BIG DATA

Too Big to Ignore for Latin America and the Caribbean

Dr. Boris Saavedra
edited by Kathleen Vaughan and Liliana Besosa

August 2017

William J. Perry
Center for Hemispheric Defense Studies

Viviana Edwards
Multimedia Specialist
Cover and Book Layout

Big Data:
Too Big to Ignore for Latin America and the Caribbean

Dr. Boris Saavedra
edited by Kathleen Vaughan and Liliana Besosa

INTRODUCTION:

Data is among the most valuable commodities of the 21st Century and Cyber Security is its guardian. Any nation hoping to successfully compete in the global ecosystem must be able to access, collect, process, analyze, make sense of and secure Big Data. Nations with a deep understanding of how to make sense of data to inform strategic decision making will have a distinct advantage in the decades to come. Data-informed decision making will be enabled by advanced analytics, augmented by artificial intelligence all within a comprehensive nation state competitive frameworks.

This paper addresses Big Data and Cyber Security from strategic, public, and private perspectives with a focus on Latin America and the Caribbean.

In this paper we will attempt the following:

Explain how current methods of data storage have changed from the first computers to the emergence of the NoSQL databases linked to big data analytics.[1]

- Examine how data centers function.
- Discuss how governments and major companies utilize big data.
- Explain what the Cloud is and what it enables.
- Evaluate and explain Hadoop and Mongo databases; two most popular technologies for analyzing big data.
- Consider the challenges and opportunities associated with regional governmental policies and strategic solutions.
- Analyze the impact of big data surveillance on individuals' daily lives, communities, nations and society in terms of security and defense.
- Analyze the rapidly changing future of big data analytics.

1 Rafael Caballero and Enrique Martin, Las Bases de Big Data (Madrid: Los Libros de la Catarata, 2015), 7-9.

Evidence of our highly technological world and its rapid evolution lies in the palm of your hand in the form of a cell phone. This small, sleek, and incredibly powerful tool has become central to both our professional and personal lives. Our phones are constantly generating data. Every digital application, sensor, GPS device, credit card transaction, and social media exchange produces data. Our phones are no longer limited to conveying messages and conversations but instead allowing human beings to produce vast sums of information about their actions They are constantly emitting data that reveal our location, our interests, our friends, and our preferences.

Every time we surf the web, we produce huge amounts of information that is used to improve and adapt webpages to our wants and needs. This information also allows for targeted advertising, of either beneficial or malicious intent. When we purchase public transportation tickets, new information is added to a huge database that helps managers decide how to resource transportation lines based on demand. When we make a credit card purchase, we send information to our banks about spending habits. If one prefers cash transactions, it still generates new data. The recorded sales are aggregated to indicate trends that influence restocking decisions. The data generation cycle never sleeps. In one day in 2014, 204,166,667 e-mails were sent worldwide, Google's search engine was queried 2 million times, 684,000 pieces of content were shared on Facebook, 100,000 tweets were posted on Twitter, 47,000 apps were downloaded from the Apple App Store, 48,000 hours of new video was uploaded on YouTube, 36,000 new photographs were posted on Instagram, and 34 million WhatsApp messages were sent. In other words, every ten minutes, modern technology generates the same amount of information as was as was created in the first ten thousand generations of human history.[2]

However, there are numerous risks associated with this growing sea of data. Not only do we face an onslaught of data mining from Internet companies, marketers, and third-party data brokers, but criminals, terrorists, and rogue governments are exploiting this information as well. The data trails are now growing exponentially longer due to the computers we carry with us at every turn—our mobile phones.

