Building digital trust:
The role of data ethics in the digital age

High performance. Delivered.

The digital economy is built on data—massive streams of data being created, collected, combined, and shared— for which traditional governance frameworks and risk-mitigation strategies are insufficient. In the digital age, analyzing and acting on insights from data can introduce entirely new classes of risk. These include unethical or even illegal use of insights, amplifying biases that exacerbate issues of social and economic justice, and using data for purposes to which its original disclosers would not have agreed, and without their consent.

Provided organizations prioritize ethical data practices throughout their decision-making processes, risks such as these can be identified, managed, and contained. The alternative? Left unchecked, they can permanently damage consumer trust in a brand. The following example puts the scale of these risks into perspective. The developers of a dating app were tasked with increasing the amount of time users spend with the app. In their data analysis, they discovered a strong correlation between engagement and ethnic and racial biases. Under pressure to improve business metrics, a new match recommendation algorithm predicting and reinforcing these biases went into production. This true story illustrates how the digital economy can scale poor judgement to a massive degree and lead to serious ethical failures.

There are many similar risks involving the ethical use of data, where today's best practices are simply insufficient to provide a guide for practitioners. These new vectors for risk demand the development of robust ethical controls throughout data supply chains. With such controls, organizations can create "digital trust"—a widely accepted belief that a brand is reliable, capable, safe, transparent, and truthful in its digital practices. Digital trust is difficult to build, but startlingly easy to lose. This makes it a key differentiator in the digital economy. Confidence in a brand facilitates growth through product development, collaboration with partners, and expansion into new and existing markets.

In the past, the scope for digital risk was largely limited to cybersecurity threats. These threats remain omnipresent, but leading organizations must now also recognize risks from lackluster ethical data practices. Mitigating these internal threats is critical for every player in the digital economy, and cannot be addressed with strong cybersecurity alone. These new risks require their own frameworks and best practices at every step of project and service delivery lifecycles, and should be integrated into every project, offering, or new endeavor.
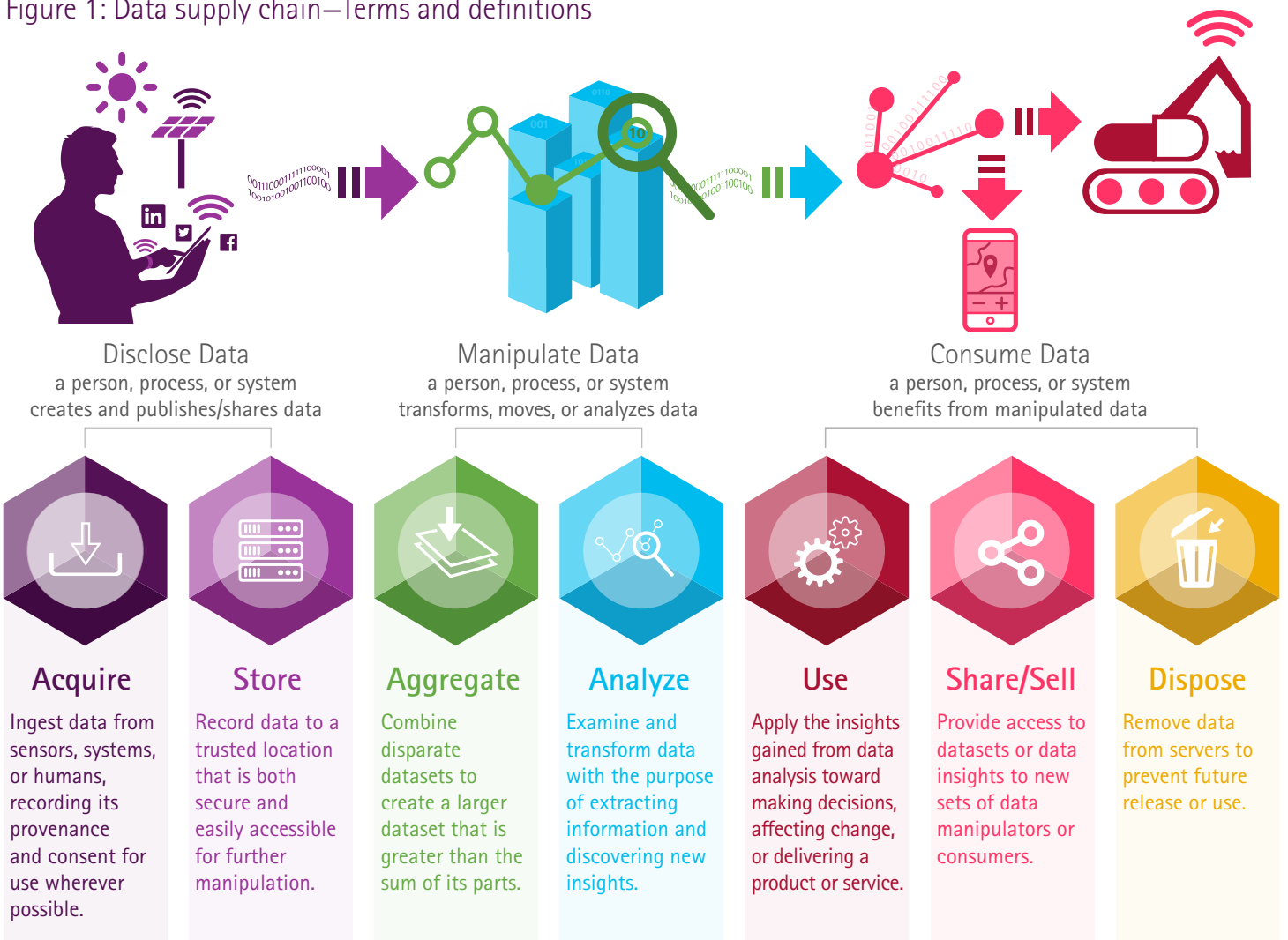
Focus on ethics

