



## STRATEGIC FORESIGHT INITIATIVE

By Banning Garrett

# Big Data Is Changing Your World... More than You Know

“This is some of the best driving I’ve ever done,” Steve Mahan joked at the end of a ride in the Google self-driving car. Mahan’s 2012 drive—to buy tacos for lunch and pick up his laundry—was especially remarkable since he is 95 percent blind. His hands-free test drive (accompanied by Morgan Hill Police Department Sergeant Troy Hoefling and recorded in a YouTube video<sup>1</sup>) would have been impossible not only without advanced sensors, computers, and software, but also without big data, which both enabled development of the driverless car and inform its movement along the streets and freeways of California. The Google car itself gathers nearly 1 gigabyte of data per second as it scans and analyzes its environment. Think of the potential data gathering of 100 million self-driving cars on the roads of the United States. How will that data—100 million gigabytes per second—be transmitted, stored, and analyzed?

“Big data” may not be as exciting as a driverless car, but is an increasingly important facet of nearly all aspects of life. It will be largely invisible to most people, even as it increasingly informs and shapes decisions of business, government, and individuals. Big data may exist only as digital bits, but it comes from the material world and serves the material world.

Big data is enabling quantification and analysis of human life, from the macro down to the micro level—

<sup>1</sup> See <http://www.youtube.com/watch?feature=endscreen&v=cdgQpa1pUU> E&NR=1.

### Emerging Technologies and Society

The Emerging Technologies and Society project is a collaboration between Singapore’s Risk Assessment Horizon Scanning Programme Office (RPO) in the National Security Coordination Secretariat (NSCS) and the Atlantic Council Brent Scowcroft Center on International Security’s Strategic Foresight Initiative (SFI). Initiated by RPO, the project focuses on the political, economic, and societal impacts of significant innovations arising from the science and technology fields. Through a series of meetings with leading researchers and private enterprises in the Silicon Valley, the project explores topics ranging from ubiquitous robotics and its impact on human capital developments, to algorithmic risk, quantum computing, and their challenges to national security.

Through horizon scanning efforts, RPO enhances policy making capabilities through engaging analysis, robust processes, and leading-edge systems. The SFI, which strives to forge greater cooperation on futures analysis among its main partners around the world, has rapidly become a hub for an expanding international community of strategic planners in government and the private sector.



**Banning Garrett** is strategic foresight senior fellow for global trends and innovation in the Atlantic Council’s Brent Scowcroft Center on International Security. The author would like to thank Peter Haynes, Jay Wettlaufer, and Paul Saffo for their invaluable assistance in preparing this report.



Google's driverless car races around a test track. Photo credit: Flickr, Steve Jurvetson

that is, from the global scale down to the individual. It will present new opportunities for governments, businesses, citizens, and organizations. Big data could be used to help solve or manage critical global problems, lead to new scientific breakthroughs and advances in human health, provide real-time information and analysis on wide areas of life, wire up the planet's natural systems for monitoring and environmental remediation, greatly improve efficiency and resource use, and enhance the decision-making and daily operations of society.

Big data, powered by algorithms,<sup>2</sup> can inform decision-making and peer into the future. It can help with decisions even when the problem is not entirely understood, such as climate change, resource scarcities, species die-off, polar ice melts, and ocean degradation. Big data can be used to monitor performance and maintenance requirements for transportation systems, bridges, buildings, and other infrastructure. It can also be used to better understand supply and demand patterns in energy, food, and water—interdependent and highly non-

linear systems with asymmetric feedback loops. Policymakers and the public can better understand both the global situation and the local situation.

