

INTRODUCTION

DEFINING THE DATA MOVEMENT

BY JUSTIN HIENZ

Today, there is a rapidly growing capacity to collect, store and analyze massive amounts of data, far more than an individual mind could process on its own. This enormous volume of information has been called Big Data, a term that is widely used, sometimes to the point of cliché. Yet, while the term can be trite, the dramatic potential in exploring large datasets for new insights, trends, and correlations is anything but.

The data movement is a force for good. It is fodder for research and a catalyst for innovation. It is the bedrock of informed decision-making and better business and the key to unlocking more efficient, effective government and other services. It unleashes economic growth, competition, profitability, and other breakthrough discoveries. And it is at once a product of an ever-more technologically sophisticated world and a tool to advance, enhance, and shape all of its domains going forward. This widespread emergence and use of Big Data is revolutionary, and history will record the early 21st century as the beginning of a data revolution that defined a century.

There is no shortage of examples of data-driven decision making and innovation. Less common, however, is scholarship that looks at myriad examples to extrapolate the ideas, themes and potential that define the data movement and the changes it will bring. This report begins to fill that knowledge gap, with leading scholars and practitioners looking to the horizon to describe the data-driven future.

The world is but a few steps down the data road. In time, the very notion of “Big Data” will fade as data-driven decision making becomes a ubiquitous and unquestioned piece of everyday life. Yet, the way we understand and embrace the data movement now will shape how it impacts all of our futures. This report informs the ongoing discussion to reveal how data impacts our lives, economies, societies, the choices we make, and, inevitably, changes everything for the better.

The Rise of Big Data

Because this is only the beginning of the data-driven movement, the terms, definitions, and ideas associated with Big Data are still evolving. This is a subject area that is as dynamic as it is amorphous. One can no more wrap their arms around a tidal wave than

“THE AMOUNT OF DATA GENERATED IN TWO DAYS IS AS MUCH AS ALL DATA GENERATED IN HUMAN HISTORY BEFORE 2003.”

they can nail down precisely all that is grouped under the Big Data epithet. That said, there are some clear properties of the data landscape.

Most definitions of Big Data draw from Doug Laney’s often-cited “three Vs,” each of which describe a component quality of Big Data: volume (the amount of data); velocity (the speed at which data is created); and variety (the types of data).¹ Of late, Big Data definitions have come to include a fourth V: veracity. As shown throughout this report, data accuracy is as important to realizing value as the size, type, and generation of information.