Cybersecurity has become a major concern around the world. In the last decade, it has deeply affected Latin American nations. The most notable conclusion emerging from the 2016 Cybersecurity Report for Latin America and the Caribbean (LAC), is that the majority of the region's countries are not yet prepared to counter cybercrime. This finding

2 Marc Goodman, Future Crimes. Everything is connected, everyone is vulnerable and what we can do about it (New York: Doubleday, a division of Random House, 2015) 84-85.

is a call for action. Latin American and Caribbean nations must start taking the necessary steps to protect key regional 21st century databases and infrastructure.[3]

According to the 2016 report, estimates place the cost of cybercrime worldwide at US$575 billion a year, which represents 0.5% of the global GDP. That is almost four times the sum of international donations for development annually. In Latin America and the Caribbean, the report estimates a yearly cost of around US$90 billion from cybercrime.[4] This is the money that private and public entities are paying to fix their vulnerabilities. Since region falls short in the prevention and mitigation of criminal or malicious activity in cyberspace; this money is lost that could have gone to increasing regional scientific research fourfold. The report proposes cybersecurity capabilities based on cooperation by all countries under a common regional policy and framework.

Evolution of Automatic Data Processing

The first commercial computer UNIVAC I, was built in 1952 for the U.S. Census Bureau. One of its biggest successes was accurately predicting Eisenhower's victory in the 1952 presidential elections. The most notable consequence was that the general population became aware of the possibilities offered by a machine capable of storing 12 Kb of data.[5] Today, 50,000 UNIVACs would be required to match the internal memory capability of a modest cell phone.

According to Whitchorn, in the field of information technology, "Big Data" is any data that does not fit well into tables, and that generally responds poorly to manipulation by SQL language[6]. Big data is characterized by three features, also known as the 3-Vs: Volume, Velocity and Variety. The first refers to the large amount or volume of data stored. The second refers to the speed at which data arrives from the web without being blocked, and the third refers to the diversity of information that automatically enters the system and is sorted for eventual use.

Although the capacity of databases has expanded very quickly, there are basic principles

3 Organization of American States & Inter-American Development Bank. Observatory Cybersecurity in Latin America and the Caribbean. Annual Report. 2016

4 Patricia Pandrini and Marcia L. Maggiore, Panorama Del Ciberdelito En Latinoamérica. Working paper. (Montevideo: Latin America and Caribbean Network Information Centre, 2011.)

5 Bruce Schneier, Data and Goliath. The Hidden Battles to Collect Your Data and Control Your World. (New York: W.W. Norton & Company, Inc., 2016) According to Scneier: "A note on storage units. We should remember that information is usually stored in a byte. A Kilobyte kB equals 1,000 bytes or characters, one megabyte (MB) is 1000 kB, one gigabyte (GB) is 1000 MB and a terabyte (TB) is 1000 GB. A petabyte (PB) is 1000 TB. A petabyte is a one followed by 15 zeros. For a better understanding of this capacity, a book of 700 pages typically take an average of 500 kB. However, there are names for bigger numbers. A thousand petabytes is an exabyte (a billion, billion bytes). A thousand Exabyte is a zettabyte, and a thousand zettabytes is a yottabyte. To put it in human terms, an Exabyte of data is 500 billion pages for text."

6 Mark Whitchorn, Big Data Bites Back: How to handle those Unwieldy Digits when You can't Just Cram It into Tables, August 27, 2012, www.theregister.co.uk/2012/08/27/how_did_big_data_get_so_big/

that have remained constant. Decisions made at the time of design and storage format have enormous repercussions on cost, speed, efficiency, consultation mechanism, security, and maintenance. The concerns about information integrity that have existed since the 1950s continues in today's systems such as "the cloud."

The "cloud" enables users to rent data storage space, website management, and data analysis; called cloud computing. The general idea is to rent computers that are housed in data remote processing centers through the cloud using programs on our own computers. Thus, cloud computing provides the resources we need to process data without large personal investments in infrastructure. The result of the great success of this system are the emergence of companies like Google, Facebook, and Amazon which are the main providers of the cloud for public and private sectors. For users at home, cloud computers are not physically accessible, but rather virtually (via the internet). The space, resources, and hardware of cloud computers (hard-disks etc.) are shared with other cloud customers.