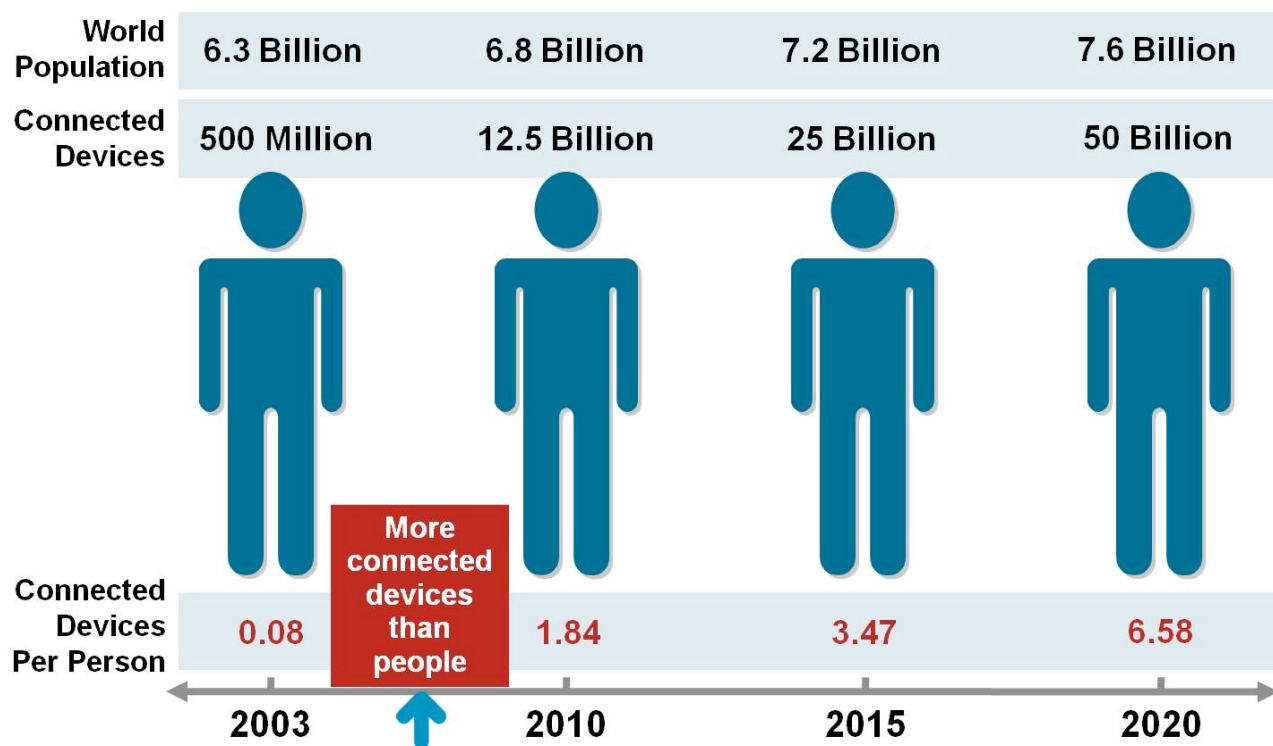
Much of the digital universe is transient, however, including phone calls that are not recorded (unless by governments), digital TV images that are watched (or “consumed”) but are not saved, packets temporarily stored in routers, digital surveillance images purged from memory when new images come in, and so on. Then there is stored but unused (at least in the short run) data that is forecast to grow by a factor of eight between 2012 and 2020, but will still be less than a quarter of the total digital universe by that time.<sup>3</sup>

Big data can also be misused. Spurious correlations made by big data analytics can affect an individual's future, such as employers making hiring decisions on the basis of what kind of Internet browser was used to file an online application,<sup>4</sup> or Google Flu Trends analysis being used in a pandemic to quarantine an individual who made a search correlated with flu symptoms because the Googler feared coworkers

2 For discussion of the key importance of algorithms to big data, see Banning Garrett, “A World Run on Algorithms?” Atlantic Council (Washington: Atlantic Council, July 2013), <http://atlanticcouncil.org/publications/issue-briefs/14346-a-world-run-on-algorithms>.

3 John Gantz and David Reinsel, op.cit.

4 “Robot Recruiting: Big Data and Hiring,” *The Economist*, April 6, 2013, <http://www.economist.com/news/business/21575820-how-software-helps-firms-hire-workers-more-efficiently-robot-recruiters>.



Source: Cisco IBSG, April 2011, [http://www.cisco.com/web/about/ac79/docs/innov/IoT\\_IBSG\\_0411FINAL.pdf](http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf).

exhibited those symptoms and wanted protection.<sup>5</sup> Even more ominous could be a “Minority Report” scenario in which correlations of certain behaviors and traits are seen as indicative that a particular person might engage in criminal or other anti-social behavior, thus justifying preventive action against that individual.<sup>6</sup> Big data analytics, when they rely on discovering correlations rather than determining causality, can lead to mistaken analyses and harmful decisions based on the “fallacy of numbers” and the “dictatorship of data.”<sup>7</sup> Big data also presents huge challenges for privacy, security, regulation, and the power relationship between citizens and the state, potentially leading to “big data authoritarianism”<sup>8</sup> rather than being used to enhance “big data democratization”—a concern heightened by the revelations about the National Security Administration (NSA)’s big data gathering, storage, and analysis programs. The potential misuses of big

data by the NSA and other governments has spurred new efforts in the United States, Europe, and beyond to regulate what big data can be acquired, how it can be used, and who will have access to it.

### Exponentially Growing Digital Traces of Human Activity

“Big data” has no official definition. It refers to huge volumes of data that are created and captured and cannot be processed by a single computer, but rather requires the resources of the cloud to store, manage, and parse. This size of data is on the order of a petabyte,<sup>9</sup> or one million gigabytes. The larger ecosystem of big data includes analysis with the results used to make decisions about the material or online world—whether these decisions are made directly by people or by other machines (such as the Google self-driving car sending and receiving data to guide its driving decisions milliseconds at a time). This digital universe is “made up of images and videos on mobile phones uploaded to YouTube, digital movies populating the pixels of our high-definition TVs, banking data swiped in an ATM, security footage at airports and major events such as the Olympic

5 Kenneth Neil Cukier and Viktor Mayer-Schoenberger, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (New York: Houghton Mifflin Harcourt, 2013) 169.

6 *Ibid.*, 157-158; 163.

7 Kenneth Neil Cukier and Viktor Mayer-Schoenberger, *op.cit.*, 166.

8 Kenneth Neil Cukier and Viktor Mayer-Schoenberger, “The Rise of Big Data: How It’s Changing the Way We Think About the World,” *Foreign Affairs*, May-June 2013.

