Dr. Joseph Kennedy is president of Kennedy Research, LLC and a senior fellow at the Information Technology and Innovation Foundation. His main areas of experience include the impact of technology on society, macroeconomic policy, and finance. Previous positions include general counsel of the U.S. Senate Permanent Subcommittee on Investigations and chief economist for the Department of Commerce.
Knowledge has always played a crucial role in economic activity and higher living standards. Human civilization is closely linked with the ability to transmit and record information. Similarly, scientific advancements depend on the increased use of objective data (rather than subjective dogma) as the best guide to understanding the world. The close link between data and our ability to intelligently shape our lives remains strong. Fortunately, our capacities are about to radically improve as new technologies and greater access to more and better data makes it possible to understand, control, and change much more of the world. This will have significant effects on the economy and living standards.

For millennia, massive amounts of oil and natural gas lay trapped within shale deposits underneath the United States, making no contribution to economic growth. Within the last decade, new technology has allowed us to exploit these previously inaccessible resources, even as some voices are decrying risks that come with taking advantage of this energy resource.¹

Big Data, as a resource, presents similar opportunities—and corresponding, less-than-rational concerns over potential consequences of putting data to work. These concerns are misplaced. More than any other newly tapped resource, Big Data has the potential to deliver large economic gains.

The benefits of this resource (but not its costs) increase rapidly as data is shared. The central challenge for public and private sector leaders is to apply this resource to the large variety of problems that now confront us while minimizing the relatively manageable risks associated with the greater availability of data. Just as the builders of the first oil well in Titusville, Pa., could not have envisioned the combustion engine and airplanes, we cannot foresee all the uses of cheap, abundant data. Yet, we can expect a wide variety of new products and processes that add economic value. Some of these improvements will increase traditional measures of economic growth. Others will have the primary effect of reducing costs and increasing consumer surplus. In every case, however, the advent of the Big Data era is bringing with it enormous economic potential.
The economic impact of Big Data will take a number of forms. A recent McKinsey report estimates that improved use of data could generate $3 trillion in additional value each year in seven industries.

Estimating Big Data’s Economic Impact
While Big Data is having a significant impact on the economy, that impact is difficult to measure. One reason is that the domestic use and international exchanges of data do not always show up in economic statistics. If access to large amounts of data is used to build a new business that sells consumer data to advertisers, the fees advertisers pay for the information will be counted in national income. But if better data allows hotels to meet the individual preferences of their guests without charging them more, all of the benefit will be captured as consumer surplus. The customer is better off, but because the value of economic transactions remains the same, national income is unchanged.

If improved visibility into its supply chains lets a retailer cut its prices in half, national income and perhaps employment would actually fall, at least until customers spent the savings on other items. As discussed in Chapter 4 of this report, Big Data can also increase competitiveness. Companies that increase value to customers without increasing cost will likely gain market share from their rivals. If U.S. companies take market share from foreign businesses, national income would rise.

The value of Big Data is closely tied to the growing Internet of Things—the integration of sensors and transmission capability into a wide variety of objects. It therefore benefits from continued progress in making sensors, transmission capacity, storage, and processing power significantly cheaper and better over time. This progress has led to a significant increase in data generation and capture from a number of diverse sources, including financial transactions, social media, traffic patterns, medical treatments, and environmental conditions.

As data becomes more accessible, it will affect the economy in a number of ways, all of which can be loosely encompassed as being part of the data-driven economy. These impacts include:

- Generating new goods and services, such as GM’s OnStar or custom-tailored clothing, in which information is either the product itself or it contributes significantly to the quality of another product.
- Optimizing production processes and supply chains, such as what Walmart has done with its stores.
- Targeted marketing, including the integration of customer feedback into product design.
- Improved organizational management often in the form of using data to make better decisions.
- Faster research and development, which shortens the trial and error process of innovation.

