
accenture
Strategy | Consulting | Digital | Technology | Operations
Universal principles of data ethics
12 guidelines for developing ethics codes
High performance. Delivered.

Professional ethics codes serve a surprisingly broad range of purposes. On their face, ethics codes set out the standard for acceptable behavior within a profession. However, the acts of assembling, deliberating, distributing, training and enforcing ethics codes ultimately result in much broader impacts. A set of shared norms helps to define the boundaries of a professional community and identifies the standards that the public should demand of practitioners within that profession. It establishes the type of relationship that professionals have with the rest of society—including with their clients, research subjects, users of their services, and governments.

In many professions, learning about and complying with ethical obligations is a marker of accountability. Indeed, the act of becoming licensed in professions such as accounting, law and medicine includes swearing to uphold specific ethical norms. In some professions, such as journalism (and arguably science and engineering), the ethics of professional practice are the *only* commonality defining the boundaries of what is an extremely varied group. Though such codes typically lack the force of law, they do ultimately shape legal and regulatory dynamics by establishing expectations about responsibility and liability. In the long run, ethics codes can play a major role in defining a community of professionals.

Public discussions about the ethics of "big data" analytics are rapidly gaining prominence in public discourse. The insights derived from data already permeate much of our lives, and promise to shape even more of the opportunities, limits, and major and minor life decisions we encounter moving forward. In other words, data professionals will, in all likelihood, play a role in our lives that's as intimate as medical, fiduciary, and legal professionals. This is why establishing a shared set of norms is critically important for data scientists and practitioners (and those making requests of them). It's good for the profession and good for society.

This report discusses the dynamics involved in generating a code of ethics that could guide the profession of data science as it grows and evolves, and immediately help organizations shape their own internal guidelines related to data. A broad set of principles is proposed and intended to inform the development of domain-

specific codes of ethics for specific organizations or industries. Developing a code of ethics should be a collaborative effort that involves all of the stakeholders in a community and builds from the proposed principles. Additionally, the uses of data science are so diverse (and many are still unforeseen) that not every scenario can be accounted for in a code of ethics. Nor does "data science" adequately capture the many facets of the data ecosystem. There is a diversity of practitioners that utilize the techniques of data science to provide analysis, insights and advice about a breadth of human activities; all of these actors may have specific obligations that differ from data scientists. Nonetheless, these principles are intended to function as a foundation or outline of what a universal code of ethics for the data science field should emphasize.

## Framing data ethics

The foremost practical question for data ethics is whether there is anything special about data such that collecting, manipulating, and applying it requires a distinct code of ethics. The history of science and engineering ethics suggests that ethical regimes often track new ways of knowing. As new ways to know the world are developed, appropriate rules governing those approaches are helpful. And so the question of whether data scientists and practitioners need a special focus on ethics is ultimately a question of whether data science represents a distinctly new way of knowing.

The way data is used today is more than just a technical phenomenon. It's a political, social, and even mythological phenomenon that has consequences for how we organize our lives and express our values.[1] Whatever ethical principles are developed in connection with data, they should account for dynamics that extend beyond technical limitations. Data analytics should be viewed as a phenomenon with consequences beyond technology, and the community should demand that data scientists and practitioners consider those consequences.

Data analytics is an emerging form of knowledge production that provides the ability to cheaply and easily connect and analyze datasets, often drawn from highly disparate contexts.[2] The capacity to continually re-analyze and correlate data collected from a broad range of contexts has proven challenging to ethically conceptualize and regulate.[3,4] In the past it could be assumed that data collected in one context—medical, political, genetic, social, financial, census, behavioral, geographic, etc—would stay in that context and could be regulated as such. Furthermore, many familiar ethical controls, such as informed consent, occur only at the point of collection. But the power and peril of data science is that data is most valuable when it can be reused and repurposed in many different contexts and in combination with other datasets.