9 For examples of the use of petabyte to describe data sizes in different fields, see Wikipedia, “Petabyte,” <http://en.wikipedia.org/wiki/Petabyte>.

Games, subatomic collisions recorded by the Large Hadron Collider at CERN, transponders recording highway tolls, voice calls zipping through digital phone lines, and texting as a widespread means of communications.”<sup>10</sup> Big data also includes billions of search engine queries, tweets, and Facebook profiles. While the amount of information individuals create themselves—writing documents, taking pictures, downloading music, making searches, posting tweets and “likes,” etc.—is huge, it remains far less than the amount of information being created about them in the digital universe, often without their knowledge or consent.<sup>11</sup>

Big data also includes data flows from potentially billions of sensors on natural and human-created objects, as well as on humans themselves, who are carrying data generators. For example, smartphones send information on an individual’s location, determined by GPS satellites, which returns to the smartphone as analyzed traffic information and appears on Google Maps. Virtually everyone will be connected by 2020, Google Chairman Eric Schmidt predicts, thus creating a world of 7 billion-plus unique nodes producing data through use of the Internet over smartphones, tablets, and other digital devices. Data is also produced through the use of machines that leave digital traces, from ATMs to airport kiosks, and from smart homes to offices and cars that will be awash with sensors monitoring every aspect of our lives.

Big data is increasingly critical in the day-to-day operations of society, connecting billions of sensors with each other, cloud computing, and the humans who are operating systems. This may include, for example, smart cars talking to the traffic system and to each other, or robots transmitting data to the cloud, using the cloud-computing capability in lieu of local computing power, and then communicating with human beings to perform tasks or address problems.

Nevertheless, big data is providing massive and exponentially growing digital traces of human activity

that is becoming a laboratory for understanding society. All kinds of data—videos, blogs, Internet searches, social media postings, tweets, and so forth—can be analyzed to better understand and predict human behavior. Businesses can analyze customer preferences, as demonstrated by a big data analytics company applying its algorithms to millions of Facebook pages and tweets to assess attitudes toward nine aspects of their experience with a clothing store, from the quality and style of the product to the shopping experience and returns policies.

Big data is making possible advances in nearly all areas of science and technology. In combination with computational capabilities and genomic data, big data is leading to advances in bioengineering. Moreover, big data can be used to identify correlations of chronic diseases with other factors such as the environment, geography, dynamics, and genomics to identify points of health intervention. The cloud and big data are enabling collaboration on a global scale to both interpret data and find solutions.

### Information Ecosystem

“Big data” is not just the data itself but rather an ecosystem from sensors to storage to computational analytics to human use of the information and analyses. Big data has been enabled by the exponential increase in cloud storage and a three-million-fold decrease in storage costs since 1980,<sup>12</sup> while the cost of computational power and software has also plunged dramatically over the past thirty years. Today, big data is managed by literally millions of servers in massive “server farms” operated by Google, Amazon, Microsoft, Yahoo, and other high-tech companies, with software algorithms that, for example, enable instant Google searches, Bing predictive airfare analytics, and Amazon personal book suggestions based on analysis of past purchases.

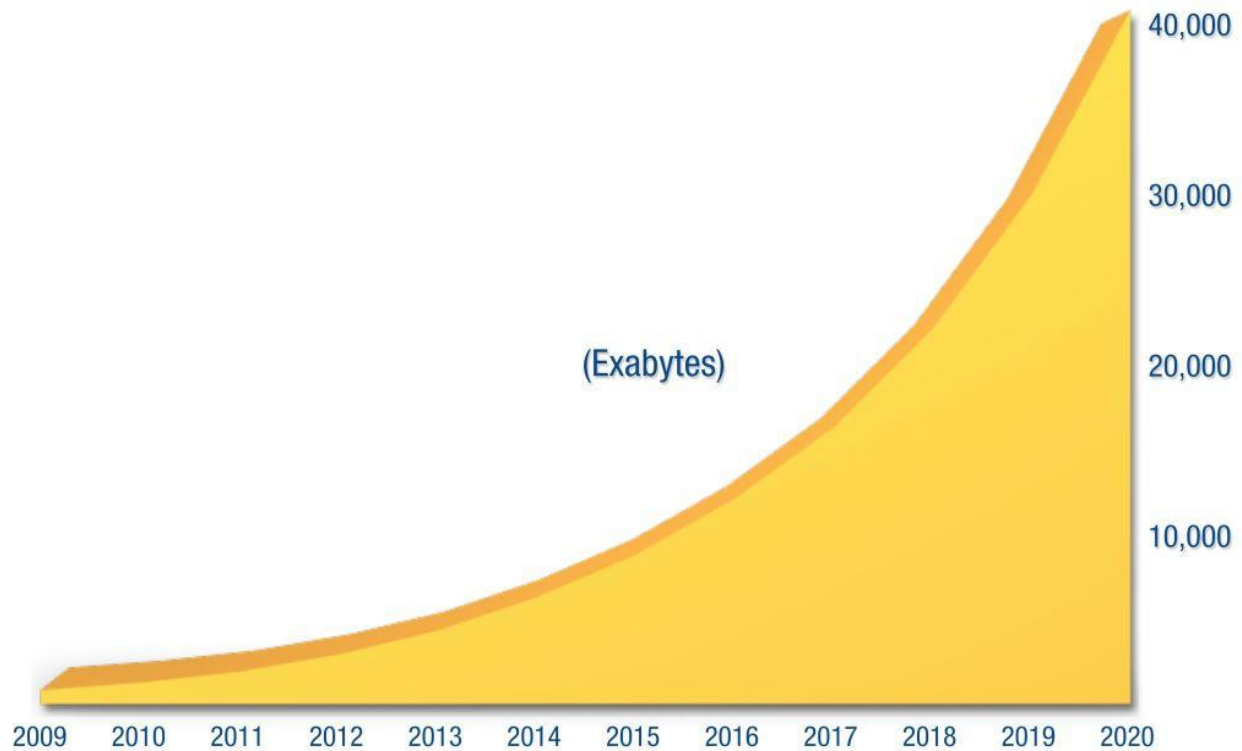
In the “wired world,” there is an increasingly dense and ubiquitous networking, not only of people, but also of things. Big data draws on and helps make useful the exponentially growing “Internet of Things” (IoT), which is linking machines to machines, ranging