The economic impact of Big Data will take a number of forms. A recent McKinsey report estimates that improved use of data could generate $3 trillion in additional value each year in seven industries. Of this, $1.3 trillion would benefit the United States. McKinsey also estimates that more than half of this value will go directly to consumers in the form of things like shorter wait times in traffic, improved ability to comparison shop, and better matching between schools and students. The rest will go to companies that either create new products centered around the use of data or use data to gain an edge over their competitors. Walmart, GM, and other companies are already using Big Data to offer new products, improve their margins, and take market share from their rivals. Walmart’s use of Big Data to streamline and improve its supply chain, for example, has led to a 16% increase in revenue over the last four years.
Access to better data may also improve the economic climate within which businesses operate. At present, macroeconomic policy is hobbled by the limitations of official government data. Accurate data is often associated with long lags, making it difficult for policymakers to know where the economy is, let alone where it is going. Although the most recent recession officially began in December 2007, the Bureau of Economic Analysis reported as late as June 2008 that the economy had grown during that quarter. Official data also often cover only a small part of the actual economy. It is possible that, by giving policymakers access to real-time information covering a much larger portion of actual transactions, Big Data could improve the ability of fiscal and monetary officials to avoid policy errors and allow businesses to time their investments more accurately.

Big Data is also having an enormous impact on international trade. Data flows are the fastest growing component of international trade. Another McKinsey report found that global flows of trade, finance, people, and data increased world GDP between $250 billion and $450 billion each year. This report also found that economies with more international connections received up to 40% more benefit than less connected economies.

**Private Sector Value and Potential**

Big Data will have a disproportionate impact on many industries. A 2013 report commissioned by the Direct Marketing Association measured the size of the data-driven marketing economy (DDME), defined as the set of firms that produce marketing services focused on individual-level consumer data for marketing firms. It found that in 2012, producers spent about $156 billion on these services, creating employment for about 676,000 people. The implication is that the buyers of this information derived at least this much value from it. Importantly, it found that roughly 70% of this value and employment depended on moving data between firms.

The study also found that the main benefit of the DDME was that it made marketing more efficient, allowing companies to avoid sending solicitations to individuals who are unlikely to buy their products and instead target prospective customers with offers that better match their needs and interests. A second benefit is that sellers are able to improve their effectiveness by matching specific marketing efforts with results. The DDME also reduces the barriers to entry for small manufacturers because it lowers the cost of obtaining and using high-quality consumer data. This benefit would not be available unless a robust market was allowed to exist in consumer data.

Because of their increased importance as an economic resource, restricting data flows can seriously hurt national welfare. A study by the European Centre for International Political Economy and the U.S. Chamber of Commerce concludes that implementation of the European Union’s proposed General Data Privacy Regulation would reduce EU exports to the United States by between 0.6% and 1%, undoing much of the potential impact from the proposed Transatlantic Trade and Investment Partnership. The negative results were reduced because the regulation would replace national data restrictions that are already in effect and allow for workarounds, such as model contract clauses and binding corporate rules to substitute for direct regulation. Eliminating these workarounds would have an even larger effect, reducing EU exports to the United States by 4.6% to 6.7% and EU GDP between 0.8% and 1.3%.

Further emphasizing the importance of data mobility, the Omidyar Network recently released an economic analysis of adopting the type of Open Data policies discussed in Chapter 6 of this report. The study concludes that implementation of Open Data policies could boost annual income within the G20 by $700 billion to $950 billion. Significantly, the benefits come in a wide variety of forms, including reducing corruption, improved workplace conditions, better energy efficiency, and a reduction in the regulatory costs associated with international trade.

The quest to gather and use consumer data has also generated a large increase in Internet advertising. A recent McKinsey study found that these ads underwrote the delivery of a range of free Internet services that delivered significant benefits to Internet users. The study estimated...
that in 2010, these services generated a social surplus (the excess of benefits over costs) of €120 billion. Significantly, 80% of this surplus went to consumers. Consumers will only continue receiving these benefits so long as advertisers receive value from funding them.