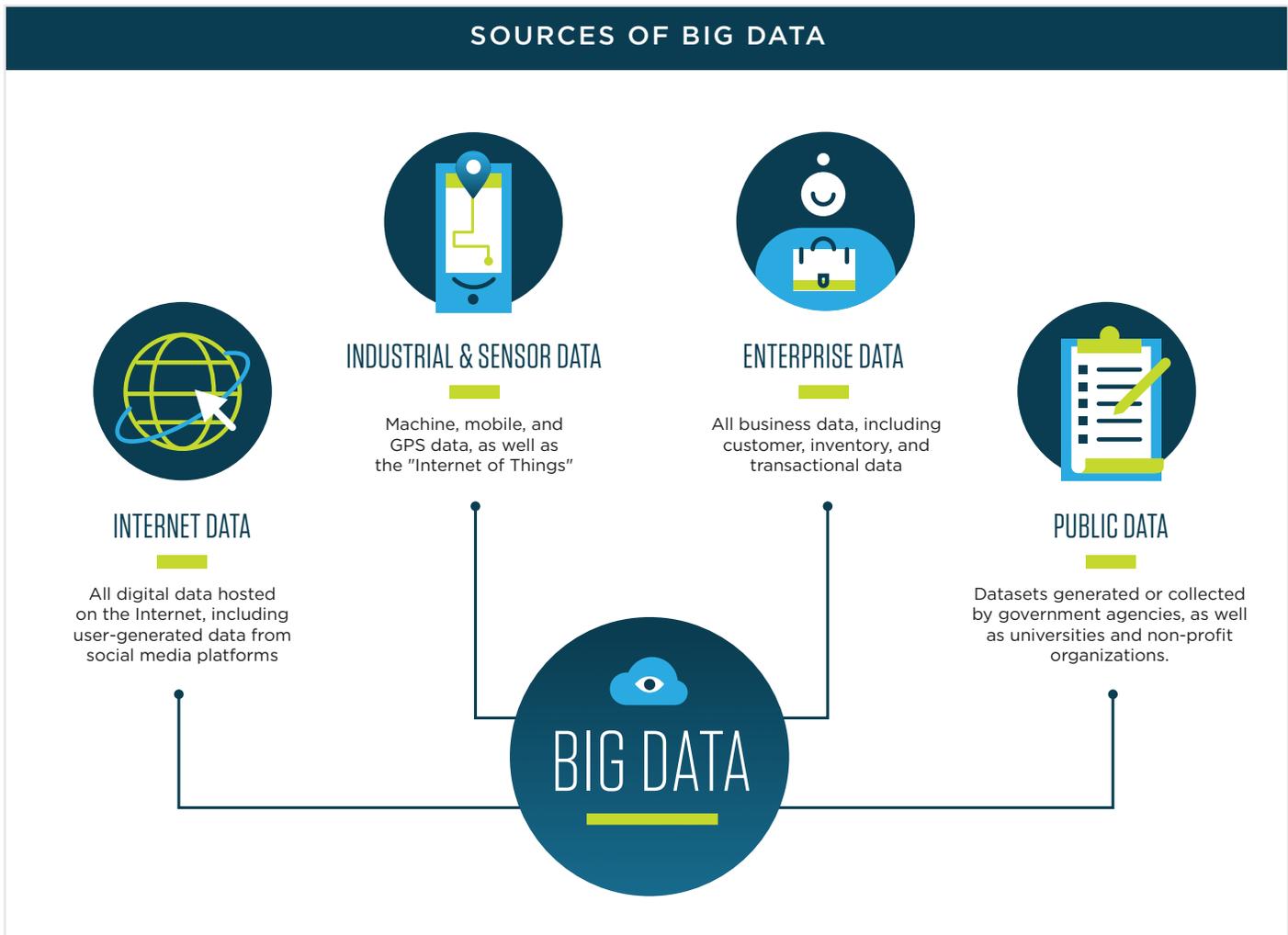
Big Data is so voluminous and generated at such a rapid pace that it cannot be effectively gathered, searched, or even understood by the human mind alone. As such, technology is the medium through which the data movement thrives. Data’s volume and variety owes to the growing number of sensors and connected devices that permeate every aspect of industrialized society (the so-called “Internet of Things”). In addition, the digitization of commercial transactions, medical records, online social communication, and other information also contributes to the amount of data. And then there are datasets from governments, research groups, and international organizations, all of which create and consume data in their activities.

Collectively, in 2012, these and other sources generated 2.5 quintillion bytes of data every day.² To put that in perspective, as Google CEO Eric Schmidt said at a 2010 Techonomy conference, the amount of data generated in two days is as much as all data generated in human history

before 2003.³ Thousands of years of civilization, millions of books, every piece of information from the ancient Library of Alexandria to the modern Library of Congress, all of that data together is but a miniscule drop in the proverbial bucket.

Given the volume, growing alongside data generation has been the capacity to hold and access it. Ever-more ubiquitous Internet makes it possible to easily share information without regard to geographic distance or observed borders, a capability unique to our modern age that allows all professionals, businesses, and organizations to share, collaborate, and advance knowledge like never before. These online advantages, as well as advances in computing power, facilitated innovative approaches to storing the exponentially growing volume of data. This has made it practical to keep and analyze entire datasets (rather than just down-sampled portions), dramatically expanding the power and promise of Big Data.