There are two different types of databases in use today: the relational Structural Query Language (SQL) and non-relational Structural Query Language (NoSQL) databases. These are languages that allow you to communicate and manage databases. Until 2013, 90% of the world's databases were relational databases. By 2015, that percentage fell to 80%. According to experts, this downward trend will continue in the coming years as more models enter the database market.

The study of cyber issues did not appear in Latin America and the Caribbean until the late 1980s and early 1990s. The subject was not given priority until 1999 when the Organization of American States (OAS) established its first transnational cybercrime alliance. The objectives were to increase cooperation among its members, intensify technical and legal efforts, advise on the possible enactment of a cybercrime agreement, and apply legislation aimed at combating this kind of crime. The result of the OAS initiative was that several countries began to realize the magnitude of cyber threats. Consequently, Argentina, Chile, Colombia, Costa Rica, the Dominican Republic, Mexico, Panama, Paraguay, and Peru joined the Council of Europe's Convention on Cybercrime as Non-Members.[7]

The Data Broker Industry

Tracking customer data is much older than the Internet. Before the internet, there were four basic data gathering lines of effort. This first involved companies keeping records of

7 Its main objective is to pursue a common criminal policy aimed at the protection of society against cybercrime, by adopting appropriate legislation and fostering international co-operation.

their frequent customers' preferences, such as airlines, hotels, rental cars, etc. Eventually, this evolved into databases that enabled the companies to track their sales. The second data tracking line was direct marketing through physical mail. The objective was to provide companies with lists of people most likely to be interested in the products being advertised. The third line came from credit bureaus. These companies collected detailed credit information about people and sold that information to banks trying to determine whether to lend to individuals and the optimal rates for such transactions. The fourth line was done by the government through various public records: birth and death certificates, driver's license records, voter registration records, various permits and licenses, and so on. The depth of information that data brokers have collected is astonishing. Data brokers sort this information into profiles that are divided into various marketing categories. For example, the data broker Equifax sold lists of people who were late on their mortgage payments to a discount loan company. They were fined by the Federal Trade Commission for their action.[8]

According to the OAS & IDB 2016 report, in Latin America and the Caribbean two out of three countries do not have command centers or cybersecurity control operating under national legislation. Moreover, most prosecutors lack the legal tools to prosecute cybercrimes. Instead of having national regulations for cybersecurity; companies in the private sector employ technical recommendations from software providers to maximize the efficiency and profits of their business.[9]

Government Database Monitoring and Control

Government monitoring cannot be democratic if it is not transparent and does not follow the rule of law.[10] In strong democratic states, surveillance is difficult to do because of regulations about transparency. The democratic state's ability to conduct surveillance is limited because rule of law is robust politically, legally, and technically. Additionally, the public-private surveillance partnership means governments don't conduct surveillance, censorship, and control operations alone. They are supported by a vast public-private surveillance partnership (PPSP) an array of for-profit cooperation. For example, in 2013, the Washington Post reported that 70% of the U.S. intelligence budget is spent on private firms and that 483,000 government contractors hold top-secret clearances, a third of the 1.4

8 Federal Trade Commission. FTC Settlements Require Equifax to Forfeit Money Made by Allegedly Improperly Selling Information about Millions of Consumers Who were Late on their Mortgages. 2012

9 Organization of American States & Inter-American Development Bank. Observatory Cybersecurity in Latin America and the Caribbean. Annual Report 2016

10 Weber, Max. "Politics as a Vocation". http://anthropos-lab.net/wp/wp-content/uploads/2011/12/Weber-Politics-as-a-Vocation.pdf

million people cleared at that level.[11] On the other hand, democracy is relatively weak in Latin America so information about surveillance is not always available to the public due to a lack of laws about transparency.

All over the world, more intense attention was given to eavesdropping after the terrorist attacks of September 11, 2001. This also resulted in significant advancement in big database technology. The logic was that the only way to have a chance of preventing something from happening was to know everything that is happening. Terrorism today is networked, with independent terrorist cells and individuals who can operate anywhere. Modern government surveillance monitors everyone – domestic and international alike – to attempt to detect these cells and disrupt their plans.

