

Louvain School of Management

Data-driven governance: application assumptions to the State and public services of Big Data practices used by the electoral sector.

Author : Verbiest Alexis
Supervisor(s) : Kolp Manuel
Academic year 2019-2020

First of all, I would like to express my heartfelt thanks to Mr. Kolp, my promoter, for his availability, support and valuable advices. His expertise in Big Data has undeniably helped me in the realization of my work. I would also like to thank him for having placed his trust in me and for having supported my subject from the start.

Secondly, I would like to thank all my professors at the Louvain School of Management. Their teaching has enable me to acquire knowledge that has contributed to the development of my thesis.

In addition, I would like to warmly thank Matthieu Aladenise, Stéphane Faulkner, Nicolas Dombret and Philippe Maes, for giving me their time through an interview. The completion of my work would not have been possible without their input.

Double thanks Martin de Waziers, for the careful rereading of this work, as well for your constructive advices and for your interview.

Finally, I would like to thank my parents for their undeniable support and encouragement on a daily basis. I would also like to thank my family and friends as well all the people who contributed, from near or far, to the accomplishment of my work. Your curiosity about my subject, as well as your advice and encouragement have motivated me enormously.

Table of Contents

Introduction	1
Part 1: The Big Data Revolution.....	3
1. The different definitions of Big Data	3
2. The characteristics of Big Data	5
2.1 The three basic "V's"	5
2.1.1 Big Data Volume	5
2.1.2 Big Data Variety	6
2.1.3 Big Data Velocity	6
2.2 More and more "V's".....	7
2.2.1 Big Data Veracity	7
2.2.2 Big Data Value	7
2.2.3 Big Data Visualization	8
2.2.4 Viscosity, Variability, and Volatility of Big Data	8
2.2.5 Big Data Vulnerability	9
3. Big Data analytics.....	9
3.1 Defining Big Data analytics.....	9
3.2 The Power of Algorithms	10
3.3 The major challenges of algorithms faced by the public sector	11
3.3.1 Verifying the source of the data.....	11
3.3.2 Trusting algorithms.....	13
3.3.3 Continuing to look for causes	14
4. The sources of the Big Data	15
4.1 The Three Categories of Sources	16
4.1.1 Directed Sources.....	16
4.1.2 Automated sources.....	16
4.1.3 Volunteered sources.....	17
4.2 Open Data	17
5. The tools of Big Data	19
5.1 The Apache Hadoop Platform	19
5.2 Non-relational databases	20
5.3 Cloud Computing.....	20
5.4 Data mining systems	20
6. From Big Brother to Big Data	21
7. Our own definition of Big Data	22
PART 2: The 4 major trends identified from Big Data in the Electoral sector	23
1. The Origin of Electoral Big Data: American Campaigns	23
1.1 The Beginnings of Electoral Big Data in the United States	23
1.2 Major differences between campaigns in France and the United States	25
1.2.1 The difference in financial resources.....	25
1.2.2 The difference in legislation	26
1.2.3 Differences in organizational methods	26