10 John Gantz and David Reinsel, “The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East,” *IDC Iview*, December 2012, <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.

11 Ibid.

12 M. Komorowski, “A History of Storage Cost,” 2009, <http://www.mkomo.com/cost-per-gigabyte>.

# The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



This IDC graph predicts exponential growth of data from around 3 zettabytes in 2013 to approximately 40 zettabytes by 2020. An exabyte equals 1,000,000,000,000,000,000 bytes and 1,000 exabytes equals one zettabyte. Source: IDC's Digital Universe Study, December 2012, <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>.

from smartphones, tablets, and laptops, to intelligent sensors, autonomous vehicles, and home heating and air conditioning systems. There are now twice as many “things” connected to each other through the Internet—some 15 billion—as there are people on the planet. Cisco predicts that there will be 25 billion Internet-connected devices by 2015 and 50 billion by 2020.<sup>13</sup>

HP is even more expansive in its predictions, forecasting that there will be trillions of sensors providing data by the 2020s.<sup>14</sup> This will enable big data to take the pulse of the planet in real time

with a continuous flow of information on sea levels, agricultural production, impact of climate change, environmental degradation, energy use, food production and distribution, and water availability—all simultaneously at global and granular levels. This could enable global and local monitoring, pinpointing challenges and threats and mobilizing governments and NGOs to respond in more or less real time. The Internet of Things will also be vulnerable to potentially catastrophic hacking.<sup>15</sup>

## Boosting Productivity

Sensors will enable big data to monitor individual health as well. Individuals will be able to swallow sensors that cost a few pennies each to provide daily updates on their health as the data is transmitted and

13 Dave Evans, “The Internet of Things: How the Next Evolution of the Internet Is Changing Everything,” CISCO Internet Business Solutions Group, April 2011, [http://www.cisco.com/web/about/ac79/docs/innov/IoT\\_IBSG\\_0411FINAL.pdf](http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf).

14 “The Internet of Things with Trillions of Sensors Will Change Our Future,” *Big Data Startup*, March 20, 2013, <http://www.bigdata-startups.com/internet-of-things-with-trillions-of-sensors-will-change-our-future/>.

15 See Peter Haynes and Thomas A. Campbell, “Hacking the Internet of Everything,” *Scientific American*, August 1, 2013, <http://www.scientificamerican.com/article.cfm?id=hacking-internet-of-everything>.



*Flickr user Eric Fischer used geotagging data to map popular tourist and local hotspots around San Francisco.*

analyzed in the cloud and the results returned to the individual or their physician on their smartphone. There are also estimates that use of big data could reduce healthcare errors, including a 30 to 40 percent improvement in diagnoses, and lead to a potential 30 percent reduction in healthcare costs. Agriculture is another area that is being enhanced by sensors and big data. McKinsey notes that “precision farming equipment with wireless links to data collected from remote satellites and ground sensors can take into account crop conditions and adjust the way each individual part of a field is farmed—for instance, by spreading extra fertilizer on areas that need more nutrients.”<sup>16</sup>

Robots increasingly will be sources of big data, some of which will be mined for improving the performance of those machines, from warehouse loaders to robotized surgery to elder care to self-driving cars and aerial drones. Moreover, big data will play a key role in energy development. It has already had an immense impact on the revolution in oil and gas drilling over the past five years, which, by one estimate, will enable a 200 to 300 percent improvement in the productivity of the typical oil or gas rig in US shale fields.<sup>17</sup>

But this is just the dawn of the big data era. In this next phase of the Internet, the IoT, the sheer scale of data washing over the world to be stored, shared, and

<sup>16</sup> Michael Chui, Markus Löffler, and Roger Roberts, “The Internet of Things,” *The McKinsey Quarterly*, March 2010, [https://www.mckinseyquarterly.com/High\\_Tech/Hardware/The\\_Internet\\_of\\_Things\\_2538](https://www.mckinseyquarterly.com/High_Tech/Hardware/The_Internet_of_Things_2538).

