by Marc Kosciejew

The Individual and Big Data

Big Data is transforming the individual. Many aspects and features of personal lives are being datafied; that is, turned into bits of data by public and private entities, as well as by individuals themselves, in order to record, monitor, track, observe, and learn about our personal actions, behaviours, emotions, and preferences. This article explores this transformation by examining the data-driven life, self-tracking, data exhaust, and (the need for) data rights. The aim is to start a conversation on the meaning of Big Data for us as individuals, and on our interactions and relationships with the digital world and, ultimately, our (digital) selves.

The Data-Driven Life

Until recently, attempts to understand one's individuality, such as trying to attain self-knowledge, self-improvement, or self-actualization, was the domain of introspection and psychoanalysis. The main method of self-tracking was to write down personal observations and in some cases talk about them. But, as Gary Wolf argues in "The Data-Driven Life," four things changed in the past decade or so: "first, electronic sensors got smaller and better. Second, people started carrying powerful computing devices, typically disguised as mobile phones. Third, social media made it seem normal to share everything. And fourth, we began to get an inkling of the rise of a global superintelligence known as the cloud."[1] These four changes fundamentally altered the ways in which people engaged in attempts to understand themselves. Today, data is "infiltrating the last redoubts of the personal. Sleep, exercise, sex, food, mood, location, alertness, productivity, even spiritual well-being are being tracked and measured, shared and displayed."[2] The individual and (ideas of) individuality are being transformed into quantifiable, statistical, and objective data points that can (allegedly and supposedly) help reveal new possibilities, novel insights, and hidden facts about ourselves. A data-driven life, in other words, is datafying the individual.

Self-Tracking

This individual datafication is best expressed through the self-tracking, or self-quantification, movement in which individuals datafy their lives and, in turn, use those data points to record, measure, compare, and share themselves with third parties such as friends, interest groups, companies, governments, and so on. Wolf argues that "when the familiar pen-and-paper methods of self-analysis are enhanced by sensors [and other digital devices and services] that monitor our behavior automatically, the process of self-tracking becomes both more alluring and more meaningful." Further, these technologies "also remind us that our ordinary behavior contains obscure quantifiable signals that can be used to inform our behavior, once we learn to read them."[3] In *The Virtual Self*, Nora Young similarly argues that this data-driven life can reveal novel or unexpected information about ourselves. She states that "while it seems trivial—silly, even—what is remarkable about self-tracking is the surprising power it offers. At the personal level, it can help us change Mr. Mohammed Taher behaviour, but it also offers insight, providing us with an undeniably clear picture of how we behave."[4]

Young provides a detailed description of the self-tracking movement, including the characteristics, motivations, and neuroses of self-trackers. She admits that "what would once have been an unwieldy and bizarre commitment to personal information-gathering is now a casual and painless process;"[5] additionally, with the help of Big Data devices, technologies, and services, this process is resulting in many people "keeping track of the statistical minutiae of daily life, leading lives that are increasingly numerically documented."[6] It is also becoming increasingly common for self-trackers to share their self-tracking practices and results with others on smartphone applications (there's an app for that!), social media networks, interest forums, and other similar specialized online services.

Data Exhaust

Yet, self-tracking and other online activities generate data exhaust. According to Viktor Mayer-Schönberger and Kenneth Cukier in *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, "a term of art has emerged to describe the digital trail that people leave in their wake: 'data exhaust.' It refers to data that is shed as a byproduct of people's actions and movements in the world."[7] Further, "for the internet, it describes users' online interactions: where they click, how long they look at a page, where the mouse-cursor hovers, what they type, and more."[8] Like a vehicle spewing out pollution into the natural environment, individuals are sputtering out data exhaust into the digital environment, unexpectedly and unknowingly leaving behind personal information across a variety of databases, devices, platforms, and services.

Data exhaust can be used by any company or agency that can gather it. In "We Post Nothing about Our Daughter Online," Amy Webb further illuminates some of data exhaust's possible (unknown and unexpected) uses by third parties. She describes how many people willingly share personal information about their children online. Using the example of her friends' daughter "Kate," she warns that "with each photo upload, Kate's parents are, unwittingly, helping Facebook [and other similar social media networks] to merge her digital and real worlds. Algorithms will analyze the people around Kate, the references made to them in posts, and over time will determine Kate's most likely inner circle [. . .]. The problem is that Facebook is only one site. With every status update, YouTube video, and birthday blog post, Kate's parents are preventing her from any hope of future anonymity."[9] The more personal information an individual records, measures, tracks, shares, and uses online, the more data and data exhaust third parties have at their disposal to use and reuse today and in the future.