Much of the data impact is and will continue to be in the information technology industry, as the demand for sensors, data storage, processing capacity, and software increases. McKinsey cites studies that global data generation will increase by 40% per year. Nearly 80% of this is apparently copies of existing data. From 1986 to 2007, data storage and computing capacity increased by 23% and 58%, respectively. Virtually all of this information is now in digital form, making it much easier to copy, analyze, transmit, and store.

Then there is the impact on the labor market. Data analysis has been labeled “the sexiest job of the 21st century.” One estimate finds that there are already around 500,000 Big Data jobs in the United States. Still, the McKinsey studies point to a serious shortage of managerial talent capable of understanding and acting on Big Data. Most visible are the data experts with advanced degrees in statistics, computer engineering, and other applied fields. McKinsey finds a national shortage of between 140,000 and 190,000 people. But just as serious is the shortage of 1.5 million managers and analysts in traditional jobs who are capable of asking the right questions about the data and acting on the answers. Also important are the line employees who must properly implement data strategies. In fact, the inability to find and keep workers with even moderate math and statistics skills is already placing limits on business profitability. The demands of the data-driven economy will only exacerbate the current shortage of well-educated workers.

The impact on labor markets will not be totally positive. Some have expressed concern that a data-driven economy will eliminate jobs through a combination of automation and increased competition. Technology has frequently produced highly disruptive changes to the economy, and the pace of these changes may well increase as a result of future advancements in information technology. Yet, an international study by McKinsey found that within small- and medium-sized enterprises, the Internet (and by implication, the data-driven economy) created 2.6 jobs for every 1 it eliminated.

Finally, cloud computing has increased the power of the information system. The ability to lease cheap storage and processing power has two important economic impacts. First, it transforms a large fixed cost into a variable cost. Companies can avoid having to purchase and maintain their own data centers and write or purchase their own software programs and instead lease both on an easily scalable, as-needed basis. Sophisticated data strategies no longer require large up-front capital costs and deep expertise in computer maintenance. Second, even the smallest companies now have access to the fastest servers and most sophisticated processing power at affordable rates. By making it easier for all companies to enter new and existing markets, cloud computing should increase both the diversity and competitiveness of markets.

The Role of Data in the Economy
The challenge for businesses will be to find the necessary talent that allows them to discover the true causal relationships within the data. They must then use these relationships to implement profitable business strategies in ways that do not violate public expectations about the proper uses of data. Developing and implementing successful innovation based on data insights will often be the hardest challenge. And they must often do this in real-time, responding even as the causal relationships change.

Much of the value in Big Data is likely to come from combining Big Data with the Internet of Things. Cheap sensors and transmission capacity can be used to generate enormous amounts of fresh data, which can then be fed into a system capable of analyzing and acting on it to solve existing problems. Figuring out how to act on the resulting flow of information, however, may not be easy or cheap. Consider parallels to the introduction of electricity into manufacturing plants, which forced a significant reengineering of manufacturing activity as plants that were built to harness other forms of power tried to optimize the value of this new resource.

Companies can create value by using data to solve problems. Some problems are unsolvable in the sense that the data needed to solve them does not exist, anywhere. A good example at present is
how to cure late-stage cancers. For most, however, the challenge is that the data needed to solve a problem have often been difficult to collect or are mixed in with other, unrelated data.

The promise of the information system is that it makes it possible (and increasingly affordable) to collect the right information, process it into actionable knowledge, transmit that knowledge to the right person, and act on it. In doing so, it allows us to solve an increasing number of problems, including many that we had never thought of. Sometimes the border between solvable and unsolvable problems is fuzzy. For example, we currently do not often know much about why some students learn at different speeds. Yet, it may be that if we had more data about all students, including individual students’ strengths and weaknesses, we could design more effective educational software.