2. The interaction between political parties and the market players of the electoral Big Data	27
3. The electoral Big Data market players	28
3.1 Who are the commercial actors of Elections 2.0?	28
3.1.1 The emergence of the market under US influence	28
3.1.2 Heterogeneity of entrepreneurial forms and logics	29
3.2 What do market actors do in elections 2.0?	29
3.2.1 A multiplicity of means for a duality of services	30
3.2.2 The Three Types of Actors in Elections 2.0	30
3.2.2.1 Specialized actors.....	31
3.2.2.2 The integrated actors.....	31
3.2.2.3 The relay actors.....	31
3.3 The partisan dimension of market players	32
4. Big Data, political parties, and elections	32
4.1 Campaign team data.....	33
4.1.1 Data collection.....	33
4.1.2 The creation of electoral software in France	34
4.2 The use of data during electoral campaigns	36
4.2.1 Mapping.....	36
4.2.2 Co-constructing the programs.....	37
4.2.3 Candidates' websites, which have become interactive Big Data platforms	38
4.2.4 Mailing.....	38
4.2.5 Phoning.....	40
4.2.6 Door-to-door contact	40
4.2.7 Predicting voter behavior.....	41
4.2.8 Improving Internal Party Management.....	42
4.2.9 Renewed activism.....	43
4.2.10 Managing and Creating Community	44
4.2.11 Voter profiling	44
5. General Process of the Electoral Big Data in France.....	45
6. The prospects of the electoral Big Data	46
6.1 Standardization of the commercial actors of elections	46
6.2 Towards an American-style two-party system	47
6.3 Lower and lower barriers to the Presidential Office	48
6.4 Information disruption and polarization of Society.....	48
6.5 Change in power and influence in the data economy	49
6.6 A regularization of the use of data	49
7. The major trends identified.....	50
7.1 The application of electoral Big Data summarized by 4 major trends	50
7.1.1 Personalization.....	50
7.1.2 Predictability.....	50
7.1.3 Optimization	51
7.1.4 Objectivation.....	51
7.2 Summary Table of Electoral Big Data Techniques in Relation to the Four Major Trends	51
Part 3: Data-driven Governance	52
1. How can the State use Big Data?	53
1.1 Personalization of service to the citizen	54
1.1.1 From standardization to personalization.....	54
1.1.2 The case of teaching and education.....	55
1.2 Predictability: preventing fraud and threats	58
1.2.1 From retrospective to prospective	58
1.2.2 Cases of prevention against tax and social fraud.....	59

1.3 Internal optimization	61
1.3.1 From organization to optimization	62
1.3.2 The case of road traffic	62
1.4 Objectivation of decision making and judgement	64
1.4.1 From objectivity to objectivation thanks to Open Data	65
1.4.2 The case of connected objects “secured by design”	66
2. What are the challenges and risks of Data-driven Governance?	68
2.1 Challenges for the State	68
2.1.1 A technical and infrastructure challenge	69
2.1.2 A skills challenge	69
2.1.3 A challenge of mentality	69
2.2 Risks for the citizen	70
2.2.1 A data security risk	70
2.2.2 A risk to individual liberties	70
3. Audit of the four sectors studied	71
3.1 The evaluation grid	72
3.2 The criteria used	72
3.3 The final audit table	74
Conclusion	75
Bibliography	81
Appendix	92
Interviews	112

Introduction

Big data is, without us realizing it, changing our view of the world. It is becoming so important that entire courses are now devoted to it in university curricula. Big data has become a new economic resource for businesses. For example, it is estimated that Amazon derives a third of its revenue from its recommendation system, which is based on Big Data techniques. We now have more data at our disposal than ever before in human history. Volumes of data are estimated to have increased 100-fold between 1987 and 2007, and have doubled every year since then. It is estimated that in 2020, Big Data will generate 203 billion dollars revenue, compared to only 130.1 billion dollars in 2016. IDC and SAP have predicted that 60% of the world's manufacturers will use Big Data, particularly to analyze manufacturing processes and identify optimization opportunities. In addition to the enormous opportunities for economic development, especially for emerging countries, Big Data will become an increasingly political issue in the future. It will constitute a new issue of rivalry between States. The importance of data in the field of politics and governance is of particular interest to us. In politics, everyone has heard about the “Cambridge Analytica” case. The company boasted that it had collected nearly 5,000 data points on every American voter. We are at a turning point in our history where we are just starting to realize how much data companies and States hold on us. Now, we will have to decide how best to manage these infinite volumes of data, whether it is for marketing, during election campaigns or for governance. Big Data is everywhere, impacting all sectors of the economy and individuals, while being almost invisible. As complex as the problem is, it is important to inform people as much as possible about the use of data, to explain its qualities and opportunities, but also its risks and dangers.