E-book reader devices, for example, consume huge amounts of data on the individuals who use them. This data constructs a profile of each reader's literary preferences and habits: "how long they take to read a page or section, where they read, if they turn the page with barely a skim or close the book forever. The devices record each time users underline a passage or take notes in the margins." Then, "once aggregated, the data exhaust can tell publishers and authors things they could never know before in a quantifiable way: the likes, dislikes, and reading patterns of people. This information is commercially valuable. One can imagine e-book firms selling it to publishers to improve the content and structure of books."[10] Reading, in other words, is transformed from an individual solitary act into a datafied, quantifiable, and monitored experience to be analyzed and used by third parties.

As Ronald J. Deibert argues in *Black Code*, "in a very real sense we no longer move about our lives as self-contained beings, but as nodes of information production in a dense network of digital relations involving other nodes of information production. All of the data about us as individuals in social network communities is owned, operated, managed, and manipulated by third parties beyond our control, and those third parties are, typically, private companies."[11] Let us now briefly examine three cases—Acxiom, Google, and the National Security Agency (NSA) — to draw attention to how individuals are affected by their personal data's exploitation.

Acxiom

Data brokerage firms are information companies that capture, collect, store, use, analyze, and sell personal information typically gleaned from consumer data to other parties, including banks and financial institutions, credit card companies, insurance companies, and retail services. In "A Data Broker Offers a Peek Behind the Curtain," Natasha Singer profiles Acxiom, one of the most prolific data brokerage firms, which "has amassed details on the household makeup, financial means, shopping preferences and leisure pursuits of a majority of adults in the United States."[12] While data brokerages are typically secretive enterprises, Acxiom recently created a service allowing people to access, review, correct, and/or delete the data that the company currently has about them. To access and review this personal information, however, individuals must divulge even more data about themselves by undergoing an identity verification process, which includes submitting their name, their address, and the last four digits of their Social Security Number.

As Singer notes in "Getting a Glimpse of Your Own Marketing Data Online," "some privacy and security analysts said they were uncomfortable with an identity verification system that requires consumers to give personal details like their birth dates and partial Social Security numbers."[13] There are also various possible security flaws with such a service; for instance, "family members like ex-spouses—as well as employees of banks and credit card companies—may have access to the same personal identity details about consumers and could use that information to impersonate them on the site."[14]

Google

Mayer-Schönberger and Cukier argue that Google is the undisputed leader in exploiting both personal information and, especially, the resulting data exhaust. They explain how every interaction an individual has with Google services is captured, analyzed, and fed back into the system to be reused by the company's algorithms to improve existing features, develop new ones, and personalize search results. They show, for example, how "Google is acutely aware of how many times people searched for a term as well as related ones, and of how often they clicked on a link but then returned to the search page unimpressed with what they found, only to search again. It knows whether they clicked on the eighth link on the first page or the first link on the eighth page—or if they abandoned the search altogether."[15]

Young also explains Google's data exhaust advantage, stating that "I'm not trying to help Google or my fellow computer users when I search, but [. . .] the byproduct of my searching is an improvement in the speed of search results for all."[16] Further, "because one is searching from a computer [. . . with] an IP address, Google can tell where searches are likely coming from. Once the volume of searches is large enough to be statistically significant, it can be very revealing of what's going on in the real, physical world."[17] Google, in other words, not only improves and expands its corporate advantages with an individual's various uses of its search engine and other services, but also develops and stores a detailed personal profile of that same individual for unknown and future reuses.

Google additionally partners with other actors in the management and use of all of this personal information. Deibert describes how Google works closely with the NSA in order to help improve and maintain its security apparatus. Thus, the world's largest data collection company has a close relationship with arguably the world's most powerful spy agency and, according to American courts, no one outside of either institution has the right to know the details of this arrangement.[18]

