DATA, NOW BIGGER AND BETTER!

EDITED BY TOM BOELLSTORFF AND BILL MAURER

Data, Now Bigger and Better!

Edited and with an Introduction by
Tom Boellstorff and Bill Maurer

Data, Now Bigger and Better!

Contents

Introduction

Tom Boellstorff and Bill Maurer

This pamphlet brings together five authors responding to current debates regarding the new salience of big data in society. By big data, we refer to the mobile and digital computational systems that permit the large-scale generation, collection, and analysis of information about people's and devices' activities, locations, and transformations. We refer as well to as the social and technical effects of those systems and data, and the speculative hype, hopes, and futures that accompany them. While our individual responses address different aspects of these debates, four guiding principles link them together.

First, we challenge the strong tendency for discussions regarding digital technology to be shaped by obsessions with "trending." The value of an analysis becomes measured by claims about the future rather than accuracy in explaining the past or present. From public health to climatology, trying to predict the

future is not always wrongheaded—it is good to act on signals indicating the outbreak of a disease, or to model what might happen to coastal cities as more carbon is dumped into the air. But with regard to the matter at hand, focusing on the unknowable often diverts critical attention from the emerging present. This is, for instance, why we refer to "big data" in this pamphlet, despite attempts by many commentators to claim the phrase is already antiquated. In our view the planned obsolescence of terminology leads to term-coinage standing in for insight. No term ever perfectly captures its referent. "Big data" helpfully denotes a set of sociotechnological facts and connotes an atmosphere of simultaneous hype and legitimate innovation that represents the subject of our collective inquiry.

Second, we challenge the cultures of expertise that characterize discussions of big data. The move to data is often characterized as a move away from narrative, from ethnography, from the qualitative and interpretive. But as many coders and number-crunchers themselves assert, we cannot treat data as a purely quantitative phenomenon. Our analyses are interpretive and critical in character, yet shaped by a range of deep engagements with data. We are located in both academia and industry, and see the complicities and collaborations that emerge from that double location as a source of conceptual strength.

Third, all of our contributions to this pamphlet draw on anthropological perspectives, but not specifically ethnographic ones. As Tim Ingold has argued in his aptly titled article "Anthropology Is Not Ethnography," sociocultural anthropology is more than a basket of ethnographic methods like participant observation and interviewing. The value of anthropology lies also in its conceptual frameworks, frameworks

that are comparative as well as field-based. These theoretical contributions of anthropology date back more than a century, and turning to these classic debates provides an effective counter to the future-oriented hype and speculation so characteristic of discussions regarding big data. In this pamphlet, names like Malinowski, Lévi-Strauss, and Mauss appear as centrally as those of present-day thinkers. Discussions of kinship, exchange, and even cooking shape our analyses.

Fourth, our contributions are shaped by what our colleague Carl DiSalvo has termed "speculative civics"—which in his analysis links the emphasis on unknowable futures mentioned above with the emergence of entrepreneurialism and precarity in a public life shaped by data technologies not always open to scrutiny. The public is important, in both its classic and its newly emergent political senses. This is why we attend to activists like Edward Snowden and public intellectuals like Jaron Lanier, as well as broader debates over the nature of data and the meaning of "economy."

Through these guiding principles, we seek to contribute to a vitally important debate over the relationship between data, power, and meaning in the contemporary world. Big data experts often speak of the "three Vs" when characterizing crucial parameters for understanding what makes data "big": volume, variety, and velocity. In our view, "three Rs" might prove more significant: relation, recognition, and rot. We emphasize how data is formed through relations that extend beyond "data" itself; how what counts as data (and data's referent) is a social process with political overtones; and that data is always in real-time transformation in ways that cut across notions of nature and

culture. Geoffrey Bowker and Lisa Gitelman remind us, to quote the title of the latter's recent book, that *"Raw Data" Is An Oxymoron* (2013). But like fruit and food, data can be transformed by decomposition as much as by "cooking it up" for analysis.

Our pamphlet begins with Genevieve Bell's "The Secret Life of Big Data." Drawing on her experiences as an anthropologist and Vice President & Intel Fellow at Intel Labs, Bell playfully anthropomorphizes and personalizes "big data" so as to unpack its meanings and implications from an anthropological perspective. In the chapters that follow, each author builds from Bell's provocations to explore aspects of big data through the guiding principles discussed above.

Nick Seaver explores claims that big data will submerge other modes of social inquiry, invoking Chris Anderson's influential argument that the "data deluge" renders qualitative research and critical theory moot. To do so, he turns to other moments when computational formalism seemed to overwhelm social analysis, from Bronislaw Malinowski's criticisms of the "bastard algebra" that kinship studies in anthropology had become by 1930, to Clifford Geertz's complaints against the cognitive anthropologists of the 1960s and 1970s. Seaver then draws on the work of Marilyn Strathern to examine how the very ways in which methods are taken to "relate" (in both a cognitive and kinship sense) have consequences for anthropological analysis and knowledge production more broadly.