<sup>17</sup> Mark P. Mills, “Big Data And Microseismic Imaging Will Accelerate The Smart Drilling Oil And Gas Revolution,” *Forbes*, May 8, 2013, <http://www.forbes.com/sites/markpmills/2013/05/08/big-data-and-microseismic-imaging-will-accelerate-the-smart-drilling-oil-and-gas-revolution/>.

analyzed is virtually incomprehensible. Technology is enabling human activity and nature to generate massive and increasing amounts of digital data. One estimate is that nearly three zettabytes ( $3 \times 10^{21}$  bytes, or three billion terabytes) of data was created in 2012 alone and that the volume is growing by 50 percent per year. From now until 2020, the digital data produced will roughly double every two years.<sup>18</sup>

The number of human experts to design the systems and to make sense of the analytical results will have to grow dramatically as well. In the United States alone, it is estimated that by 2018 there will be a shortage of 140,000 to 190,000 people with deep analytical skills, as well as 1.5 million managers and analysts with the ability to leverage big data for effective decision making. In the quantum field, fewer than one in 10,000 scientists, and even fewer engineers, have the education and training necessary to exploit quantum tools, even when they are enabled by a quantum machine.

Big data can be stored and analyzed later or analyzed in real time whether or not the data is saved. That data can be about human behavior and choices (web searches, personal location, tweets, purchases, phone calls, emails, and credit card purchases), machine activity (robots, transportation vehicles, manufacturing equipment, and satellites), behavior of human artifacts (structural integrity of bridges and buildings), and natural activity (temperature increases, water pollution, ocean acidification, carbon dioxide accumulation, or animal location and health). Only a tiny fraction of the digital universe has been explored for analytic value, according to IDC, which estimates that by 2020 as much as 33 percent of the digital universe will contain information that might be valuable if analyzed.<sup>19</sup> By 2020, nearly 40 percent of the information in the digital universe—

---

18 From 2005 to 2020, the digital universe will grow by a factor of 300, from 130 exabytes to 40,000 exabytes, or 40 trillion gigabytes (more than 5,200 gigabytes for every man, woman, and child in 2020). John Gantz and David Reinsel, "The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East," *IDC Iview*, December 2012, <http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>. Gantz and Reinsel also note: "Much of the digital universe is transient—phone calls that are not recorded, digital TV images that are watched (or "consumed") that are not saved, packets temporarily stored in routers, digital surveillance images purged from memory when new images come in, and so on. Unused storage bits installed throughout the digital universe will grow by a factor of eight between 2012 and 2020, but will still be less than a quarter of the total digital universe in 2020."

19 John Gantz and David Reinsel, *op.cit.*

approximately sixteen zettabytes or more than five times the current total data flow—will be stored or processed in a cloud somewhere in its journey from originator to disposal.<sup>20</sup>

Big data is often, though not always, generated at high velocity and includes an increasingly wide range of data as more and more human and natural activities are "datified"—that is, turned into data that can be collected and analyzed. Finally, big data is only meaningful if it can now (or in the future) be analyzed and produce value from the insights and knowledge gained from application of advanced machine learning and analytics to the datasets.<sup>21</sup>

Some of our personal as well as societal digital exhaust could last longer than carbon dioxide in the atmosphere, which is expected to be there for a century or longer. While much of the data produced everyday may be immediately assessed as detritus and discarded, a large amount of the data each of us produces every day is gathered and could be stored indefinitely, to be mined and recalled whenever those with access to the archive become interested. Those Facebook posts, tweets, and search engine queries—as well as phone calls, GPS records of our travel uploaded by our smartphones, and every credit card purchase—will be stored and available. Such personal data is way overshadowed, of course, by all the other data sources contributing to the massive daily accumulation of data.

And as we learned in June 2013, some of that data is being siphoned and stored, along with metadata on phone calls, by the NSA. Much of the data and metadata will presumably be kept at the NSA's new Data Center in Bluffdale, Utah, which will be able to store and mine 1,000 zettabytes or 1 yottabyte (1 septillion or  $10^{24}$  bytes) of data. To get an idea of the immensity of this challenge, to store a yottabyte on terabyte-sized hard drives would require 1 million city-block-sized data centers, equivalent to covering the states of Delaware and Rhode Island.<sup>22</sup>

---

20 *Ibid.*

21 For a discussion of limitations of Big Data, see Christopher Mims, "Most data isn't 'big,' and businesses are wasting money pretending it is," *Quartz*, May 16, 2013, <http://qz.com/81661/most-data-isnt-big-and-businesses-are-wasting-money-pretending-it-is/>.

