

[Home](#)[Shop Video Training & Books](#)[Radar](#)[Safari Books Online](#)[Conferences](#)[Data](#)[Design](#)[Emerging Tech](#)[IoT](#)[Programming](#)[Web Ops & Performance](#)[Web Platform](#)

We're in the process of moving Radar to the new oreilly.com. Check it out.

What's Up With Big Data Ethics?

Insights from a business executive and law professor

by O'Reilly Strata | [Comment](#) | March 21, 2014

by [Jonathan H. King](#) & [Neil M. Richards](#)

If you develop software or manage databases, you're probably at the point now where the phrase "Big Data" makes you roll your eyes. Yes, it's hyped quite a lot these days. But, overexposed or not, the Big Data



Photo provided courtesy of Jonathan H. King.

Photo provided courtesy of Jonathan King

revolution raises a bunch of ethical issues related to privacy, confidentiality, transparency and identity. Who owns all that data that you're analyzing? Are there limits to what kinds of inferences you can make, or what decisions can be made about people based on those inferences? Perhaps you've wondered about this yourself.

[We're obsessed by these questions.](#) We're a business executive and a law professor who've written about this question a lot, but our audience is usually lawyers. But because engineers are the ones who confront these questions on a daily basis, we think it's essential to talk about these issues in the context of software development.



Photo provided courtesy of Neil M. Richards.

Photo provided courtesy of Neil M. Richards.

While there's nothing particularly new about the analytics conducted in big data, the scale and ease with which it can all be done today changes the ethical framework of data analysis. Developers today can tap into remarkably varied and far-flung data

sources. Just a few years ago, this kind of access would have been hard to imagine. The problem is that our ability to reveal patterns and

new knowledge from previously unexamined troves of data is moving faster than our current legal and ethical guidelines can manage. We can now do things that were impossible a few years ago, and we've driven off the existing ethical and legal maps. If we fail to preserve the values we care about in our new digital society, then our big data capabilities risk abandoning these values for the sake of innovation and expediency.

Consider the recent \$16 Billion acquisition of WhatsApp by Facebook. WhatsApp's meteoric growth to over 450 million mobile monthly users over the past four years was in part based on a "No Ads" philosophy. It was [reported](#) that SnapChat declined an earlier \$3 Billion acquisition offer from Facebook. Snapchat's primary value proposition is an ephemeral mobile message that disappears after a few seconds to protect message privacy. Why is Facebook willing to pay Billions for a mobile messaging company? *Demographics and Data*. Instead of spending time on Facebook, international and younger users are increasingly spending time on mobile messaging services that don't carry ads and offer heightened privacy by design. In missing this mobile usage, Facebook is lacking the mobile data. With WhatsApp, Facebook immediately gains access to the mobile data of hundreds of millions of users and growing. While WhatsApp founder Jan Koum promises "no ads, no games and no gimmicks" and has a board seat to back it up, Facebook has a pretty strong incentive to monetize the WhatsApp mobile data it will now control.

Big Data is about much more than just correlating database tables and creating pattern recognition algorithms. [It's about money and power](#). Big Data, broadly defined, is producing increased powers of institutional awareness and power that require the development of what we call Big Data Ethics. The Facebook acquisition of WhatsApp and the whole NSA affair shows just how high the stakes can be. Even when we're not dealing in national security, the values we build or fail to build into our new digital structures will define us.

From our perspective, we believe that any organizational conversation about big data ethics should relate to four basic principles that can lead to the establishment of big data norms:

- ***Privacy isn't dead***; it's just another word for information rules. Private doesn't always mean secret. Ensuring privacy of data is a matter of defining and enforcing information rules – not just rules about data collection, but about data use and retention. People should have the ability to manage the flow of their private information across massive, third-party analytical systems.
- ***Shared private information can still remain confidential***. It's not realistic to think of information as either secret or shared, completely public or completely private. For many reasons, some of them quite good, data (and metadata) is shared or generated by design with

services we trust (e.g. address books, pictures, GPS, cell tower, and WiFi location tracking of our cell phones). But just because we share and generate information, it doesn't follow that anything goes, whether we're talking medical data, financial data, address book data, location data, reading data, or anything else.

- ***Big data requires transparency.*** Big data is powerful when secondary uses of data sets produce new predictions and inferences. Of course, this leads to data being a business, with people such as data brokers, collecting massive amounts of data about us, often without our knowledge or consent, and shared in ways that we don't want or expect. For big data to work in ethical terms, the data owners (the people whose data we are handling) need to have a transparent view of how our data is being used – or sold.
- ***Big Data can compromise identity.*** Privacy protections aren't enough any more. Big data analytics can compromise identity by allowing institutional surveillance to moderate and even determine who we are before we make up our own minds. We need to begin to think about the kind of big data predictions and inferences that we will allow, and the ones that we should not.

There's a great deal of work to do in translating these principles into laws and rules that will result in ethical handling of Big Data. And there's certainly more principles we need to develop as we build more powerful tech tools. But anyone involved in handling big data should have a voice in the ethical discussion about the way Big Data is used. Developers and database administrators are on the front lines of the whole issue. The law is a powerful element of Big Data Ethics, but it is far from able to handle the many use cases and nuanced scenarios that arise. Organizational principles, institutional statements of ethics, self-policing, and other forms of ethical guidance are also needed. Technology itself can help provide an important element of the ethical mix as well. This could take the form of intelligent data use trackers that can tell us how our data is being used and let us make the decision about whether or not we want our data used in analysis that takes place beyond our spheres of awareness and control. We also need clear default rules for what kinds of processing of personal data is allowed, and what kinds of decisions based upon this data are acceptable when they affect people's lives. But the important point is this – we need a big data ethics, and software developers need to be at the center of these critical ethical discussions. Big data ethics, as we argue [in our paper](#), are for everyone.

Related Resources:

- [Data and Ethics: Etiquette and Law for an Always-On World](#) – Recorded Online Conference

- [Ethics of Big Data](#) – Book
 - [Big Ethics for Big Data](#) – Data Blog Post
 - [Big Data Ethics](#) – Paper
-