth Cukier note in *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, "with big data, the value of information no longer resides solely in its primary purpose . . . it is now in secondary [and more] uses."[3] Most reuses are not necessarily imagined, known, or planned when the data is first collected. When people are asked to provide their informed consent to the collection or generation of their personal information, it is typically for a specific and limited reason; however, the data-collector or data-holder that is providing the service or transaction can now store that information indefinitely for other purposes yet unknown, such as selling or leasing it to a third party. But how can you ask for, or give, informed consent for a possible, unimagined, unknown future purpose that is yet to exist? It remains financially, technically, and procedurally difficult for data-collectors or data-holders to find every individual to ask permission for each reuse.

Many people, moreover, remain unaware of actual or potential privacy breaches of their personal information. As Danah Boyd and Kate Crawford state in "Six Provocations for Big Data," many people "are not aware of the multiplicity of agents and algorithms currently gathering and storing their data for future use. Researchers are rarely in a [person's] imagined audience, neither are [people] necessarily aware of all the multiple

uses, profits and other gains that come from [their personal] information." Boyd and Crawford maintain that just because personal information is given, or may even be public in some form, this "does not simplistically equate with full permission being given for all uses."[4] In other words, people may agree to provide or make public their personal information for one specific and limited purpose without fully realizing that their privacy may be jeopardized if that data is reused.

In "Letting Down Our Guard With Web Privacy," moreover, Somini Sengupta discusses how many people foster false senses of security regarding many privacy policies' guarantees of control over personal information. Many of these policies shift responsibility —and therefore blame—onto individuals for whatever problems may arise from privacy violations, even though many of these problems are created or facilitated by the data-collectors or data-holders who have access to the information, especially in their reuse practices.[5] Nevertheless, such obliviousness is beneficial to data-collectors and data-holders, most of whom would rather treat personal information as an asset to use and reuse in various business deals, trades, transactions, operations, etc. to increase their profits or strategic advantages.

Imperiling Freedom

Big Data can imperil freedom. Steven Spielberg's 2002 film *Minority Report* portrayed a dystopic future of prediction-based punishments; chillingly, Big Data is, today, helping to create a similar reality. Big Data analysis can threaten individual freedom by judging people on algorithmic predictions of individual propensities or possible future behaviour. Predictive policing, for instance, is becoming a routine practice in many American police departments. These departments are using Big Data analysis to determine which neighbourhoods, groups, and people to scrutinize, target, and monitor after an algorithm indicates that they are more likely to commit or experience a transgression. Such prediction-based practices illustrate one of the many novel purposes to which Big Data is applied: in this case, crime prevention.

Proponents of Big Data's prediction-based profiling, judgment, and punishment argue that it helps increase safety, reduce delinquency, and prevent crime. They state that society already employs preventive measures against potentially reckless behaviours. For example, there are stringent regulations on smoking, rules requiring the wearing of seatbelts, security checks at borders and airports, physical tests for certain employment or healthcare considerations, and so on. These preventions constrain freedoms to varying extents, but many people consider them necessary inconveniences or commonsense practices that help foil greater harm.

Prediction-based policing, and by extension prediction-based profiling, judgment, and punishment, becomes dangerous, however, if Big Data analysis is used to predetermine one's guilt for behaviour that has not yet occurred. As Mayer-Schönberger and Cukier note, "this negates the very idea of the presumption of innocence, the principle upon which our legal system, as well as our sense of fairness, is based. And if we hold people responsible for predicted future acts, ones they may never commit, we also deny that humans have a capacity for moral choice."[6] Denying people their choices, thwarting their decision-making capabilities, and disabling their responsibility ultimately means a negation of their personal freedom and liberties.

Policing, judging, and punishing people based on probabilities, not real actions, nonetheless has perilous implications for freedom. In "More Data, More Problems: Is Big Data Always Right?" Ari Zoldan warns that "big data's algorithms may draw the wrong conclusions about who you are, and what you do. Digital marketers [and other corporate, political, and security players] are getting to know you more than you might think."[7] Further, Mayer-Schönberger and Cukier argue that the peril is much broader than policing and the judicial system, for "it covers all areas of society, all instances of human judgment in which big data predictions are used to decide whether people are culpable for future acts or not. Those include everything from a company's decision to dismiss an employee, to a doctor denying a patient surgery, to a spouse filing for divorce."[8] Without responsible information stewardship, Big Data threatens to fully realize Spielberg's prediction-based, freedom-denying world.