Treating data in an ethical manner throughout its supply chain requires a fundamental change in how data is viewed within organizations. While the perspectives of security (is the confidentiality, integrity, and availability of data adequately protected?) and privacy (do controls on data satisfy regulatory requirements?) remain relevant, added lenses for ethics and trust become critical. Organizations must begin to consider the ethics of data collection, manipulation, and use. This enables trust, but requires attention at each stage of the data supply chain and collaboration with every stakeholder.

Introducing these new perspectives will ensure an organization can simultaneously manage risk and build trust by consistently evaluating how ethics are taken into account in data-driven decisions. By focusing on ethics, organizations will improve the trust their customers have in them—a mandate for those that have undergone digital transformations and become publishers of, or participants in, digital platforms and ecosystems.

For example, Everledger set out to minimize fraud and the prevalence of conflict gems in the diamond industry. To attract investors and realize its goals, the company knew its solution would have to be completely transparent, auditable, and immutable. To achieve this, it uses a blockchain architecture that delivers on all of these requirements. Everledger also aggregates data from law enforcement and insurance companies, which in turn use the technology as a verification system, reducing fraud and its associated costs. Everledger has built a trusted, permanent ledger for diamond certification and transaction histories that can be extended to track any asset with a unique identifier. This solution is "trusted by design."

Figure 1: Data supply chain—Terms and definitions

Disclose Data
a person, process, or system creates and publishes/shares data

Manipulate Data
a person, process, or system transforms, moves, or analyzes data

Consume Data
a person, process, or system benefits from manipulated data

Acquire
Ingest data from sensors, systems, or humans, recording its provenance and consent for use wherever possible.

Store
Record data to a trusted location that is both secure and easily accessible for further manipulation.

Aggregate
Combine disparate datasets to create a larger dataset that is greater than the sum of its parts.

Analyze
Examine and transform data with the purpose of extracting information and discovering new insights.

