

Ethics of Big Data



RICHARD MATEOSIAN

..... In this issue, I look at a book that anyone who has ever done anything online should read.

Ethics of Big Data by Kord Davis with Doug Patterson (O'Reilly, 2012, 82 pp., ISBN 978-1-4493-1179-7, \$19.99)

Kord Davis played with technology from an early age, and he loved learning its underlying principles and mechanisms. He chose to study philosophy because it gave him the best of both worlds: rigorous analysis and uncovering the way things work. Now, with a philosophy degree from Reed College, he advises high-tech companies about how to align their business practices with their values and principles. Doug Patterson teaches business ethics. His discussions with Davis led to many of the ideas in this book.

For many years society has struggled with the implications of gathering and analyzing personal data. The vastly increased speed and quantity of such activities have created a qualitatively new situation, generally known as *big data*. Some of the hardest questions businesses face today arise out of big data. For example, with enough information about your environment, someone can know a lot about you without knowing your name or Social Security number. And, by combining such troves of information from disparate sources, an organization can build an intrusive dossier

about you without actually violating the well-intentioned privacy policies of the organizations that originally collected the data.

Davis defines the forcing function of big data as the push, whether we like it or not, "to consider serious ethical issues including whether certain uses of big data violate fundamental civil, social, political, and legal rights." Companies were once thrilled at the prospect of knowing which color car is purchased most often in Texas in the summertime. Now they can know how much toothpaste your family bought from them last year. Does it matter if Target knows you are pregnant before your husband does? What if your boss does? Considering such issues falls outside the usual business discussions, even at the strategic level. Davis presents a framework for such consideration. Because each business situation is different, he includes both abstract and specific elements.

Davis begins with the basic concepts: identity, privacy, ownership, and reputation. These concepts are central to many ethical discussions, so he wants to establish their definitions and scopes. Identity concerns the relationship between our online and offline lives. Privacy issues boil down to who should control access to data. Ownership concerns not just who owns collected data but which rights can be transferred and what obligations collecting or receiving

such data entails. Reputation is about what you can trust. Big data provides both sources of error and checks and balances.

In addition to these abstract elements, each business has its own values and principles. Big data is ethically neutral—that is, the ethical questions arise from aligning organizational and societal values with organizational actions. Davis presents a system for carrying out this alignment. He believes that by gaining competence in this area, organizations can help to shape public opinion, not just react to it.

Many people are concerned with their right to privacy. Davis feels that using the term "privacy rights," because of its connotation of absoluteness, can prejudice some of the ethical issues that arise as organizations try to align their policies with their values. He prefers the term "privacy interest," which covers the gamut from no interest to absolute rights. That is, a right is the strongest kind of interest.

Similarly, Davis avoids the term "personally identifying information," because the technological limitations that define that term are constantly changing. What was once not considered personally identifying can, with new technology, be easily associated with a person. Davis uses the term "personal data" to refer to any data generated in the course of a person's activities. Digital transactions in business or social areas capture

related information distinct from the transaction data, and by this definition, all of it is personal data.

Davis contrasts the digital and non-digital situations. If you show someone a picture at a party, the event leaves little residue. Post the same picture on your Facebook page, and the permanent record includes plenty of ancillary personal information. Deciding which personal information is ancillary and which is not has an ethical impact. Davis advocates explicitly and transparently evaluating that impact. Doing so starts with articulating organizational values. Values can change over time, and trying to align values with actions can lead to reconsidering those values. Thus, Davis defines an ethical decision point as a cycle, iterated indefinitely, consisting of the following activities: inquiry, analysis, articulation, and action.

Inquiry means discovering and discussing the organization's core values. Sometimes organizations have one set of values in their founding documents or public relations literature but show by their actions that they have a different set of unstated values. Inquiry aims at discovering the organization's actual values. An example of a value is, "We value transparency in our use of big data."

Analysis means reviewing the organization's actual or proposed data-handling practices and determining how well they align with the identified values. For example, analysis might ask whether deducing that a woman is pregnant and placing ads for nursery furniture in front of her aligns well with a stated value of telling customers how the organization uses their personal information.

Articulation means writing the results of analysis—that is, stating explicitly where values and actions align and where they do not. Action means producing and executing plans to close alignment gaps and to prevent new ones from developing. For example, you might decide to place a button

labeled "Why am I seeing this ad?" next to the ad for nursery furniture.

Deciding which organizational activities require this ethical decision point process is tricky. One method, which Davis recognizes as not entirely satisfactory, is to look for the "creepy" factor. Like the Supreme Court's definition of pornography, this determination involves the phrase, "I know it when I see it." For example, it strikes many people as creepy when a retail firm determines whether a woman is pregnant based on her online behavior, so the organization should use the ethical decision point process to evaluate any actions it takes in this area. A health maintenance organization engaging in similar behavior might seem less creepy, so the ethical decision point process might be less important in that case.

Davis talks about ethical decision points in the context of big data, but the methodology can apply to all sorts of situations. We all have values, and we all do something like ethical decision point analysis in our everyday lives. Big data technology's speed and scale make it essential for organizations to carry out the process explicitly and transparently as a core competency. This reduces the risks of unintended consequences and provides a starting point for a clear and immediate response when things go wrong.

To determine the current practices of large organizations, Davis visited the websites of the top 50 of the Fortune 500 companies in the fall of 2011. Few of these organizations understand and articulate sets of values that customers can use to interpret the organizations' imprecise policy statements. Most focus on "privacy" with little mention of identity, ownership, and reputation. Most are concerned with personally identifying information, but fail to define that term. Most say that they don't sell personal data, but none claim that they don't buy such data. There is much more information in Davis's summary of his findings. Things might have changed for the better since this survey, but what he found was not encouraging.

Davis sees hope in the fact that the organizations he studied have many strong capabilities in place: leadership and management, strategic planning, a product development process, communication, education and training, and a process for initiating change. But he also notes that not everyone in an organization has the same values and that different roles within an organization have conflicting interests in transparency and alignment. And in the face of tactical pressures, it's tempting to kick the can down the road by avoiding ethical discussions entirely. Davis hopes to overcome these difficulties by taking a cookbook approach. The final third of this short book lays out his alignment methodology framework in more detail, complete with forms and a case study. The forms help you define value personas, an analog of customer personas, which most executives and marketing personnel understand.

As Davis points out, if you ask people what they want, they will tell you. By insisting on definitions of terms and explicit statements of values and how actions align with those values, he creates a framework in which people can discuss difficult issues without unnecessary confusion or rancor. Davis's method focuses on a critical area for every company that deals with big data and every person who engages in online transactions. Big data is coming at us like a runaway freight train, and we must methodically examine the issues and take appropriate actions, or we won't like the mess it makes when it crashes. I highly recommend this book to everyone.

MICRO

Richard Mateosian is a freelance technical writer in the San Francisco Bay Area. Contact him at xrmxrm@gmail.com.



Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.