Imperiling Agency

Big Data can imperil agency by threatening to discard our control and decision-making in favour of numbers and statistics. We begin to fetishize information, creating a datafied world in which people increasingly rely on data to make minor and major decisions without much consideration for other contextual, circumstantial, casual, or quality-related factors.

Mayer-Schönberger and Cukier call this danger a "dictatorship of data," whereby data becomes a crutch for every decision as policymakers, businesspeople, politicians, and others gradually seek solutions only in more and bigger data. They argue that "we are more susceptible than we may think to the 'dictatorship of data'—that is, to letting the data govern us in ways that may do as much harm as good. The threat is that we will let ourselves be mindlessly bound by the output of our analyses even when we have reasonable grounds for suspecting something is amiss. Or that we will become obsessed with collecting facts and figures for data's sake. Or that we will attribute a degree of truth to the data which it does not deserve."[9] Our agency may therefore become usurped by a fixation on data: a situation in which we become enamoured of its power and promise, dependent on its solutions, and ignorant of its limitations.

Our agency is further jeopardized by Big Data's possible mistakes if we focus and rely too much on the data while ignoring its possible negative consequences. "When we botch the facts, our ability to create damage is greatly magnified because of our enhanced technology, global interconnectivity, and huge data sizes," Zoldan explains. "More often than not," he continues, "we are too trusting of statistics, and fail to examine the data with a critical eye. Oftentimes, we are quick to conclude that the data presented to us is factual, which is entering risky waters in the context of big data."[10] Big Data, in other words, can cause catastrophic miscalculations if the underlying data is poor, mis-analyzed, or used misleadingly or unwisely.

Combating Perils

To help minimize the perils of Big Data, more robust information policies are required. In "Big Data Is Opening Doors, but Maybe Too Many," Steve Lohr presents the three tenets of Alex Pentland's "new deal on data": you have a right to 1) possess your data, 2) control its uses, and 3) distribute it as you wish and dispose of it as you see fit. According to Pentland, "personal data . . . is like modern money—digital packets that move around the planet, traveling rapidly but needing to be controlled. 'You give it to a bank, but there's only so many things the bank can do with it,' [Pentland] says."[11] Information policies must incorporate these tenets by specifically recognizing you as the owner of your personal data and, consequently, as the authority on how to handle, use, save, share, and deploy it as you want.

Additionally, in "Data, Data Everywhere," Joseph Janes asks about humility and humanity in dealing with Big Data, arguing that information professionals who understand and appreciate Big Data's limitations are needed. He asks, "How about professionals who work to humanize this field? Those who think about questions of privacy, authority, quality, authenticity, rationality, and ethicality. Who center these processes in efforts to better the human condition and the lives of individuals. Who build tools to gyre and gimbal in the taffeta of data to find just the right thread for a person in need. Somebody like . . . a reference librarian."[12] Indeed, the era of Big Data presents exciting opportunities for established information professionals to shape and implement the much-needed new information policies that will affect our emerging Big Data interactions and relationships.

Big Responsibilities

Big Data can imperil privacy, freedom, and agency, especially if we fetishize it and rely too heavily on it. As Big Data becomes increasingly more influential in countless aspects of life, society, and institutions, it is crucial to recognize, understand, and address these perils. As Boyd and Crawford implore, "it is imperative that we begin asking critical questions about what all this data means, who gets access to it, how it is deployed, and to what ends. With Big Data come big responsibilities."[13] Approaching Big Data in a responsible manner requires answering these questions in order not only to better understand and appreciate, but also to address and combat its risks.