The NSA

Many powerful security agencies collect, store, and analyze the personal information that individuals willingly disclose and the trails they unwittingly leave behind. The chilling revelations of the NSA surveillance program, for example, show how this personal information and resulting data exhaust can be used in the surveillance of individuals; for example, James Risen and Laura Poitras recently reported on how the agency gathers data on the social connections of American citizens. They describe how it "has been exploiting its huge collections of data to create sophisticated graphs of some Americans' social connections that can identify their associates, their locations at certain times, their travelling companions and other personal information." It also augments this information with so-called enrichment data, that is, "material from public, commercial and other sources, including bank codes, insurance information, Facebook profiles, passenger manifests, voter registration rolls and GPS location information, as well as property records and unspecified tax data."[19]

All of this data allows NSA analysts not only to learn about individuals but also "to identify people's friends and associates, detect where they were at a certain time, acquire clues to religious or political affiliations, and pick up sensitive information like regular calls to a psychiatrist's office, night messages to an extramarital partner or exchanges with a fellow plotter."[20] Thus, the agency can paint a rich picture of an individual's life, creating an almost-complete model of his or her actual, and probable, daily routines, habits, preferences, purchases, communications, interactions, and relationships.

Data Rights

In "Better Engineering, and Better Laws," Aleecia McDonald observes that "most people want what a data-driven future can provide, but we have learned the hard way that we cannot trust companies or governments to exercise basic decency and restraint in collecting our data."[21] Robust rules and regulations are required in order to protect individuals' rights to their privacy as well as their rights to access to and ultimate ownership of their personal information regardless of how it is wittingly or unwittingly disclosed or shared. We, as individuals, have particular data rights to our personal information that must be recognized, respected, and enforced.

Let us conclude with Alex Pentland's "new deal on data" to highlight our data rights:[22] first, you have a right to possess your data (you can open an account, deposit, and remove data whenever you like); second, you must have full control over your data's use (everything must be opt-in with regular reminders that you can opt-out); and third, you have a right to distribute or dispose of your data at your pleasure. Personal data, after all, is an extension of the individual; that is, your data is yourself.

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. Gary Wolf, "The Data-Driven Life," *The New York Times*, April 28, 2010, accessed November 12, 2013, http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html?pagewanted=all.
2. Ibid.
3. Ibid.
4. Nora Young, *The Virtual Self: How Our Digital Lives Are Altering the World Around Us* (Toronto: McClelland & Stewart Ltd., 2012), 3.
5. Ibid, 2-3.
6. Ibid, 2.
7. 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), 113.
8. Ibid.
9. Amy Webb, "We Post Nothing about Our Daughter Online," *Slate*, September 4, 2013, accessed November 12, 2013, http://www.slate.com/articles/technology/data_mine_1/2013/09/facebook_privacy_and_kids_don_t_post_photos_of_your_kids_online.html.
10. Mayer-Schönberger and Cukier, *Big Data*, 114.
11. Ronald J. Deibert, *Black Code: Inside the Battle for Cyberspace* (Toronto: McClelland & Stewart, 2013), 63.
12. Natasha Singer, "A Data Broker Offers a Peek Behind the Curtain," *The New York Times*, August 31, 2013, accessed November 12, 2013, http://www.nytimes.com/2013/09/01/business/a-data-broker-offers-a-peek-behind-the-curtain.html.
13. Singer, "Getting a Glimpse of Your Own Marketing Data Online," *The New York Times*, September 4, 2013, accessed November 12, 2013, http://bits.blogs.nytimes.com/2013/09/04/getting-a-glimpse-of-your-own-marketing-data/.
14. Ibid.
15. Mayer-Schönberger and Cukier, *Big Data*, 113.
16. Young, *The Virtual Self*, 128.
17. Ibid.
18. Deibert, *Black Code*, 62.
19. James Risen and Laura Poitras, "N.S.A. Gathers Data on Social Connections of U.S. Citizens," *The New York Times*, September 28, 2013, accessed November 12, 2013, http://www.nytimes.com/2013/09/29/us/nsa-examines-social-networks-of-us-citizens.html.
20. Ibid.
21. Aleecia McDonald, "Better Engineering, and Better Laws," *The New York Times*, September 8, 2013, accessed November 12, 2013, http://www.nytimes.com/roomfordebate/2013/09/08/privacy-and-the-internet-of-things/laws-can-ensure-privacy-in-the-internet-of-things.
22. Alex Pentland, "Reality Mining of Mobile Communications: Toward a New Deal on Data," in *The Global Information Technology Report 2008-2009: Mobility in a Networked World* (World Economic Forum, 2009), accessed November 12, 2013, http://www.insead.edu/v1/gitr/wef/main/fullreport/files/Chap1/1.6.pdf.