Big Data, Analytics and Ethics: Lawyering in the Information Age

Bennett B. Borden[1]

"The only thing that is constant is change."

- Heraclitus of Ephesus (c. 500 BCE)

Each day, more than 2.5 billion gigabytes of new data is created.[1] 90 percent of all data created throughout the entirety of human existence was created only in the past few years. The dramatic and accelerating growth of digital information, and our reliance upon it, is an unprecedented challenge (and opportunity) for humanity as a species.[2]

For organizations with responsibility for governing and extracting value from this information, it creates a revolutionary and generational shift on par with the Industrial Revolution. However, in this revolution, data is the steam and the engine is analytics. But in this revolution, information is not just the raw material, it is **the asset**. In this revolution, the difference between winners and loser will be defined by who is the best steward of that asset.

Why Now?

"Over the past two decades (depending on the application) data availability has increased as much as 1,000-fold, key algorithms have improved 10-fold to 100-fold, and hardware speed has improved by at least 100-fold."

Harvard Business Review, July 2017[3]

It is critical for all practitioners who share responsibility for protecting, using, and extracting value from information to understand what is driving this monumental shift. There are two primary – but very different – sources.

---

[1] Bennett B. Borden is Chief Data Scientist at the law firm, Drinker Biddle & Reath. He is a partner in the litigation section and Chair of the Information Governance and eDiscovery Group.

Human Data

Until very recently, all recorded information was human generated. 40,000 years ago, even before complex spoken languages existed, humans were capturing information about themselves and their world in the form of cave paintings, like those in Sulawesi, El Castillo, and Lascaux. Our ability to capture and record information accelerated with the invention of the printing press, and then a comparatively short time later, the microprocessor.

Not only was all data created by humans, until recently it was created on purpose. A human sat down with the intention of capturing a piece of information, whether for posterity, communication, entertainment, or other purposes.

However, this is no longer the case.

Machine Data

As this revolution unfolds, more and more information will be created by machines, or what the industry calls "things." These "things" include the plethora of computers, sensors, devices, and other non-human entities capable of creating and recording and/or transmitting information. This includes, for example, the FDA-approved ingestible sensor that transmits data to the internet indicating whether or not a patient has taken prescribed medication.[4] This includes the jet engine fitted with 5000 sensors generating 20 GB of data per second, and nearly a petabyte of data per flight.[5] It includes the trail of "data exhaust" we all create each day simply by going to work, eating, and sleeping while carrying our smartphones.

In 2009, the number of machines connected to the Internet surpassed the number of people.[6] There will soon be 26 times more things connected than people.

The Management Challenge

Human and machine-generated data come from different sources, and they represent different challenges and opportunities. The technology and techniques used to manage and use the data is different. Whereas legal and regulatory regimes addressing rights, obligations, and protections for human-generated information are well establish, a legal framework for machine data and IoT (Internet of Things) that encourages innovation while protecting society is still emerging.

Passively-generated data about human behavior have given us unprecedented insight into human behavior and conduct. That is a boon to providers of products and services, or any

organization seeking to understand how to most effectively fill the needs of the market. However, it is also fraught with ethical and other novel implications.

If an employer had the ability to automatically monitor and algorithmically assess the performance of each employee, and use that insight to drive recruiting, training, compensation, promotions, discipline, and dismissal, is that good for the employee? Is it good for society? Does it create a world we want to live in? How would this be treated by the law? Can the rights of the employee be balanced with those of the employer?

These questions are not theoretical. New technologies and techniques give employers an increasingly powerful set of tools to understand exactly what employees are doing at work. Simply using machine learning techniques to evaluate email and other digital communications can generate a myriad of insights from who is performing well to who is likely to commit fraud in the next two weeks. Although new forms of communication and collaboration continue to inundate the enterprise, trusty old email is not going anywhere – the volume of email sent and received by business users and consumers continues to grow, and is projected to reach over 300 billion daily messages by 2021. [7]

But employment relations is but one of an unfathomable number of use cases across sectors where the information revolution creates new opportunities and threats.

Every piece of information owned by an organization is a double-edged sword. Each is an asset, with the potential to drive better and faster decisions. But each is also a potential liability, driving management costs and representing the next embarrassing leak or piece of damaging evidence in a lawsuit. Given these opportunities, costs, and risks, one might expect organizations to manage their information well. However, for a variety of reasons, both organizational and technical, this is rarely the case. In fact, most organizations do not even realize how poorly they are doing.