22 See Wikipedia, "Yottabyte," <http://en.wikipedia.org/wiki/Yottabyte>.

According to veteran NSA-watcher James Bamford, who was first to describe the massive Bluffdale facility, its purpose is “to intercept, decipher, analyze, and store vast swaths of the world’s communications as they zap down from satellites and zip through the underground and undersea cables of international, foreign, and domestic networks.”<sup>23</sup> Bamford claimed that Bluffdale will store “all forms of communication, including the complete contents of private emails, cell phone calls, and Google searches, as well as all sorts of personal data trails—parking receipts, travel itineraries, bookstore purchases, and other digital ‘pocket litter.’” In addition, according to a former official cited by Bamford who was previously involved in the program, code-breaking will be a crucial function of the NSA facility since “much of the data that the center will handle—financial information, stock transactions, business deals, foreign military and diplomatic secrets, legal documents, confidential personal communications—will be heavily encrypted.”

### Value of Huge Data Sets

Big data analytics are the key to making use of data and to data mining for correlations on human behavior to gain market intelligence. The more the data, the easier it is to solve the problem, from traffic congestion to identifying the best airline seats to the empowering of individuals. With a combination of networks plus data, human behavior can be identified, modeled, influenced, and mined for potential economic value. The use of big data will become a key basis of competition and growth for the private sector, as it has been for Amazon, Google, and others, thus impacting overall economic growth. McKinsey has estimated that the use of data analytics will enable retailers to increase their operating margins by more than 60 percent and that services using personal location data will generate an estimated \$600 billion in additional global consumer spending.<sup>24</sup>

Big data also often holds the promise of analyzing complete data sets rather than a sampling of data.

---

23 See James Bamford, “The NSA Is Building the Country’s Biggest Spy Center (Watch What You Say),” *Wired*, March 15, 2012, [http://www.wired.com/threatlevel/2012/03/ff\\_nsadatacenter/all/](http://www.wired.com/threatlevel/2012/03/ff_nsadatacenter/all/).

24 James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers, “Big data: The Next Frontier for Innovation, Competition and Productivity,” McKinsey Global Institute, March 2011, [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation).

Google has experimented with a Google Price Index (GPI) that would determine the average price for all sales of a given item on the Internet in contrast with the US government’s Consumer Price Index (CPI), which is based on taking a periodic and small sample of the same items to extrapolate an average price.<sup>25</sup> And unlike the CPI, the GPI is analyzed and presented in real time, not analyzed and presented days or weeks after the actual sampling.

To keep up with and analyze this exponential flow of data will require further massive growth in storage capacity and computational power, along with rapidly increasing development of algorithms and software. Big data analytics require new techniques and algorithms that are yet to be developed, requiring new research and development. They will also, in time, benefit from the near-real-time analytical abilities of quantum computing. For example, scientists at Microsoft have calculated that a system of  $10^{21}$  linear equations that would take 31,000 years to solve on a PC would take a few seconds to calculate on a fairly basic quantum computer.

While the portion of the digital universe that holds potential analytic value, including economic value, is growing—from about 25 percent today to about one third of an exponentially larger universe in 2020—only a very small portion has so far been explored, according to IDC. “This untapped value could be found in patterns in social media usage, correlations in scientific data from discrete studies, medical information intersected with sociological data, faces in security footage, and so on.” IDC noted, however, that the amount of information in the digital universe “tagged” for analysis in 2012 accounted for only about 3 percent of the big data gathered, and only about one half of one percent was actually analyzed. Thus, there are large untapped pools of data in the digital universe from which big data technology can potentially extract value.<sup>26</sup>

Protection of data from theft, misuse, and loss is and will continue to be a major problem in the digital world.

---

25 Briefing of the author by Google Chief Economist at Google, November 2010. See also Jacob Goldstein, “The Google Price Index,” *NPR*, October 12, 2010, <http://www.npr.org/blogs/money/2010/10/12/130508576/the-google-price-index>. Google apparently decided to not to make the GPI public.

26 John Gantz and David Reinsel, op.cit.



The proportion of data that requires protection is growing faster than the digital universe itself, from less than a third in 2010 to more than 40 percent in 2020, but only about half the information that needs protection has protection. That may improve slightly by 2020, as some of the better-secured information categories will grow faster than the digital universe itself, but it still means that the amount of unprotected data will grow by a factor of twenty-six. Emerging markets have even less protection than mature markets.<sup>27</sup> IDC predicts that by 2020, 40 percent of the forty zettabytes of data—up from one-third of data collected in 2010—will need to be securely stored. Security is critical to big data, and data protection regulations and practices must be at the forefront of digital agendas to ensure a trustworthy foundation for the data-driven economy.

Big data can help governments improve policymaking in different sectors—including city planning, traffic management, efficiency improvements in use of energy and other resources, greening and managing the built environment, infrastructure maintenance monitoring, epidemic tracking, disaster preparedness and response, healthcare improvement and cost control, economic forecasting, and resource allocation. Anticipated uses of big data include population movement patterns in the aftermath of a disaster or a disease outbreak, early warning upon detection of anomalies, real-time awareness of a population's needs, and real-time feedback on where development programs are not delivering anticipated results. Public access to big data and big data analytics will also put pressure on governments for transparency and responsiveness. In addition, governance can be improved through crowdsourcing identification of problems and development of solutions. This will take some burdens off government by shifting them to the population and will help government identify innovative solutions to intractable problems.