Personal and sensitive data now travels unpredictably and will be reused indefinitely for unforeseeable purposes. Because our "data selves" are no longer compartmentalized, many different actors can learn intimate details about the lives of anybody who leaves a digital trail. For this reason, the ethical infrastructures, concepts, and norms that have been developed to handle compartmentalized data are often neither salient nor applicable—how data moves in time and space is no longer synchronized with our temporally and geographically constrained ethical regimes.[5] The language of medical and scientific ethics has long emphasized *respect for persons* and *informed consent* as core values. But it is a daunting proposition to explain how such principles can hold when data about individuals is persistently shared, transformed, and aggregated and when future uses of datasets are so unknowable that "informed consent" is a misnomer at best—and impossible at worst.

## Professional codes of data ethics

Analyses of professional ethics codes show that the articulation of shared values is often a key stage in the professionalization of a field: It establishes who is a member of the field and what can be expected of them by colleagues, clients and society at large.[6,7] Mark Frankel offers a taxonomy of professional ethics codes as *aspirational, educational, and regulatory,* noting that most codes are an admixture and serve multiple goals. He argues that the process of establishing a code provides opportunities for critical reflexivity that are perhaps more important than the final product: "This process of self-criticism, codification, and consciousness-raising reinforces or redefines the profession's collective responsibility and is an important learning and maturing experience for both individual members and the profession."[8]

In an analysis conducted for the Council for Big Data, Ethics & Society, Jacob Metcalf identified the inward—and outward-facing goals of professional ethics codes that may be applicable for data ethics:[9]

### Inward-facing goals:

- Provide guidance when existing implicit norms and values are not sufficient; essentially, guidance for a novel situation
- Reduce internal conflicts; strengthen the sense of common purpose among members of the organization
- Satisfy internal criticism from members of a profession
- Create generalized rules for individuals and organizations that have responsibilities for important "human goods"
- Establish role-specific guidelines that demarcate general principles as particular duties
- Establish standards of behavior toward colleagues, students/trainees, employees, employers, clients
- Strengthen the sense of common purpose among members of an organization
- Deter unethical behavior by identifying sanctions and creating an environment in which the reporting of unethical behavior is affirmed
- Provide support for individuals who come under pressure to behave in an unethical manner

### Outward-facing goals:

- Protect vulnerable populations and individuals who could be harmed by the profession's activities
- Protect and enhance the good reputation of and trust for the profession
- Establish the profession as a distinct moral community worthy of autonomy from external control and regulation
- Provide a basis for public expectations and evaluation of the profession
- Serve as a basis for adjudicating disputes among members of the profession and disputes between members and the public
- Create institutions that are resilient in the face of external pressures
- Respond to past harms done by the profession.

There are already some ethics codes that cover most computing and data scientists and engineers. In the US, four major computing professional societies have substantially different codes for their members due to their different missions.[10] The Association of Computing Machinery (ACM), the largest professional organization for computer scientists and engineers, distributes an ethics code for members of its organization.[11] However, that code was adopted in 1992 at the beginning of the internet age, predating many of the technologies that define the ethical conflicts faced by data and computing professionals today. Although the ACM's ethics code contains some principles that do still hold up—such as striving to maintain the integrity of data about individuals—it lacks the specificity that would make the code optimally useful to current and future generations of data and computing professionals. Other professional groups that are more closely associated with the data revolution have more recent codes. The recently founded Data Science Association offers a relatively detailed ethics code that is notable for detailing how members should adhere closely to scientifically sound statistical methods.[12] For example, rule 8(d) reads:

> "If a data scientist reasonably believes a client is misusing data science to communicate a false reality or promote an illusion of understanding, the data scientist shall take reasonable remedial measures, including disclosure to the client, and including, if necessary, disclosure to the proper authorities. The data scientist shall take reasonable measures to persuade the client to use data science appropriately."