Unfortunately, government and private sector response to cybersecurity challenges is often shockingly inept. However, identifying what is at fault triggers defensive reactions, finger pointing, and endless disputes in policy circles. The underlying policy challenge, according to Dave Oxner of the Security Industry and Financial Markets Association, is getting the government, private sector, and the public to all realize that cyber risk can only be mitigated but not eliminated. Oxner contends that "we need to change the mindset that this can be solved."[12]

Cybersecurity as a policy issue also frequently aligns with the unsettled debate over the right level of government anti-terror surveillance. This makes the development of national policy on cybersecurity a troubled topic through issues such as privacy and civil liberties.

Time is working against the political bureaucracy and stakeholders working on cybersecurity in the face of such a monumental challenge. In most cases, there is a negligent lack of laws governing cybersecurity. As a result, cybersecurity is mostly on a voluntary basis because political efforts to establish a mandatory standard of cybersecurity control for industry have failed. Even hard-core opponents of command-and-control government regulation concede that more must be done, particularly in the private sector.

According to the 2016 OAS & IDB report, in the Latin American and the Caribbean region several countries are vulnerable to potentially devastating cyberattacks, based on 49 indicators. Four out of five countries do not have cybersecurity strategies, critical infrastructure, or database protection plans. Two out of three have no cybersecurity command and control centers.[13]

---

11 Bruce Schneier, Data and Goliath. The Hidden Battles to Collect Your Data and Control Your World. (New York: W.W. Norton & Company, Inc., 2016).
12 Charlie Mitchell, *Hacked. The Inside Story of America's Struggle to Secure Cyberspace* (Rowman & Littlefield Publishers, 2016).
13 Organization of American States & Inter-American Development Bank. Observatory Cybersecurity in Latin America and the Caribbean. Annual Report. 2016 file:///C:/Saavedrab/Downloads/Cybersecurity-Are-We-Prepared-in-Latin-America-and-Caribbean.pdf

Big Data Physical Locations

Policy makers, strategists and CEOs that use major big data systems need to know some basic information about the function and locations of these databases. They are typically found in remote places and often go unnoticed within industrial building complexes. The fundamental requirement is a significant power supply. Data processing centers in the United States consumed about 70 billion kilowatt / hours of energy in 2014. This is the equivalent to the amount consumed by about 6.4 million average American homes that year.[14] Should the power fail and the computers all turn off, an unacceptable amount of data would be lost. Some, but not all centers have back-up diesel powered generators to provide the required electricity. However, to start the generators, there is a lag-time of several minutes that is covered by thousands of batteries. Nevertheless, the generators cannot run for extended periods of time. Big data centers are now a component of national critical infrastructure. The around-the-clock operation of these centers is vital for the smooth development of nation-wide activities. They impact individuals, the community, the nation and society.

In Latin America and other small countries around the world, cybersecurity has become a major concern over the last decade. Protecting big data has deeply affected Latin American nations. In this regard, less developed countries often turn to larger and more advanced nations to help them protect big data for government and private industries. The United States and Europe have become the source of international assistance for Latin American and Caribbean nations. They assist with challenges, such as the protection of big data in poor energy services environments, inadequate computer maintenance, and insufficient infrastructure protection. One of the main obstacles is the differing perspectives of each country regarding how to deal with these vulnerabilities. This makes cyber threats difficult to combat because of the lack of common public security policy and strategy, or public-private partnership agreements to deal with these threats.

14 Yevgeniy Sverdlik. Here's How Much Energy All US Data Centers Consume. (Data Center Knowledge, 2016).

Better Known and More Commonly Used Database Systems

There are two different systems that are the most widely used for big data: Hadoop and Mongo DB.

Structured Query Language (SQL)- Hadoop

- SQL systems have an ecosystem structure. Several components work together to store and process Big Data at low-cost.
- This database allows queries involving a variety structured data using a SQL-like language.
- Hadoop has a free software license that allows any person or company use it without paying anything.
- At present it is used by a large number of institutions and companies around the world.