Use
Apply the insights gained from data analysis toward making decisions, affecting change, or delivering a product or service.

Share/Sell
Provide access to datasets or data insights to new sets of data manipulators or consumers.

Dispose
Remove data from servers to prevent future release or use.

To build solutions that are trusted by design, a new set of best practices must be created to guide practitioners through the process of embedding ethical considerations at every stage of product development, service delivery, and the data supply chain. The "pivot to trust" strategy rewards those who demonstrate a commitment to strong ethical standards and sets them apart from those who do not. This focus on data ethics requires a portfolio of considerations—outlined below—that are new to many practitioners.

Taxonomies for impact across the data supply chain

In a digital marketplace where consumers discriminate based on their ability to trust, achieving a high level of trust adds gravitational pull to a brand and is becoming a strong differentiator for companies. This is true across industries and sectors of the economy. But to move data ethics forward as a discipline, there must be a common language that professionals can use for discussing and classifying data ethics— from acquisition to sharing and throughout the data supply chain.

Given the relative youth of data ethics, such a common language does not yet exist. Accenture is proposing a taxonomy that will help enable practitioners to describe the nuances of making ethical decisions about data.[1] Having a taxonomy provides clarity to all parties involved in the exchange of data and will prove increasingly valuable as regulatory and insurance industry standards evolve. Already, companies must have policies and procedures in place to address the types of behavioral risk vectors exploited by cyber attackers. However, as insurance markets (and regulators) begin to recognize internal versus external threats, policies focused on ethical considerations throughout the data supply chain will become commonplace as a strategy for managing risk. Ultimately, Accenture sees a future where insurers will offer both cybersecurity and digital ethics policies. Being ready for these developments with a language and approach to account for these largely internal risks will be a significant advantage.

81%
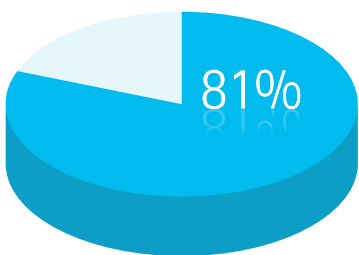
"81 percent of executives agree that as the business value of data grows, the risks companies face from improper handling of data are growing exponentially."[2]

Developing a code of ethics

With a language in place to facilitate evolution of data ethics frameworks, organizations should begin to consider the implications of working with data from an ethical perspective. Pre-existing codes of conduct are generally written for other domains or are grossly out of date, barely taking data into account (if at all). Addressing this gap, Accenture has developed a set of data-centric principles that organizations can use to develop a code of ethics.[3] Rather than attempt to deliver dozens of industry-specific codes, this framework approach lets organizations incorporate their industry knowledge and domain expertise in developing a code of ethics for their industry, ecosystem, or organization. The result will be a domain-specific code, directly applicable to each organization. This is critical because, when it comes to ethics, a community must have a generally agreed-upon set of norms that reflects its values. In civil society, these so-called "social norms" vary widely. Defining a code of ethics for a community of data practitioners is a necessary precursor to defining policies and procedures that ensure digital trust is established consistently, and in tandem with, all new products and services. When done correctly, a code of ethics helps to improve transparency for stakeholders and accountability for governance bodies.

Figure 2: Universal principles for data ethics—Guidelines for creating a code of data ethics

1. The highest priority is to respect the persons behind the data.
Where 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 into populations, but those same insights can be used to unfairly limit an individual's possibilities.

2. Account for 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: "public," "private" or "proprietary," for example. But 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 use of repurposed data in research and industry represents the greatest promise and the greatest risk of data analytics.

3. The consequences of utilizing data and analytical tools today are shaped by how they've been used in the past.
There's no such thing as raw data. All datasets and accompanying analytic tools carry a history of human decision-making. As far as possible, that history should be auditable. This should include mechanisms for tracking the context of collection, methods of consent, chains of responsibility, and assessments of data quality and accuracy.