Companies that want to use Big Data to improve their internal operations will need to identify the key problems currently holding them back. These might include poor inventory management, difficulty retaining customers and workers, or poor decision making. UPS has increased its profit margins by collecting detailed information about the location and performance of their large vehicle fleet. Data can also help businesses spot previously unrecognized problems. Gathering data on the water use of individual households, for example, can help municipalities identify anomalies that deserve a closer look.

Companies seeking to develop new products or services need to identify unsolved problems among their customers. For instance, every car driver has occasionally become lost. Every parent has wondered whether their child is learning what she should. Every diabetic needs to know his blood sugar level. To be successful, companies will have to identify these challenges and needs and then figure out what data is needed to solve them, transmit it to the right place at the right time, and act on it, all in a way that is intuitive for the customer. One key rule of technology in general, and data in particular, is that it will sit idle if it is too difficult to use.

Overall, the greatest value will go to companies that can identify unmet problems or needs, both within their own operations and among their customers. It sounds easier than it is. Executives have difficulty imagining new management opportunities (such as Total Quality Management and Six Sigma), and customers have difficulty articulating a need for products like the iPhone and FitBit. The prior infeasibility of collecting and using the necessary data is likely to have prevented the recognition of many problems. That is what made both Pandora and the Nest thermostat such innovative products.

Management theorists, including W. Edwards Deming, have broken down the process of continuous improvement into a four-part cycle: plan, execute, measure, and adapt. Data, and the ability to understand it, is critical to this process. It allows people to compare planned results with actual outcomes and then adjust their future actions to reduce the gap between the two. The key challenge often lies in deciding what to measure and integrating the right information into a process designed to improve performance. Toyota and other companies, for example, used this to great effect in the 1980s. A key part of Toyota’s process was collecting detailed information about the manufacturing process, including production rates and quality measures. The company then used this data to spot problems, identify their root causes, and implement lasting solutions. The result was a dramatic improvement in profit and market share because of better quality, lower costs, and shorter production cycles.

Big Data transforms this process in several important ways. First, the time lags involved in collecting and analyzing data often imposed a significant delay between execution and measurement in the improvement cycle. For example, the negative effects of changes in maintenance procedures might not become apparent until machines begin to fail faster than normal. With cheap sensors and rapid transmission, companies increasingly have instantaneous insight not only into their own performance but also into the performance of their products long after they have left the factory. This allows for a closer connection between plan implementation and data response and permits companies to develop and produce further iterations much faster.

Second, increasingly granular data allows companies to improve performance through A/B testing. For example, by varying product layout slightly between two stores and then measuring daily traffic and sales in each location,
management can see which of the two variations leads to better results. Results show that minor changes in the layout of a Web landing page can increase customer inquiries by more than 300%. Which version is better, however, is not always obvious until one has the data.31

Third, the ability to store and process large amounts of data allows companies to search for subtle relationships between them. Whether it is Target analyzing the collective buying decisions of millions of shoppers, geneticists looking at the combination of millions of genomes and personal histories, or educators studying data on learning outcomes, in many cases, meaningful relationships between different factors do not become apparent until researchers have lots of data points to study. This is especially true with many factors, each of which slightly influences the probability of a given outcome. For instance, in Boston, Bridj looks at a large collection of data feeds from Google Earth, Foursquare, Twitter, Facebook, the Census, and other sources to figure out where commuters are and where they want to go. It then arranges temporary bus service to meet the demand. Rather than view these new routes as fixed, Bridj tries to respond to market changes.32

Yet, the mere presence of Big Data does not guarantee economic profits. In fact, firms may become misled because the link between more data and better outcomes is not perfect. As Tim Harford has pointed out,33 enthusiasts for Big Data have made four important claims: (1) data analysis produces extremely accurate results; (2) every data point can be captured; (3) we do not need to understand why data are correlated; and (4) scientific models are unnecessary. At best, these are generalizations that, if taken for granted, can lead to poor business decisions.