Source: Code of Conduct, Data Science Association

Some data science sub-disciplines have also produced valuable ethics codes and other types of ethics guidance for their members. The Association of Internet Researchers (AoIR) developed an ethics code in 2002, updated in 2012, that addresses the obligations of social science researchers working in digital domains at a macro-level.[13] This document is notable for the extensive list of questions internet researchers should address. The National Center for Education Statistics produced a guide for appropriate use of educational data in 2010, that mixes core principles with illustrative case studies.[14]

## Challenges for a universal code of data ethics

A unique aspect of today's datasets is their sprawling, multi-disciplinary utility—data science is arguably closer to a service than a discipline because it is useful in so many industries and disciplines. The analytical tools developed in applied mathematics, statistics, and computer science are being taken up by disciplines and sectors such as medicine, marketing, finance, the humanities, social science, criminal justice, geography and geospatial imaging, manufacturing, social work, human rights, and many more. This poses a major challenge for a *universal* code of data ethics: There may be too few commonalities across the specific uses of data science to pull together a single code. Principles of data ethics that hold in medicine may not hold in finance because the social roles occupied by medical professionals and financiers differ significantly. They have meaningfully different obligations to their clients and society, and so it is reasonable to expect that their uses of big data for good and ill will similarly vary.

Furthermore, many of these fields already have their own professional ethics codes that may or may not address the changes introduced by the data age. Other fields have dealt with such problems by having professional sub-societies formulate secondary ethics codes. For example, the American College of Obstetricians and Gynecologists holds its members both to the American Medical Association's code of ethics (that applies to all physicians) and to a more specific set of obligations that apply only to their own members. If data science continues on its path to ubiquity, then it may be challenging to define a truly universal code that covers its uses in such a variety of contexts.

One of the quirks of data science is that its parent fields have traditionally fallen outside of the purview of US federal research ethics regulations. Following a long arc of infamous research scandals in the mid-20th century—ranging from Nuremberg to Tuskegee to the Stanford prison experiment—the 1974 National Research Act empowered federal regulators to identify, define, and enforce ethical standards for human-subjects research that uses federal funds. The authors of the 1979 Belmont Report commissioned by the Act identified the three primary principles of bioethics: **beneficence** (research should be carefully constructed to do good in the world), **respect for persons** (research must respect personal values such as autonomy, privacy and dignity) and **justice** (research must further social equity).

These principles subsequently informed the rulemaking process initiated by the Department of Health and Human Services that resulted in the federal regulations known as the Common Rule. The Common Rule now governs (nearly) all human-subjects research funded by federal agencies. Its most consequential outcome was establishing Institutional Review Boards (IRBs) as an obligatory milestone for most academic research. However, computer science and engineering, applied mathematics, and much quantitative sociology research has historically fallen outside of the regulatory definition of "human subjects," even when these fields involve human lives.[15] As a result, most professionals trained in the parent fields of data science do not encounter the primary research norms and regulatory apparatuses that guide other science and engineering fields.

The Common Rule and IRBs dominate conversations about practical ethics, but in some cases, even these regulated standards do not go far enough. In the humanitarian field, some academics and practitioners are beginning to call for higher standards.[16] They argue that "demographically identifiable data"—a broader classification than Personally Identifiable Information (PII), the gold standard for privacy professionals—could cause various harms to entire classes of people.

As data science matures as a field and increasingly affects the human condition, there's a chorus building among professionals and practitioners to have more guidance for the ethical decisions they are forced to make—and might be unaware they are making—on a daily basis. The set of Principles proposed below is intended to provide a baseline for those seeking such guidance and those looking to develop a group-specific code of data ethics.

## Principles for Data Ethics

Data science professionals and practitioners should strive to perpetuate these principles:

### 1. The highest priority is to respect the persons behind the data.

When insights derived from data could impact the human condition, the potential harm to individuals and communities should be the paramount consideration. Big data can produce compelling insights about populations, but those same insights can be used to unfairly limit an individual's possibilities.

### 2. Attend to the downstream uses of datasets.

Data professionals should strive to use data in ways that are consistent with the intentions and understanding of the disclosing party. Many regulations govern datasets on the basis of the status of the data, such as "public," "private" or "proprietary." However, what is *done with* datasets is ultimately more consequential to subjects/users than the type of data or the context in which it is collected. Correlative uses of repurposed data in research and industry represents both the greatest promise and the greatest risk posed by data analytics.

### 3. Provenance of the data and analytical tools shapes the consequences of their use.

There is no such thing as raw data—all datasets and accompanying analytic tools carry a history of human decision-making. As much as possible, that history should be auditable, including mechanisms for tracking the context of collection, methods of consent, the chain of responsibility, and assessments of quality and accuracy of the data.