4. Seek 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. These expectations are often context-dependent. Designers and data professionals should give due consideration to those expectations and align safeguards and expectations with them, as much as possible.

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

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

6. Be wary of collecting data just for the sake of having 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 both inclusion and exclusion.

While everyone should have access to 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 far as possible, explain methods for analysis and marketing to data disclosers.

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

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

The long-term success of this discipline depends on public and client trust. Data professionals should develop practices for holding themselves and their 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 paying close attention to 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 help to 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 regularly reviewed.

Data ethics poses organizational challenges that cannot be resolved by compliance regimes alone. Because the regulatory, social, and engineering terrains are in flux, organizations engaged in data analytics need collaborative, routine and transparent practices for ethical governance.
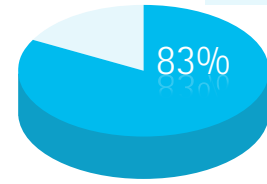
80%
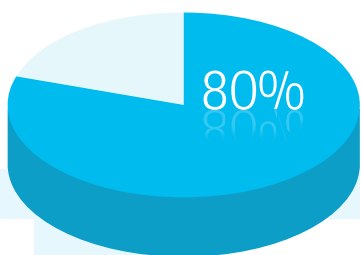
"80 percent of executives report strong demand among knowledge workers for increased ethical controls for data."[2]
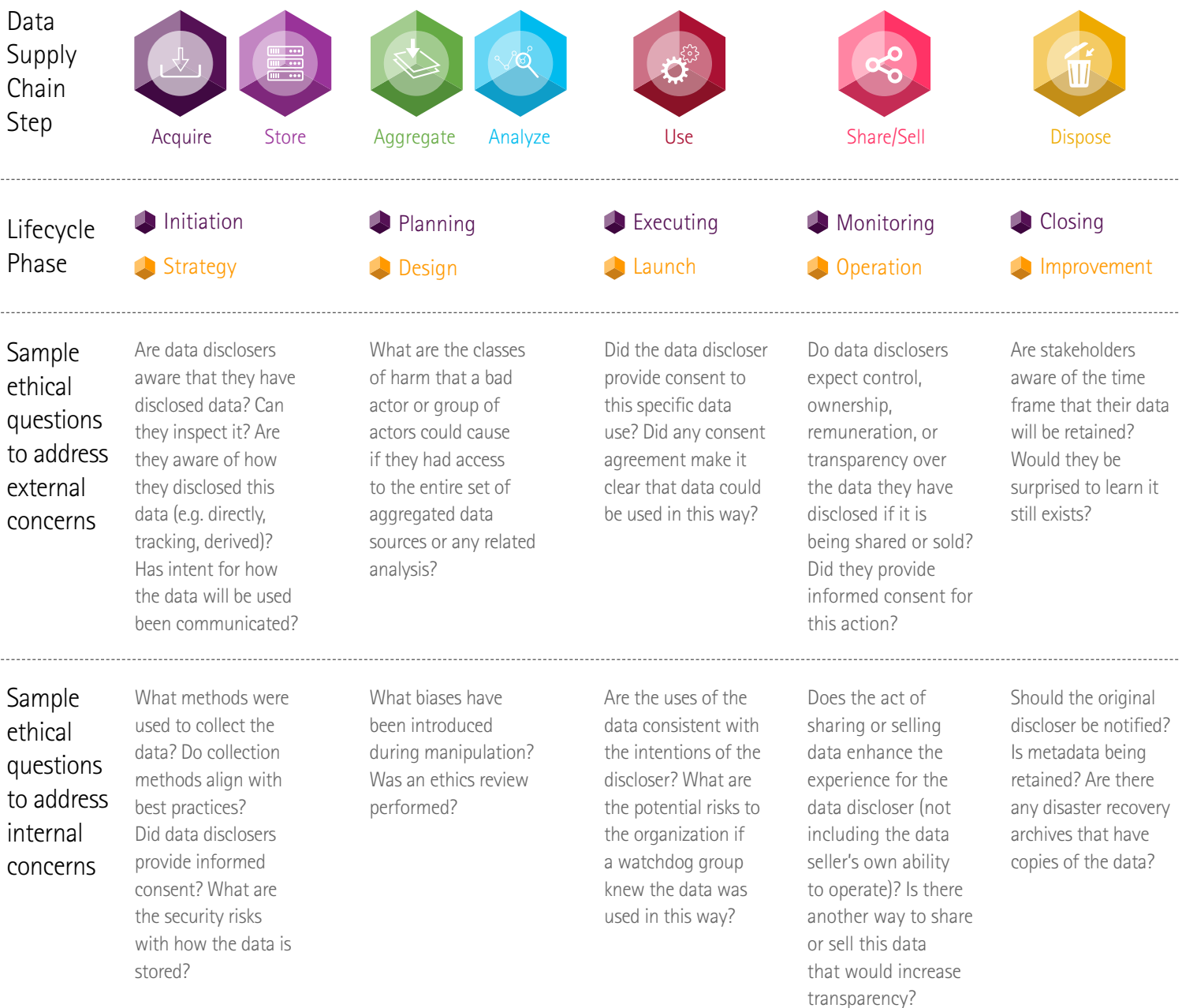
Guiding ethical decisions