Non-Relational Structural Query Language (NOSQL)-Mongo DB

- Mongo DB is a NoSQL Database.
- Free software license that allows any user, person, institution, or company to use it without restrictions.
- Allows storage of large amounts of data.
- Mongo DB features a flexible data structure that allows the integration of information from different data sources.
- Mongo DB usage has surged in the last few years.

Deciding Which Model to Use

Deciding between SQL and NOSQL requires careful consideration of what the end goals are. Some questions to consider include: What is the purpose for collecting the data? What short, medium or long-term effects are expected? How will the data collected be used? How often will the data be updated? What is the expected long-term growth rate? What analytics will be performed on the data?

The cost/benefit analysis of cloud resources versus buying a new computer should be considered. Some SQL and NOSQL databases are completely free and broadly available and therefore may be a more cost-effective option. Finally, the complexity of the queries that is desired of the system should be considered before embarking on a big data project. SQL and NOSQL data models have different characteristics which can impact the usefulness of the results of the queries.

The dependence of Latin American countries on more technologically advanced countries means that the preference for SQL or NOSQL will depend more on global trends set by advanced technology nations, as opposed to their own requirements and analysis needs.

Potential policy and Strategic Solutions for Governments

National security and law enforcement criteria should be considered when making general policy recommendations or crafting legislation. Policy is a delicate balance between giving a government the sufficient power to effectively combat cybercrime while also preventing the abuse of power. Citizens need government protection, but also in some cases, protection from the government. The following is a series of issues to be taken into consideration in the development of government solutions at the policy and strategic levels that will require a significant departure from the past.

Transparency	Government should reveal the type and amount of data they are collecting and how it is being used. They should not release the contents of personal data.
Better Oversight	There needs to be more legislation for the control and supervision of data collection by the intelligence and police communities. This will increase the strength of democracy in Latin America.
Judicial Oversight	The judicial branch needs to be more independent so that they can limit the ability to conduct surveillance without being influenced by other branches.
Fixing Vulnerabilities	We use systems that prioritize efficiency over security and now we are facing high costs to fix these vulnerabilities that hacker could take advantage of.

Separate Espionage from Surveillance	Government-on-government espionage is centuries old and is beneficial for maintaining world order since it reduces uncertainty. On the other hand, domestic surveillance has increased since 9/11 attacks but should be separated from espionage. Rules of probable cause, due process, and oversight should apply to domestic surveillance in light of rule of law.
Limit the Military's Role in Cyberspace	When Latin American countries separated the military from civilian government democracy flourished. Because cyber-attacks are not limited to physical borders it is difficult to determine who is responsible and it is easy to lump all threats together as military threats. The risk then is that military solutions for cyber security will be totalitarian at worst and extra-legal at best. Civilian government should control cyberspace to ensure human rights are upheld.

Cyber Sovereignty Movement

Twenty years ago, few governments had little to no policies regulating the Internet. In today's world, today every country does, and some of them are quite draconian. This shouldn't come as a surprise; the Internet has become too important for governments to ignore. But this change took many internet watchers by surprise, and continues to do so. The fundamentally international nature of the Internet is an enormous benefit for people living in countries that engage in surveillance and censorship. Cyber sovereignty is often a smoke screen for the desires of political leaders to monitor and control their citizens without interference from foreign governments, corporations, or international organizations.

In Latin America, most countries try to control the Internet following the traditional Westphalian sovereignty that focuses on physical borders. However, this concept is not applicable to cyberspace. The debate taking place in the developed world when it comes to governance of cyberspace is based on the technological advances, particularly with the expansion of Internet digital capabilities.