Even as the world has made great technological advances in collecting and storing data, once in hand, making sense of all this information is a challenge unto itself. Data scientists are in great demand, with many businesses and organizations creating data-specific positions and hiring data experts who can use complex algorithms and computer programs to find correlations and trends within troves of information. Indeed, while Big Data and technology are natural bedfellows, the human element is not obsolete. Leslie Bradshaw writes in Chapter 3 that until computers can “think” creatively and contextually, the human brain will remain a critical component in turning raw data into actionable insight. Indeed, data does not



replace human thinking; it enhances it. The brain is the vehicle for innovation, and data is its fuel.

A Diverse Data Landscape

Big Data is a broad, inclusive term. It may refer to spreadsheets packed with numbers, but it also includes product reviews online, the results of procedures in medical files, or the granular data that accompanies all online activity (called metadata). Shopping histories, biological responses to new pharmaceuticals, weather and agriculture trends, manufacturing plant efficiency—these and many other kinds of information make up the complex, varied data landscape.

Data is an asset. As such, much of the data generated every day is proprietary. An online retailer owns the data listing its customers' purchases, and a pharmaceutical company owns data from testing its products. This is appropriate, since businesses bear costs to generate, store and analyze data and then enjoy the innovative fruits that grow out of it.

Yet, there is also great value in providing data freely to any interested party, a philosophy called Open Data. Whereas Big Data is generally defined by its size, Open Data (which can be "big") is defined by its use. There are two core components of Open Data: data is publicly available, licensed in a way that allows for reuse; and data is relatively easy to use (e.g., accessible, digitized, etc.). Free or low-cost data offered in a widely useable format unleashes enormous potential. It exposes data to more minds, interests, goals, and perspectives. If two heads are better than one, how much better are dozens, hundreds, or even thousands of minds digging through data, looking for valuable correlations and insights?

Joel Gurin writes in Chapter 6 that there are four kinds of Open Data driving innovation: scientific; social; personal; and governmental. From genomics to astronomy, researchers are taking a collaborative approach to working with scientific data. Meanwhile, businesses and other

organizations are seeking insight via social data (e.g., blogs, company reviews, social media posts), which can reveal consumer opinions on products, services, and brands. And new digital applications are giving citizens access to their own personal data, yielding more informed consumers.

Perhaps the most common (and most robust) Open Data, however, comes from the public sector. For example, in June 2014, the U.S. Food and Drug Administration (FDA) launched OpenFDA, a portal through which anyone can access publicly available FDA data. This initiative is designed, in the words of FDA Chief Health Informatics Officer Taha Dr. Kass-Hout, to “serve as a pilot for how FDA can interact internally and with external stakeholders, spur innovation, and develop or use novel applications securely and efficiently.”⁴

A subset of the Open Data concept is what McKinsey calls MyData.⁵ This is the consumer-empowering idea of sharing data that has been collected about an individual with that individual. MyData fosters transparency, informs consumers, and has myriad impacts on commerce and cost of living. McKinsey offers the example of utility companies comparing individual and aggregated statistics to show consumers how their energy use stacks up against their neighbors. Personal healthcare data is another example where sharing MyData with an individual can yield greater understanding of one’s wellness and habits.

Yet, despite the evident benefits, it is when discussing this kind of data that concerns are sometimes voiced over the potential for personal information to be used to the detriment of the individual. In the public data discussion, this anxiety is often reduced to an overly simplistic

cry for “privacy.” When the data-privacy nexus is approached from a scholarly, objective viewpoint, however, it becomes clear that data does not present an inherent threat to privacy. Rather, the relationship between the individual and the public and private sector organizations that hold data about them is best viewed as a form of trusteeship. When the trustee (the organization) fails to uphold obligations for securing and fairly using personal data, they commit what Benjamin Wittes and Wells Bennett call “databuse” (see Chapter 7). This potential misuse of data is most deserving of discussion, rather than a fear-driven debate over vague visions of privacy.