A code of ethics also helps in defining the types of questions and concerns managers should be raising at each stage of project management and service delivery lifecycles.[4] This includes advice on how to design and implement an ethics review (much as processes exist for code reviews in software development). With this approach, organizations can be certain that trust is baked into and reinforced with all new offerings, engendering loyalty and confidence among consumers and partners.

83%

"83 percent of executives agree that trust is the cornerstone of the digital economy."[2]

Figure 3: Ethical decision-making across the data supply chain

Data Supply Chain Step	Acquire	Store	Aggregate	Analyze	Use	Share/Sell	Dispose
Lifecycle Phase	● Initiation ● Strategy		● Planning ● Design		● Executing ● Launch	● Monitoring ● Operation	● Closing ● Improvement
Sample ethical questions to address external concerns	Are data disclosers aware that they have disclosed data? Can they inspect it? Are they aware of how they disclosed this data (e.g. directly, tracking, derived)? Has intent for how the data will be used been communicated?		What are the classes of harm that a bad actor or group of actors could cause if they had access to the entire set of aggregated data sources or any related analysis?		Did the data discloser provide consent to this specific data use? Did any consent agreement make it clear that data could be used in this way?	Do data disclosers expect control, ownership, remuneration, or transparency over the data they have disclosed if it is being shared or sold? Did they provide informed consent for this action?	Are stakeholders aware of the time frame that their data will be retained? Would they be surprised to learn it still exists?
Sample ethical questions to address internal concerns	What methods were used to collect the data? Do collection methods align with best practices? Did data disclosers provide informed consent? What are the security risks with how the data is stored?		What biases have been introduced during manipulation? Was an ethics review performed?		Are the uses of the data consistent with the intentions of the discloser? What are the potential risks to the organization if a watchdog group knew the data was used in this way?	Does the act of sharing or selling data enhance the experience for the data discloser (not including the data seller's own ability to operate)? Is there another way to share or sell this data that would increase transparency?	Should the original discloser be notified? Is metadata being retained? Are there any disaster recovery archives that have copies of the data?

● Project Lifecycle Stage   ● Service Lifecycle Phase

Informed consent

Trust can be improved at the beginning of a data supply chain by making *informed consent* a priority. When consent is granted by an informed data-discloser, organizations have the added benefit of reducing their exposure to potential harm. With data being collected at an unprecedented scale, stored longer than ever, and combined with other datasets, it is critical to consider potential harm arising from its use. As data moves through its supply chain, the scope for its use often creeps further away from the consent of the initial disclosing party. This leads to uses of data that could not have been predicted at the time data was disclosed—calling into question whether or not truly informed consent is possible.