Marc Kosciejew (mkosciej@gmail.com) received his MLIS and PhD in Library and Information Science from Western University. In 2007 he conducted research in North Korea (Democratic People's Republic of Korea) on the secretive Communist state's library system, becoming one of the first English-speakers to present and publish on this specific topic. His current research interests include documentation science, records and information management, the intersections of society and technology, concepts and practices of information, and the history of libraries.

Notes

1. Marc Kosciejew, "The Era of Big Data," *Feliciter* 59, no. 4(2013), 52.
2. Ibid.

3. Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (New York: Houghton Mifflin Harcourt, 2013), 153.
4. Danah Boyd and Kate Crawford, "Six Provocations for Big Data," *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011*, accessed September 10, 2013, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431, 11.
5. Somini Sengupta, "Letting Down Our Guard With Web Privacy," *The New York Times*, March 30, 2013, accessed September 10, 2013, http://www.nytimes.com/2013/03/31/technology/web-privacy-and-how-consumers-let-down-their-guard.html?pagewanted=1&_r=0&ref=technology.
6. Mayer-Schönberger and Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, 162.
7. Ari Zoldan, "More Data, More Problems: Is Big Data Always Right?" *Wired*, May 10, 2013, accessed September 10, 2013, http://www.wired.com/insights/2013/05/more-data-more-problems-is-big-data-always-right/.
8. Mayer-Schönberger and Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, 162.
9. Mayer-Schönberger and Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, 166.
10. Zoldan, "More Data, More Problems: Is Big Data Always Right?"
11. Steve Lohr, "Big Data Is Opening Doors, but Maybe Too Many," *The New York Times*, March 23, 2013, accessed September 10, 2013, http://www.nytimes.com/2013/03/24/technology/big-data-and-a-renewed-debate-over-privacy.html?pagewanted=all.
12. Joseph Janes, "Data, Data Everywhere," *American Libraries*, May/June 2012, accessed September 10, 2013, http://www.americanlibrariesmagazine.org/article/data-data-everywhere.
13. Boyd and Crawford, "Six Provocations for Big Data," 2.

UNOFFICIAL WISDOM

Selected Contributions to Feliciter *1995-2009*

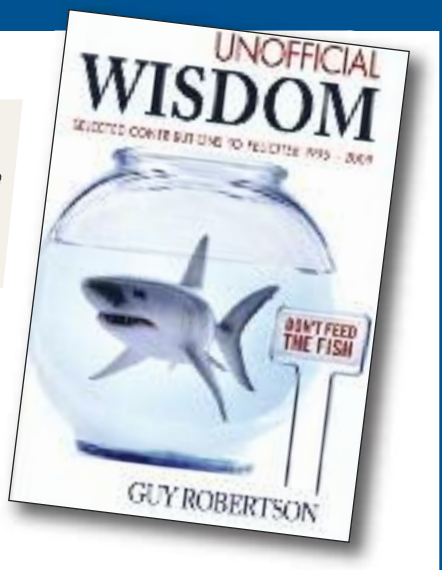
The reviews are in…

"I recommend this for libraries which have collections on library science already on their shelves, or staff collections promoting professional growth and development. It is useful for those wanting to browse through a wider range of library-based articles, and for those wanting to find a little professional enlightenment through serendipity. This book has the capacity to inform and enlighten library and information professionals with interesting stories, hidden gems and thought-provoking points."

Alison Fields, Senior Lecturer
Information and Library Studies, Open Polytechnic Kuratini Tuwhera,
LIANZA – The New Zealand Library & Information Management Journal

"Written in an easy and, at the same time very informative style, the articles made me think that the book would be a great preparation for many tests and examinations for those taking on the wonderful world of librarianship."

Jean Orpwood
ELAN – Ex Libris Association Newsletter



Discover for yourself why Guy Robertson, highly respected *Feliciter* columnist, for the past 16 years, continues to delight and share his "unofficial wisdom" with readers. Order your copy at: online at **www.cla.ca** at **Shop CLA** or **orders@cla.ca**

CLA

UNOFFICIAL WISDOM, Selected Contributions to* Feliciter *1995-2009
Guy Robertson
ISBN: 978-0-88802-334-6 • Publisher: Canadian Library Association • Price: CLA member: $26.95 (non-member: $29.95)