Being passionate about politics, particularly in France, and "fascinated" by Big Data, I wanted to make it my research topic. During the 2017 presidential campaign in France, which I followed closely, I heard several times about Big Data, as if some algorithms could predict voter behavior. So, as a citizen, I became interested in this subject, because if algorithms really do predict how we are going to act and vote, are we still in a democracy? My original intention was to do a quantitative analysis and to show how, using databases, political parties or private companies could perform predictive analyses of voter behavior. Unfortunately, after weeks of intensive research, I realized that these databases were completely inaccessible and locked. I was probably very naive about that. So I had to redirect my research question. So I finally decided to base my research on the Big Data oriented models used during election campaigns, to deduce possible applications in a Data-driven governance perspective. Thus I dedicated my thesis to the following research question: *"On the basis of the major trends identified in the use of Big Data in the electoral sector, what would be the possible applications of a Data-driven governance in some public sectors in Belgium?"*

The first part of the work defines Big Data, in order to better understand the concept, its different characteristics and the different data sources and tools used by data workers. The second part is a study of the electoral Big Data sector, in order to have a concrete example of the techniques and tools used presented in the first part. In this second part, I explain the origin of the electoral Big Data, the specific data sources used by political parties in France. I also discussed the two main actors of the electoral Big Data: market players and political parties. Finally, I have studied all Big Data oriented campaign techniques, which has allowed me to identify four major trends that I have taken up in the third part. These four main trends are: personalization, predictability, optimization and objectivation. I concluded the second part with a comparative table of electoral Big Data techniques according to the four trends. In the third part, for each trend, I then applied to a public sector a data-driven governance hypothesis. The four sectors studied are education, tax and social fraud, road traffic and finally the sector of connected objects “secured by design”. The ultimate objective is to audit the sectors studied and determine whether or not it is possible and reasonable to opt for data-driven governance.

In order to answer my research question in the best possible way, several types of sources were used, including scientific articles, newspapers, books and websites. Interviews were also conducted. Subsequently, the analysis that was carried out in the practical part was based on qualitative research, conducted through exploratory interviews, failing which a quantitative analysis with a database. Finally, I ended my study with an audit table of the four sectors studied, based on criteria taken largely from academic sources. I also asked each interviewee to fill in the audit table in order to be able to make a better synthesis. Obtaining answers to the questions was not easy because of the current circumstances.

As we all know, the world is experiencing an unprecedented health crisis. The Covid-19 crisis has had a significant impact on the population, including students. On the one hand, the Lockdown has allowed me to concentrate fully on my dissertation and avoid any distractions, but it has also had a negative impact on the development of my research. Indeed, the responsiveness of the people interviewed often took longer than expected. Not being able to access libraries was also complicated, both in terms of concentration and access to certain sources. But the major difficulty I encountered during my dissertation was really the loss of time spent researching databases, which I never obtained, and the fact that I had to rethink my subject entirely. However, I have put all the means on my side to find alternatives and to do this work in the best possible way.

Part 1: The Big Data Revolution

In this first part, we try to understand what Big Data really is, but also, what it is not. This term is used indiscriminately, and it will therefore be a question of defining the contours of the Big Data. The Big Data is radically changing the way we live, the way we do business, without us necessarily realizing it. To begin with, we will discuss the different definitions that professionals associate with this concept. Afterwards, we will try to break down Big Data to see what it is made of and to be able to identify its main characteristics. Next, we will look at the different sources of Big Data and the tools and techniques used to manipulate this mass data. We will particularly focus on the algorithms in order to realize their power but also their flaws. Thereafter, we will study the risks generated by the data revolution for our civilization. Finally, based on all that we have discussed, we will propose our own definition of big data.

1. The different definitions of Big Data

There is no unanimously accepted definition by big data experts, nor is there an official definition. There are two main reasons for this. The first is that the Big Data is changing so rapidly that definitions are becoming obsolete very quickly (Davenport, 2014). The second reason is that we know very little about Big Data compared to what it really is, as is well illustrated in Appendix 1. People who are interested in this subject are still trying to understand and define this phenomenon, which is beyond even the biggest players in technology.