Consider the growing platform economy, where organizations from different industries are partnering to create new offerings. Imagine a fitness company partnering with an insurance business, and bringing their customers' data with them. These customers may well have originally given their consent for this data to be used to tailor fitness-related offerings.

But they might have felt differently if they'd known these offerings would eventually include insurance products.

Given these circumstances, what does "consent" mean in the context of data collection? How can organizations obtain meaningful consent from their customers and, as the platform economy continues to grow, their partners' customers? Figure 4 shows a framework for analyzing informed consent as a way to meet the "do no harm" ethos for data scientists— proactively addressing new risks that are only now starting to appear.[5]

Figure 4: Guidelines for avoiding harm

|	DATA AT REST	DATA IN MOTION
Data Disclosure	Data may be sourced from archives or other backups **Guideline:** Ensure the context of original consent is known and respected; data security practices should be revisited on a regular basis to minimize risk of accidental disclosure. Aggregation of data from multiple sources often represents a new context for disclosure; have the responsible parties made a meaningful effort to renew informed consent agreements for this new context?	Data is collected in real-time from machine sensors, automated processes, or human input; while in motion, data may or may not be retained, reshaped, corrupted, disclosed, etc. **Guideline:** Be respectful of data disclosers and the individuals behind the data. Protect the integrity and security of data throughout networks and supply chains. Only collect the minimum amount of data needed for a specific application. Avoid collecting personally identifiable information, or any associated meta-data whenever possible. Maximize preservation of provenance.
Data Manipulation	Data is stored locally without widespread distribution channels; all transformations happen locally **Guideline:** Set up a secure environment for handling static data so the risk of security breaches is minimized and data is not mistakenly shared with external networks. Data movement and transformation should be fully auditable.	Data is actively being moved or aggregated; data transformations use multiple datasets or API calls which might be from multiple parties; the Internet may be used **Guideline:** Ensure that data moving between networks and cloud service providers is encrypted; shared datasets should strive to minimize the amount of data shared and anonymize as much as possible. Be sure to destroy any temporary databases that contain aggregated data. Are research outcomes consistent with the discloser's original intentions?
Data Consumption	Data analytics processes do not rely on live or real-time updates **Guideline:** Consider how comfortable data disclosers would be with how the derived insights are being applied. Gain consent, preferably informed consent, from data disclosers for application-specific uses of data.	Data insights could be context-aware, informed by sensors, or might benefit from streamed data or API calls **Guideline:** The data at rest guidelines for data consumption are equally important here. In addition, adhere to any license agreements associated with the APIs being used. Encrypt data. Be conscious of the lack of control over streamed data once it is broadcast. Streaming data also has a unique range of potential harms—the ability to track individuals, deciphering network vulnerabilities, etc.

Data-sharing best practices

Strong ethical and risk mitigation practices preclude sharing data without the consent of the people who disclose it. They also preclude sharing data with parties to whom access has not been granted. But the effective use of data demands that it should be shared—and particularly so in the digital business era, with a growing platform economy. More and more organizations are partnering to create new offerings and even new industries.

These collaborations necessitate widespread and constant data sharing, bringing new and difficult-to-predict risks. These risks are compounded by the fact that once data sets reach a large enough size, anonymity is a myth.[6] And when additional data sets are aggregated, individuals can be identified with relative ease.[7,8] Addressing the issues and damage associated with sharing data, Figure 5 shows a set of guiding principles that can be put in place to mitigate risk.[9]

94%
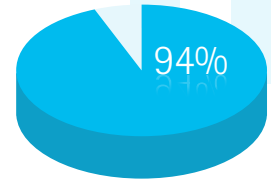
"94 percent of organizations are required to comply with ethical data handling requirements that go beyond their own protocols."[2]