Provide for Commons in Cyberspace

Unowned public spaces have enormous social value. Our public parks, sidewalks, and roads are not owned by any private establishment and remain protected by laws that reflect that public partnership. On the Internet, everything is owned by a private entity. Even websites

independently run are hosted on a corporate server somewhere. There are no commons. According to Bruce Schneier, we need places on the Internet that are not controlled by private parties—places to speak, places to converse, places to operate with special rules treating them as a true commons. There could be common-carrier social networking areas that the corporate owners or government are not allowed to monitor or censor. Whatever the solution, commons are vital to democratic society. It is critical that we work to ensure that they are always present in cyberspace.[15]

In Latin America, the concept of commons as a social value is not well understood. It is in part frequently recognized as the process of democratization. However, a more profound understanding is essential to exercising the right to common unowned public spaces and applying it to cyberspace.

Conclusions

This paper is about the use, misuse, and abuse of big data from the point of view of policy makers and strategists. It is an important topic for a more enlightened approach to national security and defense. Big data offers incredible value for individuals, communities, society, and the nation.

Again, there is value in big data to study social trends and in predicting future events. We have weighed each of these benefits against the risk of surveillance. The primary question is: How do we design systems that make use of our data collectively to benefit society, while also protecting individuals and national security? How do we find an equilibrium for data collection: a balance that creates an optimal overall outcome? These are the fundamental issues of the information age. We can address them, but it will require careful consideration of specific issues and moral analysis of how the different solutions affect our core values.

Cybersecurity has become a major concern in Latin America and the Caribbean in the last decade.. It was not given priority until 1999, when the Organization of American States (OAS) established its first transnational cybercrime alliance. Today, the region has made major advances in cybersecurity. Specifically, the need to develop policies and strategies at a regional level to address not only the threats, but also the development of confidence security measures of mutual trust, common technological advances to contribute to development, and security and defense of the region.[16]

15 Idem page 188-89
16 Organization of American States & Inter-American Development Bank. Observatory Cybersecurity in Latin America and the Caribbean. Annual Report 2016 file:///C:/Saavedrab/Downloads/Cybersecurity-Are-We-Prepared-in-Latin-America-and-Caribbean.pdf

Bibliography

"FTC Settlements Require Equifax to Forfeit Money Made by Allegedly Improperly Selling Information about Millions of Consumers Who Were Late on Their Mortgages." FTC: Protecting America's Consumers. Federal Trade Comission, 10 Oct. 2012. Web. 09 Mar. 2017. <https://www.ftc.gov/news-events/press-releases/2012/10/ftc- settlements-require-equifax-forfeit-money-made-allegedly>.

Goodman, Marc, *Future Crimes: Everything is Connected, Everyone is Vulnerable, and What We Can Do About It.* New York: Doubleday,2015.

Mitchell, Charlie, *Hacked. The Inside Story of America's Struggle to Secure Cyberspace.* Rowman & Littlefield Publishers, 2016.

Organization of American States & Inter-American Development Bank. *Observatory of Cybersecurity in Latin America and the Caribbean. Annual Report.* 2016

Prandini, Patricia, and Marcia L. Maggiore. "Panorama Del Ciberdelito En Latino America." Working paper. Montevideo: Latin America and Caribbean Network Information Centre, 2011.