Mel Gregg begins from the observation that the word "data" in English derives from the same Latin root as "gift," "that which is given." But, of course, data are never given in advance. Asking "what is big" and "what is data" pose problems of scale, scope, the epistemological claims involved in all the hype about

big data—hype which often takes the form of visual representations, infographics and the like. Gregg develops notions like "data sweat" to explore data's agency, porosity, and relation to social and ecological environments of data "trash" and "exhaust." She reminds us, too, that the colonial impulse in big data's technological infrastructures, conditions of production and modes of knowledge mean that the gift never comes for free.

Bill Maurer takes up the internet guru and critic Jaron Lanier's call for a new economy of big data, one that would reward people for the collection and use of "their" data. While Lanier's proposal is compelling, Maurer argues that it sidelines key questions regarding the nature of property and person, issues of longstanding anthropological concern. Drawing an analogy with the dawn of new reproductive technologies, Maurer asks whether we are witnessing a similar new kind of birth. He extends this line of thinking by drawing on the anthropological canon of kinship and exchange theory, asking what other relationships come into view if we treat "big data" as party to a marriage exchange, and complicate the gifts we imagine data brings with some of the bastard algebra Seaver dissects.

Tom Boellstorff seeks to open a conversation regarding the theoretical frameworks that shape the notion of "big data," despite the fact that the very term is often taken to imply a pre-theoretical or even non-theoretical perspective on the world. In his chapter (a substantially revised version of an essay that appeared in the journal *First Monday* in October 2013), Boellstorff first explores the history of the notion of "data," but also the way in which a sense of being "dated" haunts the study of internet-related technology and society.

He then turns to the notion of "metadata," showing how distinctions between data and metadata are social (and political) rather than a priori. He then develops the notion of "the dialectic of surveillance and recognition" to examine how social relationships to big data are shaped by desire as well as fear. Finally, he reframes the distinction between "raw data" and "cooked data" by introducing notions of "rotted data" and "thick data."

Our collaboration is shaped by a dense network of social and intellectual relationships stretching back two decades, but made possible in its current incarnation by the Intel Science and Technology Center for Social Computing (ISTC-Social). The ISTC-Social is an example of some of the new collaborations anthropologists are forging in hybrid industrial-academic settings. Bill Maurer and Scott Mainwaring have written for the *Journal of Business Anthropology* about the odd economic and intellectual configurations that brought the ISTC-Social into being. But this pamphlet is not "about" the collaboration so much as a demonstration of what such a venture can do. We thank in particular Paul Dourish for his support and insights in that regard, and have found the conversations and collaborations made possible by this center to have helped us advance the arguments we make throughout this pamphlet.

The Secret Life of Big Data

Genevieve Bell

"Big data" occasions a big question: How do we start talking about the socio-technical imagination? How do we even begin, given that technological history is littered with countless space-age visions for the future just around the corner that never came to pass? It is a perilous time to think about big data: business interest is at a peak, but has it already crested? Popular debate hovers between concerns over privacy and government snooping, and indifference about the surrender of vast quantities of information just by turning on one's phone. The very term big data may well disappear by the time you read this pamphlet, much as the prefix "cyber" to refer to anything internet is quickly vanishing.

There are big arguments about how to define big data. It is also already occupying a huge place in the landscape of what technology is, what it might offer, and what it could be. When this chapter was still a spoken lecture I could remark, as if it were a new thing,

"We already see big data appearing in newspaper headlines." We're beyond the point where that generates surprise. Big data now bubbles up to the surface in all sorts of places. It is already finding its space in a cartography of social imagination. An advertisement for a technology company can sport a stylized street sign indicating we're at the corner of Big Data and the Cloud. At least for a certain market, big data literally points to a destination.

Of course as an anthropologist my first response to this moment is to ask, "yeah, but is it really new?" In 1085 A.D. William the Conqueror surveyed his empire and said, "I don't actually know what everyone is doing. Could you make me a list? I'd like to know what all the men in the kingdom are up to and how many sheep they have." The Domesday Book became an entire way of framing what it meant to be in England under William the Conqueror. It is effectively big data. It may not have been connected to the internet, but it was searchable. There were people who were custodians of that book. You could go to them and say, "I'm looking for men who have more than a 100 sheep and less than 200 sheep who haven't given me a lot of money this year." And there was someone whose job it was to run the algorithm on the data to find the men who fit that category. The Domesday Book involved three quite distinct things:

- The survey—which collected information, facts;
- The Winchester Roll—that was created out of the facts;
- The "day of judgment"—when taxes, ownership and military obligations were determined for hundreds of years to come.