A fundamental assumption is that data samples are unbiased. Although tests exist to detect and treat bias, they are not perfect, and companies may still find themselves using data that are not representative of the population they supposedly represent. It is often the case that the data produced by the exponentially growing number of 21st-century technologies do not accurately reflect the characteristics of the entire population. Users of Foursquare and Twitter constitute a discrete subset of the American population, as do the consumers who take the time to register their products online. The most engaged and demanding customers are likely to have tastes that differ from those of the average shopper. In other words, the data that is easiest to collect may not be the most valuable. Companies that base their marketing strategies on data generated by a subset of the market may find that their product appeals to only that minority.

Second, firms will continue to face a trade-off between carefully targeting high-value customers and focusing on too narrow a portion of the potential market. Big Data might help them distinguish between the two but only to a point. It will still be the case that efforts to selectively target only the most promising customers will miss a lot of potential business, while campaigns that more broadly target all potential customers will spend money on people who were never potential buyers.

The size and science of Big Data can easily lend a false sense of precision to the equations that come out of it. As an example, much has been made of the fact that Target once sent ads for baby products to a teenager before her father knew she was pregnant. To know exactly how significant this was, however, we would need to know what proportion of women who received those ads were actually pregnant. If Target sent them to everyone, the significance of the event would disappear.

The bigger data is, the more likely it is to contain spurious relationships. This is especially true when the data pile contains a large number of variables and managers are mining the data for correlations. The problem is that chance correlations often exist. Because these relationships are spurious, they are unlikely to persist going forward. Thus, basing a decision on these correlations is dangerous. The next data pile may also contain lots of correlated variables, but they are likely to be different from the first set. In the meantime, companies that relied on the first set of correlations will have aimed their resources in the wrong direction.

What is more, even if correlations between data are causal, there is no assurance that these relationships will continue into the future. Indeed, the key to business success may often lie in disrupting traditional business relationships and transcending product boundaries by offering innovative products that redefine the market and play to the strengths of a particular firm. Because
these products are new, there may be little data to guide executives.

For data to add value, its use must be properly inserted into an institutional setting. For example, Street Bump is an app that mobile users can download to help the city of Boston locate potholes. The app sends a signal to the city every time its owner passes over a pothole. The idea is ingenious, but it is only likely to make a real difference if identifying potholes is the binding constraint in the system. If without Street Bump potholes would be identified within two days anyway, then the added value may be small. If the potholes are otherwise identified by city workers who are sent around to look for them, the app may save money but only if the workers are let go or assigned to more productive work. Finally, if the real constraint on street repair is a shortage of funds, equipment, or labor, Street Bump may only add to the backlog of unfilled potholes and have little impact on actual street quality.

The Social Dimensions of Big Data
Any resource’s value depends upon the social, legal, political, and economic environment surrounding it. For instance, the progress of fracking in the United States compared to other countries is at least partially the result of the facts that: U.S. landowners own and can convey the mineral rights to their property; the energy industry had a large number of small, private companies that had to innovate or die once U.S. production started declining; and companies could easily raise money in open, sophisticated capital markets. Some of the primary factors that affect the economic potential of Big Data include ownership rights, security concerns, and the level of transparency surrounding data’s collection and use.

Ownership of data raises important social issues. It will often be the case that the ownership of data will not be linked to its possession. For example, individual users of social networking sites are likely to believe that they own the content they generate, irrespective of what the Terms of Agreement state. Those views are likely to drive legislation if the companies holding the data fight this perception too much. This concept of ownership might also extend to the full extent of what McKinsey calls MyData, which it defines as all of the data generated about a person, regardless of whether he is aware of it and whether he is the one collecting and storing the data.

MyData can therefore consist of much more than a Facebook or Twitter account. It includes one’s personal shopping history, be it on websites or in brick and mortar stores. Yet, it increasingly also covers a wide variety of other data, including: health data that users generate about their weight, sleeping patterns, and diets; volume and timing of utility consumption; and personal location. A growing number of providers are likely to compete in giving individuals greater access to this information and helping them understand and act on it.

