CrossMark

Making Data Valuable: Political, Economic, and Conceptual Bases of Big Data

Anna Lauren Hoffmann[1]

The ascendency of so-called *big data* as a driving technological, economic, and political force has hinged, in part, on an understanding of people's social and behavioral data as *valuable*. Sometimes, this value is simply taken as self-evident—that is, more data *just does* lead to better knowledge, keener insights, and (for those who can harness it) social and economic gain. Other times, personal data is metaphorically positioned as a kind of natural resource—the data is simultaneously "the new oil" and something to be "mined"—that fuels scholarly and economic progress alike (Puschmann and Burgess 2014). Still further, data's value is necessarily implicit in debates around the kinds of expertise (human and machinic, centralized or distributed) big data demands (Bassett 2015).

Of course, uncritical claims to the value of personal data ultimately occlude the legal and economic structures, material conditions, and conceptual assumptions that make the capture and exploitation of digital data possible in the first place. Conceptually, for example, big data (including, but not limited to, large-scale social and behavioral data generated by and through our interactions with networked devices and online platforms) presents itself as scientifically and politically neutral, imbued with an "aura of truth, objectivity, and accuracy" (Boyd and Crawford 2012, p. 663). However, accepting this claim—that is, accepting the idea that, with big enough data, the social world can be explained from a value-neutral, objective point of view (Jurgenson 2014, n.p.)—requires willful ignorance of the theoretical moves required to conceive of something as "data" in the first place (Bowker 2014). Further, it also requires the erroneous equating of knowledge with the results of automated statistical analyses of massive datasets—as if human labor and expertise were not integral to the entire enterprise of knowledge production. Certainly, advanced techniques and technologies may be part of the knowledge production process, but they are—as Floridi (2012, p. 437) notes—insufficient by themselves.

✉ Anna Lauren Hoffmann
alho@uw.edu

[1] The Information School, University of Washington, Seattle, WA, USA

Springer

Moreover, there is nothing given or natural about viewing people's personal data as property or fuel for economic gain (Litman 2000)—as opposed to, say, treating it as an intimate part of a person's identity or being (Floridi 2005). Similarly, making data valuable is contingent upon our ability to transform subjects into something that is readily quantifiable—something that can be easily fed into the machines that convert data into meaningful information or actionable insight. But doing so means also accepting, too often without complication, the computational reductionism and decontextualization inherent in quantifying people, identities, and behaviors (Manders-Huits 2010). Consequently, big data continues to suffer from "blind spots and problems of representativeness, precisely because [they] cannot account for those who participate in the social world in ways that do not register as `digital signals" (Crawford et al. 2014, p. 1667).

To be sure, these challenges and limits do not automatically mean that large-scale, data-intensive research is necessarily bad or unimportant. Rather, they simply underscore the continued relevance of theoretical and other types of inquiry during and after the big data revolution. Demystifying data requires close political and philosophical attention to the structures, conditions, and assumptions that make the generation, collection, and exploitation of massive sets of personal data—that is, data about people and their individual and social behaviors—possible. As Tom Boellstorff (2013, n.p.) rightfully asserts: "There is a great need for theorization precisely when emerging configurations of data might seem to make concepts superfluous—to underscore that there is no Archimedean point of pure data outside conceptual worlds".

Importantly, this theorization must take into account the industries, infrastructures, and methods that have emerged to grapple with (and, often, capitalize on) the massive amounts of data generated by and through our interactions with connected devices (Floridi 2012). After all, data and the systems and tools that support their production and use do not, to repurpose part of Star and Ruhleder's (1996, p. 113) description of infrastructure, "grow de novo." Rather, they are generated by and through existing tools, methods, and practices and, further, are framed by the political and economic contexts out of which they emerge. Rather than transcending the material and the political, "big data" is firmly mired in the people and tools that make it possible. These technical, practical, and contextual dimensions are, as Floridi and Taddeo (2016, p. 4) write, "...obviously intertwined."

It is on this difficult terrain of intertwined legal, economic, practical, and material dimensions of large-scale personal social and behavioral data that the contributions to this special section are situated. Individually and combined, they represent a formidable contribution to our understanding of the political and technical arrangements that help situate data as something of economic or conceptual value.

The first two papers grapple with political economic dimensions of ever-expanding troves of personal and personally-identifiable data. In *The Biopolitical Public Domain*, Julie Cohen details the legal and technical choices that have made possible the generation and exploitation of massive amounts of personal data, helping move us toward a deeper understanding of the legal and political foundations of our "big data" economy. Weaving together insights from both liberal political economy and Foucauldian biopolitics, Cohen tracks and theorizes the mechanisms—from early design choices of the commercial Web to the (in)action of lawmakers and regulators—that have allowed personal information to be conceptualized as a kind of natural resource,

one that is both "raw" and valuable. This conception is, of course, artificial; Cohen shows how "raw" data are elicited in carefully standardized ways, thus undermining characterizations of the data science process as "protean and dynamic." Further, she shows how these choices have shaped practices of data collection and exploitation in both developed and developing nations. For Cohen, the biopolitical public domain is one that, increasingly, encompasses the entire globe, enabling the statistical construction and management of entire populations.