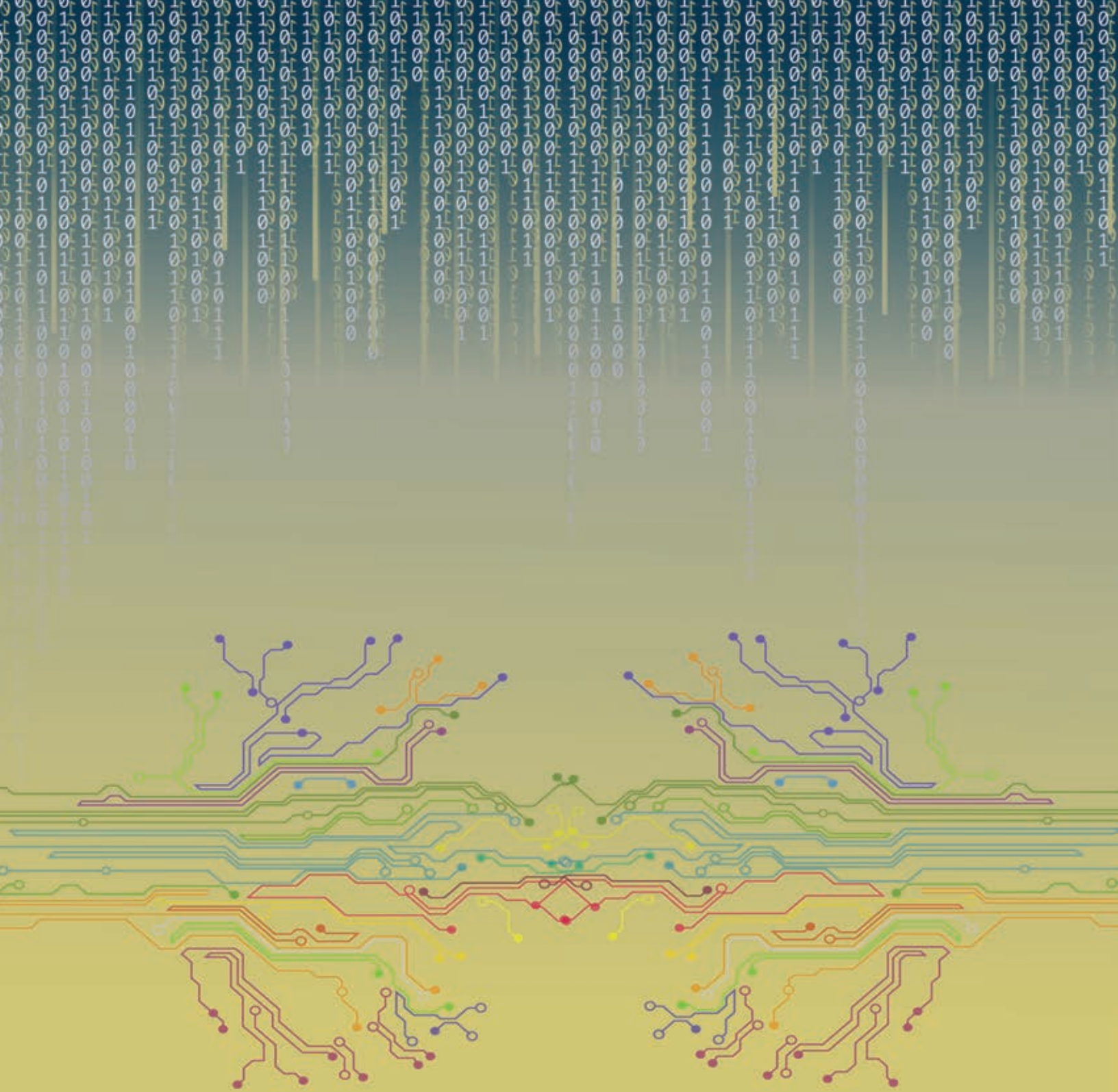
Rafael Caballero and Enrique Martin, *Las Bases de Big Data.* Madrid: Los Libros de la Catarata, 2015.

Schneier, Bruce. *Data and Goliath. The Hidden Battles to Collect Your Data and Control Your World*. New York: W.W. Norton & Company, Inc., 2016

Sverdlik, Yevgeniy "Here's How Much Energy All US Data Centers Consume."Data Center Knowledge. N.p., 27 June 2016. Web. 09 Mar. 2017. <http://www.datacenterknowledge.com/archives/2016/06/27/heres- how-much-energy-all-us-data-centers-consume/>.

Whitchorn, Mark. "Big Data Bites Back: How to handle those Unwieldy Digits when You can't Just Cram It into Tables", last modified August 27, 2012. www.theregister.co.uk/2012/08/27/how_did_big_data_get_so_big/

Weber, Max. "Politics as a Vocation". http://anthropos-lab.net/wp/wp-content/uploads/2011/12/Weber-Politics-as-a-Vocation.pdf



National Defense University
Abraham Lincoln Hall
260 5th Ave. Bldg. 64
Washington, DC 20319-5066