Surprising as it may seem, taking control and governing this asset is rarely an organizational priority, even as regulatory and legal requirements become more stringent. Responsibility for mitigating risk often falls in the cracks between different departments. New information technologies – social media, cloud computing, mobile devices, IoT – create opportunities and challenges, while legacy systems are slow to disappear. Finally, the organizational and IT frameworks for information governance are often isolated and purely reactive, rather than strategic.

What this Revolution Means for Lawyers

A fundamental function of lawyers working across all domains is to find answers. Data about human behavior is increasing exponentially in volume, velocity, and complexity. Analytics can generate insights from this data more quickly and accurately than ever before. This has tremendous evidentiary value.

Most legal scenarios involve a series of facts, with actors doing things over a period of time. In a dispute, lawyers first work to outline the basics of the story, identify the actors, and zero in on the relevant time period. As an attorney then works to discover facts to fit into this framework about actors and their actions over time, analytics are a powerful tool for deeper insight, by for example, identifying patterns that are indicative or even predictive of actions and behaviors.

Getting to deeper insights empowers a lawyer to develop a competitive edge over an opponent by mastering the facts of the case first to create strategic advantage for clients. Analytics can be used for every aspect of the law.

The Informational Advantage in Litigation

In a "bet the company" whistleblower qui tam action alleging violations of the False Claim Act, e-discovery attorneys used advanced analytics to zero in on the data most likely to reveal the central facts of the case as quickly as possible. In this case, "quickly" was not just about containing costs (a relatively minor consideration given the stakes) but rather about gaining a strategic advantage – one that could profoundly affect the outcome of the case.

Attorneys focused their attention on 675,000 documents. Then, by using advanced analytics over the course of four days, discovered with certainty that the claims were completely baseless. This was done before the answer the Complaint was even pled.

Armed with this insight, they voluntarily produced 12,500 documents that precisely illuminated the positions of each party. They then requested a meeting with plaintiff's counsel and at that meeting walked them the evidence, laying out fact after fact based on the evidence they had already produced.

Within days, the case settled for what amounted to nuisance value based on a retaliation claim—without any discovery, and at a small fraction of the cost budgeted for the litigation.

This case illustrates how the information revolution underway can profoundly affect case outcomes through an informational advantage. Getting to answers quickly and with certainty provides unprecedented strategic advantages.

The Informational Advantage in Business Transactions

The same capabilities can also be invaluable in supporting major corporate events like mergers and acquisitions. In such transactions sellers are typically required to make a standard set of representations that the financial and other material information they are disclosing is correct and has a reasonable basis under Generally Accepted Accounting Principles.

When deals fall apart, indemnity provisions protecting the buyer can trigger major valuation adjustments (potentially amounting to millions of dollars) should the information disclosed by the seller prove to be incorrect, or should the seller fail to disclose material information at all.

In reality, however, such indemnity provisions are rarely exercised, primarily because the window for the claim is short relative to the level of effort traditionally required to mount a claim. Algorithmic approaches to information retrieval, processing, and evaluation are starting to change this reality.

Attorneys have used fact development strategies based upon advanced analytics to obtain millions of dollars in indemnity claims for their clients. As soon as the merger or acquisition is completed, attorneys analyze the target's information systems and test the accuracy and validity of its disclosures. Within a matter of days, they are able to assess whether and to what extent the disclosures were inaccurate. The resulting indemnity claims are based not upon guesses but upon information from their own systems and communications. As many litigants have discovered, it is very difficult to defend against words coming out of their own mouth.

As made clear by these cases, the application of advanced data and analytics technologies and techniques to business problems can lead to profound information advantages that drive very different outcomes.

Ethical Implications of Our New Data Powers

Both public and private entities are using analytics to achieve their objectives and drive engagement. A university conducts analytics on information available about prospective

students to determine admission decisions and, when admitted, which classes, resources, and programs a student should have access to. A municipality uses analytics to deliver public services, such as power distribution, trash collection, and street maintenance. In the private sphere, an internet-based retailer makes product recommendations based on models that analyze consumer behavior. Using predictive models, a media streaming service predicts what movies or shows a consumer might enjoy, and a retail pharmacy predicts which coupons might entice a consumer to purchase a product.

A search engine, meanwhile, serves up different advertisements for different people, based on algorithms that build a profile of who it thinks you are and in what it thinks you are interested. A search engine also gives different people different search results based on algorithms that analyze our personal and behavioral information. These algorithms lead to people getting put into groups and then getting different access to goods, services, and economic or experiential opportunities based on group membership. Putting people into groups and offering them things (or refusing to offer them things) based on their group membership is the definition of discrimination.