In *Self-Tracking Practices and Digital (Re)Productive Labor*, Karen Dewart McEwan employs Marxist and feminist political economic frames to show how commercial products that allow individuals to track personal biometric and activity data (as, for example, through fitness or productivity tracking devices) enable both the exploitation and reproduction of subjects under capitalism. For McEwan, these products and practices not only capture data that can be exploited by technology companies and data brokers alike, but also work to continually (re)produce the sorts of capitalist subjectivities required if data is to be conceived of as valuable and commodifiable. In addition, she shows how this process is aided and abetted by discourses of self-discovery and self-knowledge that obscure the labor of self-tracking by casting it as, instead, a kind of personal betterment or enlightenment. In pulling together both the practical and discursive, McEwan reveals self-tracking (including, but not limited to, personal fitness and productivity tracking) as a kind of socially necessary reproductive labor that disciplines and cultivates subjects in ways that are amenable to contemporary capital accumulation—thus enabling their continued exploitation, even outside of any official labor relationship.

The final paper—*Data Science as Machinic Neoplatonism*—moves away from the political economic and toward the metaphysical. In the piece, Dan McQuillan connects contemporary data science with the Neoplatonic, two-world metaphysics that informed the early science of, among others, Copernicus and Galileo. Data science, McQuillan argues, is able to—via quantification, statistical methods, and the presumed authority of data scientists—trade on the Neoplatonic claim to revealing a hidden (mathematical) layer of reality while simultaneously evading such strong language. Instead, data science substitutes computation for mathematics and correlation for causation while still claiming to stand outside of—or apart from—the world it observes or manipulates. This move toward datafication, like mathematization before it, positions "data" as ontologically superior—put another way, it positions data as valuable precisely because supposedly exists beyond context or bias. Different from the first two papers—which make explicit the political and economic arrangements that underwrite the large-scale production of personal data—McQuillan instead shows how scientist ideology and problematic epistemological claims obscure data's social and material politics.

Though the era of "big data" has been upon us for some time, it is imperative that we continue to resist and complicate uncritical claims to the value of personal data. To this end, none of the papers are content to leave us only with diagnoses and critique. Instead, each one ends looking forward, suggesting ways in which we might reimagine or resist the politics, practices, or economics of data exploitation. For Cohen, moving forward means conducting policy discussions in full view of the critical, political, and technical choices that have made possible the commercial exploitation of massive amounts of data about individuals. For McEwan, surfacing the mechanisms and discourses by which people and their data are made to "work" under capitalism is

vital to the development of effective resistance. And, finally, for McQuillan, liberating ourselves from data science's Neoplatonic chokehold requires the cultivation of an effective counterculture of data rooted in feminist and critical epistemologies and philosophies of science. Only then might we effectively confront the big data's myths and begin to chart new directions for data science's future.

References

Bassett, C. (2015). Plenty as a response to austerity? Big data expertise, cultures and communities. *European Journal of Cultural Studies, 18*(4–5), 548–563.

Boellstorff, T. (2013). Making big data, in theory. *First Monday, 18*(10), n.p. https://doi.org/10.5210/fm.v18 i10.4869.

Bowker, G. (2014). The theory/data thing. *International Journal of Communication, 8*, 1795–1799.

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication and Society, 15*(5), 662–679.

Crawford, K., Miltner, K., & Gray, M. L. (2014). Critiquing big data: politics, ethics, epistemologies. *International Journal of Communication, 8*, 1663–1672.

Floridi, L. (2005). The ontological interpretation of informational privacy. *Ethics and Information Technology, 7*(4), 185–200.

Floridi, L. (2012). Big data and their epistemological challenge. *Philosophy & Technology, 25*(4), 435–437.

Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions A, 374*(2083), 1–5.

Jurgenson, N. (2014). View from Nowhere. Online Essay. *The New Inquiry*. http://thenewinquiry.com/essays/view-from-nowhere/. Accessed 16 April 2016.

Litman, J. (2000). Information privacy/information property. *Stanford Law Review, 52*(5), 1283–1313.

Manders-Huits, N. (2010). Practical versus moral identities in identity management. *Ethics and Information Technology, 12*(1), 43–55.

Puschmann, C., & Burgess, J. (2014). Metaphors of big data. *International Journal of Communication, 8*, 1690–1709.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: design and access for large information spaces. *Information Systems Research, 7*(1), 111–134.