According to Press (2014), the term Big data was first used in 1997 by NASA scientists because they had a problem with computer graphics, i.e. data visualization. The first time the term was used was because the data set did not fit into the main memory. In other words, computer media such as computers and their local or even remote disks were not powerful enough to store the amount of data available. It was only in 2008 that the term was popularized by university scientists Bryant, Katz and Laszowska. They refer to the term "Big-Data computing" in an article entitled: *"Big-Data Computing: creating revolutionary breakthroughs in commerce, science and society"*, without defining it (Press, 2014).

In order to browse through the different definitions of Big Data, one should refer to authoritative reference dictionaries such as the Oxford English Dictionary (OED) which defines Big Data as *"Sets of information that are too large or too complex to handle, analyze or use with standard methods"* (Oxford English Dictionary, 2020, p.1). The term "Big Data" officially entered the dictionary in 2013.

Two years earlier, in 2011, the McKinsey Global Institute (MGI) also provided a definition of Big Data, which they believe is deliberately subjective. They define Big Data as follows: *"Big Data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage*

and analyze" (MGI, 2011, p.1). They deliberately define Big Data in a subjective way because, according to them, everything is relative. They therefore define Big Data as what will always be beyond the power of our technologies.

In the same year, Grantz and Reinsel of the International Data Corporation (IDC) proposed the following definition: *"Big Data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis"* (IDC, 2011, p.6).

Another reference book is the work of Mayer-Schönberger and Kenneth Cukier. They realized that there is no rigorous definition of Big Data, and therefore proposed another approach to the definition. They define Big Data in terms of what can be done with data, as follows: *"The ability of society to harness information in novel ways to produce useful insights or goods and services of significant value and things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value"* (Mayer-Schönberger & Cukier, 2014, p.8). The authors add that Big Data will transform the way we live, work, and think.

Second, Gartner Inc. which is an American advanced technology research consulting firm, defines Big Data as follows: *"Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insights, decision making, and process automation"* (Gartner Glossary, 2020, p.1). We can already note a similarity with the definition given by Grantz and Reinsel, as we find the terms **"volume, variety, velocity"** in both definitions. In the next section, which deals with the characteristics of Big Data, we will see why these three terms are most important.

Finally, the last interesting definition to be discussed is that of the TechAmerica Foundation, which proposes this definition: *"Big Data is a term that describes large volumes of high velocity, complex and variable data that require advances techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information"* (TechAmerica Foundation's Federal Big Data Commission, 2012, p.7).

We note that there are two approaches to the definition of Big Data. A first approach that explains Big Data by its nature, and a second approach that defines it in terms of what it can do. The common point between all these different definitions is that Big Data refers to a large amount of data that is difficult to master but which has enormous potential. The potential of this data lies in the creation of value, whether from an entrepreneurial point of view, to increase the performance of a company, or from a human point of view, to improve the life of people. But apart from the large volume of data, Big Data has other characteristics that will be discussed in the next section.

2. The characteristics of Big Data

In addition to the academic definitions of Big Data cited in the previous section, many authors in the field of technology choose to define Big Data by explaining its various characteristics. Everyone has already heard about **the "3 V's" of Big Data**. But we will see that today, the characteristics have been extended beyond these basic "3V's". Not everyone agrees on the definition, but everyone agrees that the following "3 V's" are in all cases part of the main and essential characteristics.

2.1 The three basic "V's"

It was Doug Laney, in 2001, who for the first time, in an article entitled "*3D management: controlling Data Volume, Velocity, and Variety*", talked about the famous "3V's" of Big Data, which are Volume, Variety, and Velocity (Laney, 2001, p.1). Already in the early 2000s, long before the definitions cited in Section 1, the three characteristics were in the minds of technology experts. We will therefore present here the three basic "V's", the three basic characteristics of the Big Data.