Algorithmic products and services are bringing both significant advances and considerable risks to our society. Algorithms are usually always right, but they lead to significant legal and ethical concerns over how we interpret or use them. An airline, for example, uses algorithms to optimize service delivery and maximize profits. However, those algorithms can prioritize employee transportation over passenger revenue, and bump people from flights – even after the plane is boarded. Meanwhile, a retailer tracks in-store customer traffic patterns by using algorithms that pull transponder data from mobile phones and Bluetooth beacons. This raises issues of notice and consent. An advertiser monitors how consumers respond to their ads by using video recording in public spaces and algorithms that analyze facial expression. This raises significant issues of privacy and data ownership.

Laws surrounding these advances have not yet been formulated. Thi sis because the law is reactive by nature. In our *lasseiz faire* society, we encourage innovation and entrepreneurialism. We want new products and innovations to further society. We allow these new products and services to enter the market, and have tort-based remedies if they cause harm. But for the law to act, the harm has to be suffered first. When new products and services enter the market, they often clash with established products, creating new kinds of conflict the legal system has not seen before. As these conflicts make their way through the courts, reasonable conduct is defined as each case creates new contours of acceptable and reasonable conduct. For example, when the automobile was created, it

interacted on roads with horses and carriages. We didn't have laws governing such interactions, because they simply hadn't happened before. The law defining what constituted reasonable conduct for a driver and a rider developed over time, and then the horse and carriage was largely replaced by the automobile. This is how the law develops when it encounters new products and services in times of technological innovation. And thus the law lags behind current societal interactions. And, the greater the pace of technological change, the greater the law lags behind.

This is where we find ourselves in the current technological revolution. We are creating, collecting and analyzing data about human conduct that has never existed before. We are creating new products and services that were never possible before. And this is creating new risks in society that we've never experienced before. As information age lawyers, we are well positioned to understand the legal and ethical implications of using this new data and the algorithms that analyze it in order to counsel our clients appropriately.

New products and services create new risks. Companies building algorithmic products and services should champion efforts to define reasonable conduct to gain the greatest advantage possible while protecting against risk. Most law is based upon reasonable conduct, meaning, what would a reasonable person do with the information known at the time. Companies stepping into the legal greenfield of new algorithmic products and services can help define what constitutes reasonable conduct by identifying and quantifying the new risks they are introducing into the market place and acting reasonably to mitigate them. A voice-controlled digital assistant, for example, introduces a number of new risks to the marketplace involving privacy, data ownership, liability, and constitutional issues. Effective mitigation strategies bring together experts, such as privacy, consumer rights, psychology, criminal, and education specialists, to identify and quantify risks created by the always-listening assistant. Creating reasonable conduct enables companies to identify and quantify risks, identify mitigation strategies, and include a diversity of opinion and expertise.

Final Thoughts

Harnessing the value of analytics is one of the foundational requirements for information age lawyers to get to facts quickly and accurately. The use of analytics is bringing tremendous advances to our society and creating more efficient markets. But, analytics also bring us unparalleled potential for abuse. Like the advent of new technology in other

transformative periods, analytics are highly disruptive. Algorithms determine your perception of the world and your access to goods and services in that world. This introduces news risks and legal and ethical implications. Understanding and embracing analytics empowers us to create more efficient products while minimizing the associated risks.

Endnotes

[1] IBM, November 2016, online at, https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN

[2] The terms "data" and "information" have formal information science definitions that are important for discussing complex theories and practices. However, because a goal of this piece is to provide perspective on complex topics in a way that is useful to a general business audience, the terms "data" and "information" are used interchangeably throughout, as is common outside the academic context.

[3] HBR.org, "AI, For Real: What's Driving the Machine Learning Explosion?" July 18-27, 20127. Online at, https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence

[4] FDA News Release, "FDA approves pill with sensor that digitally tracks if patients have ingested their medication," November 13, 2017. Online at,

https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm584933.htm

[5] AviationWeek, "Internet Of Aircraft Things: An Industry Set To Be Transformed," Jan 18, 2016, online at, http://aviationweek.com/connected-aerospace/internet-aircraft-things-industry-set-be-transformed

[6] FTC Staff Report, "Internet of Things: Privacy and Security in a Connected World," January 2015.

[7] The Radicati Group, Inc., Email Statistics Report, 2017-2021, February 2017. Online at, http://www.radicati.com/wp/wp-content/uploads/2017/01/Email-Statistics-Report-2017-2021-Executive-Summary.pdf