Security
Security and privacy are intimately linked: the more private the data, the more securely people expect it to be held. We are coming to realize that security concerns represent a significant potential liability for firms. A large data breach can cost a company a great deal in terms of money, customer loyalty, and regulatory involvement. Companies that maintain possession of large amounts of consumer data, especially data that can be individually identified, need to implement the best security...
measures. This may require enough expertise to justify involving outside companies that specialize in data security and storage. To some extent, the problem is self-correcting; the boards of every major company are now focused on how to prevent a data breach.\textsuperscript{36}

**Transparency**
The growing use of Big Data is also likely to result in greater transparency. Companies have already put a lot of effort into identifying and tracking their best and worst customers. Big Data could allow them to have a personal relationship with each customer, tailored to that customer’s needs and individual preferences. Each individual may increasingly appear to both companies and governments as a unique entity with a full history of purchases, payments, income, goals, and more. This greater insight into people (and companies) can help providers deliver products that meet their customers’ deeper needs by identifying and responding to individual characteristics. It should also make it easier to form longer-lasting relationships, where value is based less on price and more on personal meaning to the customer.

It is true that transparency is likely to pressure profit margins by making it easier for customers to comparison shop, but it will also make it easier for buyers to verify a seller’s quality claims, strengthening the market for higher valued-added products. As the economy grows and higher incomes allow consumers to search for more personally rewarding experiences, the willingness to pay a premium for quality and tailored products and services should increase.

In the same vein, companies are becoming increasingly transparent to consumers. With the aid of social networks, third-party data aggregators, and mobile technology, customers have access to much more information about the quality, structure, and ethics of the companies with which they interact. They will increasingly be able to trace their food back to the farm or factory from which it originated. Producers may find themselves competing with each other to give consumers an open view of their processes, perhaps even by placing cameras in their production centers so that customers can verify claims of safety. This trend could make each company within a given
supply chain (particularly those whose brand is attached to the product) more responsible for the performance of the entire supply chain.

**Conclusion**

Financial derivatives are a standard business tool for managing a wide variety of risks. When they were first adopted, however, many companies were focused on their novelty rather than on the business case for using them. As a result, a number of firms suffered large losses. It took time and hard lessons for companies to learn how to properly integrate derivatives into normal business operations. The same learning process is already taking place with Big Data. Like financial innovation, it is an always-evolving concept, requiring constant education and adaptation. That said, it offers huge rewards for those who succeed in using it to create value for customers through better products and services.

By increasing the availability of data and reducing its cost, the data-driven economy promises a smarter, more efficient world. But it will not solve all problems. The path between gathering data and acting on knowledge involves many steps, not all of them subject to improved technology. Many of today’s most pressing problems involve a difference of values rather than a disagreement about facts.

Big Data represents one of the largest untapped resources yet. Thanks to continued advancements in information technology, it is finally being tapped. Together with the rise of the Internet of Things, it constitutes a general purpose technology. Such technologies have broad impacts on the economy and society. The full impact from Big Data and related technologies will be spread out over several decades. This is partially because use of any new resource or technology often requires a significant transformation of the status quo. It takes time for people to think of new ways to use the resource and implement the necessary changes. Nevertheless, the promise of Big Data is transformative and its economic impact expansive, cascading, and world changing.

**ENDNOTES**


12. Ibid.

ENDNOTES CONTINUED


15 Ibid., 15.


17 Ibid., xiv-xv.


20 Ibid., 19.

21 Ibid., 16-17.


23 Michael Mandel, “Where are the Big Data Jobs?” Progressive Policy Institute, May 2014.


28 Total Quality Management and Six Sigma were disciplined methods of improving quality within a production process. Both methods stressed careful observation of existing processes, the collection of large amounts of performance data, and investigations to discover the root causes of production problems.