For individuals, big data holds the potential of a huge new universe of useful information to enhance the quality of life, health care, and improved services of all kinds, including government services. But individuals will also be potentially haunted by their “digital

exhaust” that will in some instances last longer than their lifetimes. Grandchildren and great grandchildren and their offspring could have access to an immense amount of data on their ancestors. The amount of information individuals create themselves, however, is far less than the amount of information being created about them in the digital universe.<sup>28</sup> Individuals will have even less control over this information, and will be ever more subject to its misuse, from identity theft to misinterpretations of past behavior.

### **Critical Roles for Government**

Realizing the potential of big data for economic growth will require policies that enable data to flow and be exchanged freely across geopolitical boundaries, while at the same time minimizing risks to companies and individuals globally. This will be difficult and divisive and will probably be a continually-moving target as technology changes. It will require agreement nationally and internationally on such issues as what data can be collected and shared, who owns the data, who can access the data, how it may be used, and how disputes among nations over these issues should be resolved. It thus raises huge issues of data sovereignty, none of which are close to being resolved. Lack of sufficiently restrictive regulations can expose governments, corporations, and individuals to unacceptable risks, while regulations that restrict data collection, flow, and use can impede innovation in technology, services, business models, and scientific researching, thus slowing growth as well as harming the public good.

While there are many issues regarding limiting what data is collected and how it is used, the primary focus of government regulation should be on controlling access to and use of big data, including not only its immediate use but possible use in the future, rather than trying to restrict collection. Restricting is largely a futile effort since so much data is already being collected passively or by “consent” when users “agree” to terms of use agreements without ever actually reading the lengthy legal documents. Many sets of big data may be stored and unused at present for perceived lack of utility or current lack of software and computing power to tap its value, but they could be

---

<sup>27</sup> Ibid.

---

<sup>28</sup> Ibid.

socially or scientifically useful in the future. Moreover, the NSA is storing massive amounts of encrypted data that it currently cannot access, but the Agency expects improved computing power to eventually enable it to go back to that data and crack the codes. Quantum computers, when they become available to the NSA, will enable virtually immediate decryption.

Government needs to promote better understanding of big data and its implications for society, the economy, research and development, and governance. Both the benefits and the risks—and the importance of balancing these—must be the subject of research and broad discussions inside and outside the government. Government in particular needs to “get ahead of the curve” and not find that it is blindsided by the exponential expansion, not just of big data (and open data), but of its role throughout society and government itself. How can the implications be foreseen, the opportunities grasped, and the risks hedged against?<sup>29</sup> Answering these questions will necessarily involve representatives of all the stakeholders, including service providers, businesses, regulators, researchers, big data producers and providers, civil society, law enforcement, and the foreign policy and national security establishments.

Government itself is a prolific creator, disseminator, and repository of big data. It generates and stores immense amounts of information, both digital and physical, that should be examined and used to improve government performance, particularly in controlling epidemics, managing disasters and disaster preparedness, and improving energy efficiency throughout government and society. Government will find a potentially huge source of data and problem solving by soliciting citizen identification of problems. But government will also be challenged by demands for

transparency of data on all aspects of its performance and on issues considered relevant by citizens. The Chinese leadership found itself challenged by the US embassy in Beijing making available through the Internet daily readings on pollution, which led the Chinese government, under mounting public pressure, to monitor and publicly release pollution levels.

A confluence of technology and socioeconomic factors has fueled an exponential growth in the volume, variety, velocity, and availability of digital data, creating an emerging and vibrant data-driven economy. But this nascent economy needs an appropriate policy framework to enable its full socio-economic potential while protecting privacy, individual rights, and data security. It also will need a sustained increase in investment in data storage, computational power, software and algorithm development, and human capital to design and run the big data ecosystem and interpret and act on the results of big data analytics.

*AUGUST 2013*

---

<sup>29</sup> Big data is increasing the catastrophic risks for the financial system that is already heightened by “high frequency trading.” See Maureen O’Hara and David Easley, “The Next Big Crash Could Be Caused By Big Data,” *Financial Times*, May 21, 2013, <http://www.ft.com/intl/cms/s/0/48a278b2-c13a-11e2-9767-00144feab7de.html>: “Financial markets are significant producers of big data: trades, quotes, earnings statements, consumer research reports, officials statistical releases, polls, news articles, etc.” O’Hara and Easley note that the so-called “hash crash” of April 23, 2013 was a “big data crash” resulting from trading algorithms “based on millions of messages posted by users of Twitter, Facebook, chat rooms, and blogs” that responded to a bogus tweet about a terrorist attack on the White House.