Figure 5: Best practices for data sharing

1. Ongoing collaboration and mutual accountability are necessary between data sharing partners.

2. Build common contracting procedures, but treat every contract and dataset as unique.

3. Develop ethical review procedures between partners.

4. Be mutually accountable for interpretive resources.

5. Maximalist approaches to sharing are not always advisable.

6. Identify potential risks of sharing data within sharing agreements.

7. Repurposed data requires special attention.

8. When ethical principles or regulations are unclear, emphasize process and transparency.

9. Published research requires additional attention.

10. Treat trust as a networked phenomenon.

Ethical algorithms and automation

New risks and challenges in the digital economy extend to various types of automation that are powered by data insights. Online shoppers may be well aware that retailers will use purchase histories to drive discount offers, but the fact that ecommerce brands offer the same items at different prices based on location and other factors is only slowly becoming common knowledge. With careful deployment, this approach can deliver better marketing. But companies must carefully consider what they are using as input, how their algorithms are designed to consider that input, and how customers may react to its use.

This targeting is accomplished entirely via automated sense-and-respond systems using previously collected data to make decisions. These and other systems that operate in the physical world are subject to many of the same ethical issues as the machine-learning systems that drive ecommerce. But they also raise their own unique challenges. Take the case of the smartphone app that monitors for potholes in the road by passively collecting accelerometer data. The first cities that deployed this technology to prioritize road maintenance saw wealthy communities receive the most attention—because those were the people with the most smartphones.

The well-intended system amplified existing economic inequality issues and damaged public trust.

Building public trust in these systems, and in the decisions that result from their use, is critical to furthering their adoption (and, in the example above, the public safety improvements that could have been realized). But building that trust requires transparency and auditability, along with recourse and responsiveness when failures happen. These measures cannot be afterthoughts. They require careful foresight and planning. Accenture has a strategy for the ethical design, deployment and operation of sense-and-respond systems—each of which requires specific and tailored attention within the larger field of digital ethics.[10]

Start building digital trust today

In the digital era, data is the fundamental currency. And how organizations handle it throughout the data supply chain—from collection, aggregation, sharing, and analysis, to monetization, storage, and disposal—can have a decisive impact on their reputation and effectiveness.

New vectors of risk are scattered throughout the data supply chain. How businesses, governments, and NGOs address this risk, within and beyond the four walls of the enterprise, is critical to their ability to operate. As ethical data concerns continue to proliferate, organizations need to find a new way forward, and should embrace the opportunity: this new ethical frontier offers a way to engender trust and provide vital differentiation in a crowded marketplace.

Organizations should begin taking steps now to reduce their exposure to digital risk by integrating a wide array of data ethics practices throughout their data supply chains. In doing so, they'll gain the trust of stakeholders, reap business benefits, and position themselves for prolonged success in the digital economy.