### 4. Strive to match privacy and security safeguards with privacy and security expectations.

Data subjects hold a range of expectations about the privacy and security of their data and those expectations are often context-dependent. Designers and data professionals should give due consideration to those expectations and align safeguards and expectations as much as possible.

### 5. Always follow the law, but understand that the law is often a minimum bar.

As digital transformations have become a standard evolutionary path for businesses, governments and laws have largely failed to keep up with the pace of digital innovation and existing regulations are often mis-calibrated to present risks. In this context, compliance means complacency. To excel in data ethics, leaders must define their own compliance frameworks that outperform legislated requirements.

### 6. Be wary of collecting data just for the sake of more data.

The power and peril of data analytics is that data collected today will be useful for unpredictable purposes in the future. Give due consideration to the possibility that less data may result in both better analysis and less risk.

### 7. Data can be a tool of inclusion and exclusion.

While everyone deserves the social and economic benefits of data, not everyone is equally impacted by the processes of data collection, correlation, and prediction. Data professionals should strive to mitigate the disparate impacts of their products and listen to the concerns of affected communities.

### 8. As much as possible, explain methods for analysis and marketing to data disclosers.

Maximizing transparency at the point of data collection can minimize more significant risks as data travels through the data supply chain.

### 9. Data scientists and practitioners should accurately represent their qualifications, limits to their expertise, adhere to professional standards, and strive for peer accountability.

The long-term success of the field depends on public and client trust. Data professionals should develop practices for holding themselves and peers accountable to shared standards.

### 10. Aspire to design practices that incorporate transparency, configurability, accountability, and auditability.

Not all ethical dilemmas have design solutions, but being aware of design practices can break down many of the practical barriers that stand in the way of shared, robust ethical standards. Data ethics is an engineering challenge worthy of the best minds in the field.

### 11. Products and research practices should be subject to internal, and potentially external ethical review.

Organizations should prioritize establishing consistent, efficient, and actionable ethics review practices for new products, services, and research programs. Internal peer-review practices can mitigate risk, and an external review board can contribute significantly to public trust.

### 12. Governance practices should be robust, known to all team members and reviewed regularly.

Data ethics poses organizational challenges that cannot be resolved by familiar compliance regimes alone. Because the regulatory, social, and engineering terrains are so unsettled, organizations engaged in data analytics require collaborative, routine and transparent practices for ethical governance.

## 100/365-day Plans

Over the course of the next year, every organization can be well on its way to leveraging these 12 universal principles to develop a custom-tailored code of data ethics.

### In three months, your organization should:

- Identify the internal and external stakeholders who should contribute to or review your emerging code of ethics.
- Catalog the ecosystems, industry groups, and professional organizations where your organization is a member or participates in a meaningful way.
- Highlight the relationships in the list above where data is shared or sold.
- Collect codes of ethics from these ecosystems, industry groups, and professional organizations. Aggregate the parts of these codes that deal with the handling of data and use them as a minimum bar for your own code.

In one year (and beyond), your organization should strive to:

- Note the existing practices that require modification and consult with the process owners to understand any impediments to adopting more rigorous ethical practices.
- Once ratified, publish your code of ethics for public consumption and consider submitting it to the Center for the Study of Ethics in the Professions.
- Encourage partners to publicly publish and commit to abide by this new code of ethics.
- Circulate an early draft of your code among stakeholders and have them indicate existing practices that would require modification if the code were to be ratified.
- After incorporating insights from prior discussions, publish a code of ethics among internal stakeholders and partners who will be participating in a 12-month pilot of the draft code; once the pilot starts, interview stakeholders and partners every three months to understand how their work was impacted. With the insights from the completed pilot, make a decision to ratify or update the draft code of data ethics.
- Share outcomes of the pilot with all stakeholders and notify them when and how they will be held accountable for being able to demonstrate compliance with the code.

## References


1 Crawford K, Gray ML and Miltner K (2014) Critiquing Big Data: Politics, Ethics, Epistemology. International Journal of Communication 8(0): 10.

2 Mayer-Schönberger V and Cukier K (2013) Big Data: A Revolution that Will Transform how We Live, Work, and Think. Houghton Mifflin Harcourt.