# Atlantic Council Board of Directors

---

## INTERIM CHAIRMAN

\*Brent Scowcroft

## PRESIDENT AND CEO

\*Frederick Kempe

## VICE CHAIRS

\*Robert J. Abernethy

\*Richard Edelman

\*C. Boyden Gray

\*Richard L. Lawson

\*Virginia A. Mulberger

\*W. DeVier Pierson

\*John Studzinski

## TREASURER

\*Brian C. McK. Henderson

## SECRETARY

\*Walter B. Slocombe

## DIRECTORS

Stephane Abrial

Odeh Aburdene

Timothy D. Adams

\*Michael Ansari

Richard L. Armitage

\*Adrienne Arsht

\*David D. Aufhauser

Elizabeth F. Bagley

Ralph Bahna

Sheila Bair

Lisa B. Barry

\*Rafic Bizri

\*Thomas L. Blair

Julia Chang Bloch

Francis Bouchard

R. Nicholas Burns

\*Richard R. Burt

Michael Calvey

James E. Cartwright

Daniel W. Christman

Wesley K. Clark

John Craddock

David W. Craig

Tom Craren

\*Ralph D. Crosby, Jr.

Thomas M. Culligan

Gregory R. Dahlberg

\*Paula J. Dobriansky

Christopher J. Dodd

Markus Dohle

Lacey Neuhaus Dorn

Conrado Dornier

Patrick J. Durkin

Thomas J. Edelman

Thomas J. Egan, Jr.

Stuart E. Eizenstat

Julie Finley

Lawrence P. Fisher, II

Alan H. Fleischmann

Michèle Flournoy

\*Ronald M. Freeman

\*Robert S. Gelbard

Richard L. Gelfond

Edmund P. Giambastiani, Jr.

\*Sherri W. Goodman

John A. Gordon

\*Stephen J. Hadley

Mikael Hagström

Ian Hague

Frank Haun

Rita E. Hauser

Michael V. Hayden

Annette Heuser

Marten H.A. van Heuven

Marillyn Hewson

Jonas Hjelm

\*Mary L. Howell

Robert E. Hunter

Robert L. Hutchings

Wolfgang Ischinger

Deborah James

Robert Jeffrey

\*James L. Jones, Jr.

George A. Joulwan

Stephen R. Kappes

Francis J. Kelly, Jr.

Zalmay M. Khalilzad

Robert M. Kimmitt

Roger Kirk

Henry A. Kissinger

Franklin D. Kramer

Philip Lader

David Levy

Henrik Liljégren

\*Jan M. Lodal

\*George Lund

\*John D. Macomber

Izzat Majeed

Fouad Makhzoumi

Wendy W. Makins

Mian M. Mansha

William E. Mayer

Eric D.K. Melby

Franklin C. Miller

\*Judith A. Miller

\*Alexander V. Mirtchev

Obie L. Moore

\*George E. Moose

Georgette Mosbacher

Bruce Mosler

Sean O'Keefe

Hilda Ochoa-Brillembourg

Philip A. Odeen

Ahmet Oren

Ana Palacio

\*Thomas R. Pickering

\*Andrew Prozes

Arnold L. Punaro

Kirk A. Radke

Joseph W. Ralston

Teresa M. Ressel

Jeffrey A. Rosen

Charles O. Rossotti

Stanley O. Roth

Michael L. Ryan

Harry Sachinis

William O. Schmieder

John P. Schmitz

Kiron K. Skinner

Anne-Marie Slaughter

Alan J. Spence

John M. Spratt, Jr.

Richard J.A. Steele

James B. Steinberg

\*Paula Stern

William H. Taft, IV

John S. Tanner

Peter J. Tanous

\*Ellen O. Tauscher

Clyde C. Tuggle

Paul Twomey

Henry G. Ulrich, III

Enzo Viscusi

Charles F. Wald

Jay Walker

Michael F. Walsh

Mark R. Warner

J. Robinson West

John C. Whitehead

David A. Wilson

Maciej Witucki

R. James Woolsey

Mary C. Yates

Dov S. Zakheim

## HONORARY DIRECTORS

David C. Acheson

Madeleine K. Albright

James A. Baker, III

Harold Brown

Frank C. Carlucci, III

Robert M. Gates

Michael G. Mullen

William J. Perry

Colin L. Powell

Condoleezza Rice

Edward L. Rowny

James R. Schlesinger

George P. Shultz

John W. Warner

William H. Webster

## LIFETIME DIRECTORS

Carol C. Adelman

Lucy Wilson Benson

Daniel J. Callahan, III

Kenneth W. Dam

Stanley Ebner

Barbara Hackman Franklin

Chas W. Freeman

Carlton W. Fulford, Jr.

Geraldine S. Kunstadter

James P. McCarthy

Jack N. Merritt

William Y. Smith

Marjorie Scardino

Ronald P. Verdicchio

Carl E. Vuono

Togo D. West, Jr.

---

*\*Members of the Executive Committee  
List as of April 24, 2013*

The Atlantic Council is a nonpartisan organization that promotes constructive US leadership and engagement in international affairs based on the central role of the Atlantic community in meeting today's global challenges.

© 2013 The Atlantic Council of the United States. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means without permission in writing from the Atlantic Council, except in the case of brief quotations in news articles, critical articles, or reviews. Please direct inquiries to:

**1030 15th Street, NW, Washington, DC 20005**  
**(202) 463-7226, [AtlanticCouncil.org](http://AtlanticCouncil.org)**