"This new ethical frontier offers a way to engender trust and provide vital differentiation in a crowded marketplace."
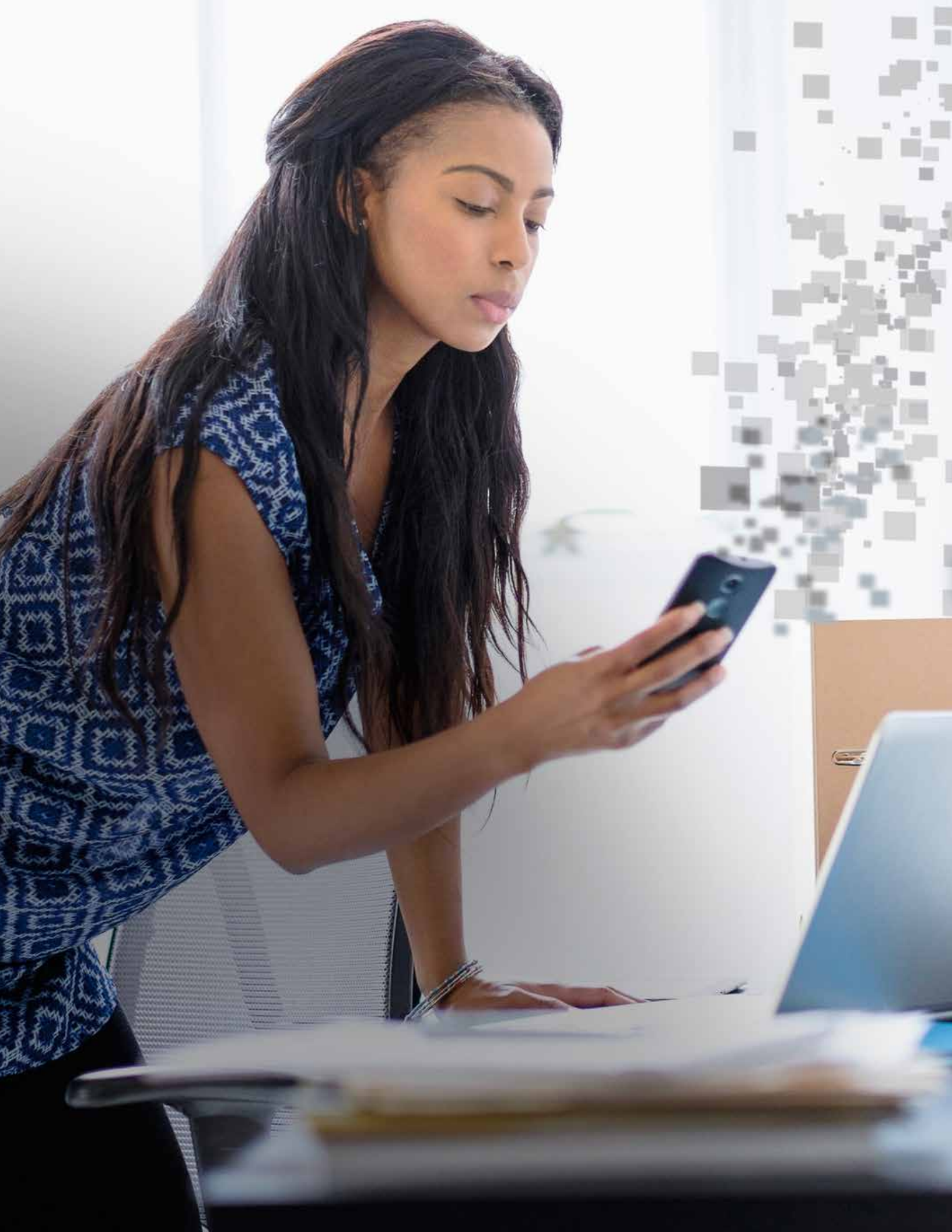
Contact Us

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

Lisa O'Connor
Managing Director—Security R&D
Accenture Labs
lisa.oconnor@accenture.com

Contributors

Harrison Lynch, Accenture

Richard Bartley, Accenture

Jacob Metcalf, Ethical Resolve

MJ Petroni, Causeit, Inc.

Aman Ahuja, The Data Guild

Scott L. David, University of Washington

References

1 "Taxonomies for impact along the data supply chain"

2 Accenture Technology Vision 2016 Survey

3 "Universal principles of data ethics"

4 "Ethical decisions throughout the data supply chain"

5 "Informed consent/implications of doing no harm"

6 de Montjoye, Y.-A., Radaelli, L, Singh, V. K., & Pentland, A. "Sandy." (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. Science, 347(6221), 536–539. http://doi.org/10.1126/science.1256297

7 Berinato, S. (2015, February 9). There's No Such Thing as Anonymous Data. Retrieved June 1, 2016, from https://hbr.org/2015/02/theres-no-such-thing-as-anonymous-data

8 Sweeney, L. (2015). Only You, Your Doctor, and Many Others May Know. Technology Science. Retrieved from http://techscience.org/a/2015092903/

9 "Data sharing best practices"

10 "Ethical algorithms for sense and respond systems"

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