3 Zwitter A (2014) Big Data ethics. Big Data & Society 1(2). DOI: 10.1177/205395171455925.

4 Metcalf J, boyd danah and Keller EF (2016) Perspectives on Big Data, Ethics, and Society. Council for Big Data, Ethics, and Society. (accessed 31 May 2016).

5 Metcalf J and Crawford K (2016) Where are human subjects in big data research? The emerging ethics divide. Big Data & Society 3(1): 1–14. DOI: 10.1177/2053951716650211.

6 Metcalf J (2014) Ethics Codes: History, Context, and Challenges. Council for Big Data, Ethics, and Society. (accessed 21 October 2015).

7 The Illinois Institute of Technology maintains a thorough collection of professional ethics codes as part of their Center for the Study of Ethics in the Professions,

8 Frankel MS (1989) Professional codes: why, how, and with what impact? Journal of business ethics 8(2-3): 109–115.

9 Metcalf, J (2014). See also: Frankel MS (1989); Gaumnitz BR and Lere JC (2002) Contents of Codes of Ethics of Professional Business Organizations in the United States. Journal of Business Ethics 35(1): 35–49; Kaptein M and Wempe J (1998) Twelve Gordian Knots When Developing an Organizational Code of Ethics. Journal of Business Ethics 17(8): 853–869.

10 Oz E (1993) Ethical standards for computer professionals: a comparative analysis of four major codes. Journal of Business Ethics 12(9): 709–726.

11 "ACM Code of Ethics and Professional Conduct." ACM Code of Ethics and Professional Conduct. ACM, Inc. 16 October 1992. Web. 31 May 2016.

12 "Code of Conduct | Data Science Association." Data science code of professional conduct. Data Science Association. Accessed 31 May 2016.

13 http://aoir.org/reports/ethics2.pdf. Accessed 31 May 2016.

14 https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010801. Accessed 31 May 2016.

15 Metcalf J and Crawford K (2016) Where are human subjects in big data research? The emerging ethics divide. Big Data & Society 3(1): 1–14.

16 Karunakara U (2014) Data Sharing in a Humanitarian Context: The Experience of Médicins Sans Frontières. In: Moore SA (ed.), Issues in Open Research Data, London: Ubiquity Press.




## Contact Us


**Steven C. Tiell**
Senior Principal—Digital Ethics
Accenture Labs
steven.c.tiell@accenture.com

**Jacob Metcalf**
Ethical Resolve
jake@ethicalresolve.com


## About Accenture Labs

Accenture Labs invents the future for Accenture, our clients and the market. Focused on solving critical business problems with advanced technology, Accenture Labs brings fresh insights and innovations to our clients, helping them capitalize on dramatic changes in technology, business and society. Our dedicated team of technologists and researchers work with leaders across the company to invest in, incubate and deliver breakthrough ideas and solutions that help our clients create new sources of business advantage.

Accenture Labs is located in six key research hubs around the world: Silicon Valley, CA; Sophia Antipolis, France; Arlington, Virginia; Beijing, China; Bangalore, India, and Dublin, Ireland. The Labs collaborates extensively with Accenture's network of nearly 400 innovation centers, studios and centers of excellence located in 92 cities and 35 countries globally to deliver cutting-edge research, insights and solutions to clients where they operate and live. For more information, please visit www.accenture.com/labs.

## Data Ethics Research Initiative

Launched by Accenture's Technology Vision team, the Data Ethics Research Initiative brings together leading thinkers and researchers from Accenture Labs and over a dozen external organizations to explore the most pertinent issues of data ethics in the digital economy. The goal of this research initiative is to outline strategic guidelines and tactical actions businesses, government agencies, and NGOs can take to adopt ethical practices throughout their data supply chains.

## About Accenture

Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. Combining unmatched experience and specialized skills across more than 40 industries and all business functions—underpinned by the world's largest delivery network—Accenture works at the intersection of business and technology to help clients improve their performance and create sustainable value for their stakeholders. With approximately 373,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives. Visit us at www.accenture.com.

**Learn more:** www.accenture.com/DataEthics