Data-Driven Innovation

Data is a resource, much like water or energy, and like any resource, data does nothing on its own. Rather, it is world-changing in how it is employed in human decision making. Without data, decisions are guesses; with it, decisions are targeted, strategic, and informed. These lead to better business, better government, and better solutions to address the world’s woes and raise its welfare.

Data has attracted the excitement and attention it has because of the massive potential in its application. Data-driven innovation, as Dr. Joseph Kennedy describes in Chapter 2, has enormous economic value, with Big Data product and service sales exceeding \$18 billion in 2013, expected to reach \$50 billion by 2017.⁶ This value comes in the form of: new goods and services; optimized production processes and supply chains; targeted marketing; improved organizational management; faster research and development; and much more. It could include companies developing consumer products based on customer surveys, energy producers using geological studies to find oil, or

“DATA IS A RESOURCE, MUCH LIKE WATER OR ENERGY, AND LIKE ANY RESOURCE, DATA DOES NOTHING ON ITS OWN. RATHER, IT IS WORLD-CHANGING IN HOW IT IS EMPLOYED IN HUMAN DECISION MAKING.”

financial firms using corporate data to advise investors.

As well as new products and services, Big Data also yields value through increased competitiveness. As John Raidt writes in Chapter 4, the United States enjoys numerous attributes that will allow the country to take the fullest advantage of the data movement, more so than any other country. The United States' longstanding technological leadership, its free market system, its research and development infrastructure, its rule of law, and a host of other national qualities make the country the most fertile for data-driven innovation and all the economic, societal, and competitive benefits that come with it.

Accessing all this value, however, depends in part on the policies guiding data gathering, usage and transmission. Matthew Harding writes in Chapter 5 that the United States (and the world) requires public policies that foster innovation and growth

while protecting individual freedom and restricting potential “databuse.” Many of these policies already exist in an effective form. In any case, the policies we set and uphold today will in part define how data is used in the future. There is an important role for public policy in this emerging phenomenon, but to extract the most value from data, policies must be developed carefully and in collaboration with private sector partners.

The data discussion today is less about where it started and what it is and more about where it's going. With the above definitions and descriptions commonly agreed across industries, this report takes the next step, gathering ideas and examples to describe all the ways data is contributing to a stronger economy, improved business and government, and a steady flow of world-changing innovations. The Big Data landscape is an exciting place, and the chapters that follow offer a window into how precisely data is changing everything around us—for the better.

ENDNOTES

- 1 Doug Laney, “3-D Data Management: Controlling Data Volume, Velocity and Variety,” *Gartner, Inc.*, 6 Feb. 2001.
- 2 IBM, “Bringing Big Data to the Enterprise,” <<http://www-01.ibm.com/software/sg/data/bigdata/>> (16 Aug. 2014).
- 3 Marshall Kirkpatrick, “Google, Privacy and the New Explosion of Data,” *Techonomy*, 4 Aug. 2010.
- 4 Taha Kass-Hout, “OpenFDA: Innovative Initiative Opens Door to Wealth of FDA’s Publicly Available Data,” *openFDA*, 2 June 2014.
- 5 James Manyika et al., “Open Data: Unlocking Innovation and Performance with Liquid Information,” *McKinsey Global Institute*, Oct. 2013.
- 6 Jeff Kelly, “Big Data Vendor Revenue and Market Forecast,” *Wikibon*, 12 Feb. 2014.



ABOUT THE AUTHOR

Justin Hienz is the owner of Cogent Writing, LLC, a strategic content company working with businesses, organizations and thought leaders to advance their goals through writing. Hienz’s self-authored and ghostwritten work has appeared in *Foreign Policy* magazine, *The Chicago Tribune*, *The Newark Star-Ledger*, *CNBC*, *Roll Call*, *Homeland Security Today*, and numerous trade and company-owned blogs and websites. He holds master’s degrees in journalism and religion from the University of Missouri.