2.1.1 Big Data Volume

As presented in the introduction, it is, first of all, the immensity of the amount of data that characterizes Big Data. While in 2011, 1.8 zettabytes represented all the information worldwide, in 2019, it is estimated that the volume of data on global IP traffic was 41 zettabytes. To give a scale, one zettabyte represents 10^{21} bytes (Statistica, 2020). It is estimated that by 2025, the amount of information and data created worldwide will reach 175 zettabytes, as illustrated in Appendix 2 (Bastien, 2015). This data growth is due to digital developments, and more. It is also due to the development of emerging countries, particularly in Asia, such as China, which has overtaken the USA in terms of data production. The second factor that explains this increase in data is the effervescence of IoT (Internet of Things), i.e. the desire to extend the Internet network and thus data exchanges to objects in the physical everyday life such as watches, glasses or even cars (Rouse, 2015). It is also important to note that this increase in data goes hand in hand with a change in infrastructure. Indeed, to cope with these evolutions, data storage infrastructures will have to adapt. Likewise, the power of servers and the bandwidth of fixed and mobile IP networks will have to be increased to support this massive volume of data.

2.1.2 Big Data Variety

Data is considered varied when it comes from multiple sources, such as sensors, smart objects, the Internet, and social collaboration technologies. Data no longer includes only traditional relational data, but also raw, semi-structured or unstructured data from web pages, online files, blogs, social networks, e-mails, videos, and sensors installed on all kinds of objects (Zikopoulos et al., 2012). Since the data have different origins, they also have different formats (text, image, audio, video, etc.).

According to the journalist Beal, today 80 to 90% of data is unstructured data (Beal, 2020). The challenge for organizations is therefore to try and sort data so that they can extract information that produces value. According to Grantz and Reinsel, sorting data to only keep useful one is even the major challenge of Big Data (Grantz & Reinsel, 2011). Indeed, today, everything has been put in place by organizations to collect as much data as possible to utilize information from specific categories. For example, in business, human resources software is designed to manage all internal personnel data, while customer relationship management (CRM) software holds data on all customers.

Organizations can no longer be content to manage only one or the other category of data, but are forced to juggle with heterogeneous databases, in which the data are different in both origin and nature. For example, a company's challenge will be to mix data from social networks with customer data from a CRM program. Then, it will have to analyze the mix, which will allow it to draw conclusions in order to move to personalized marketing (Diez, 2016).

2.1.3 Big Data Velocity

According to the TechAmerica Foundation's report, velocity refers both to the speed at which data is produced and transformed, and the speed at which data needs to be understood and processed by organizations (TechAmerica Foundation's Federal Big Data Commission, 2014). We can therefore speak of Big Data when data is collected in 'real-time' (Davenport et al., 2012). In terms of data speed, the increase in the number of Radio Frequency Identification (RFID) sensors and other information transfer tools has led to a continuous flow of data whose pace has become unmanageable for traditional data management systems. This is what Zikopoulos et al. call "*data in motion*" (2012, p.8).

Moreover, in today's world, real-time data analysis is essential because of the speed of its availability. To stay ahead of competition, it is foremost to be the first to identify a trend, a problem, or an opportunity. This shows the increasing importance of speed. Analyzing data in real time is a real means of creating value for companies and improving their economic performance (Zikopoulos et al., 2012).

A very good illustration lies in Appendix 3, a computer graphic that shows everything that happens in one minute on the Internet:

- **Volume** is represented by numbers, that is, quantities of data.
- **Variety** is represented by the 17 different listed sources generating data in a wide variety of formats (texts, photos, videos, queries, connections, actions, etc.).
- **Velocity** is represented by the short time it takes to generate all this data on the Internet, in this case 60 seconds.

2.2 More and more "V's

As knowledge about Big Data has grown over time, technology experts have found other features, also beginning with the letter "V".

2.2.1 Big Data Veracity

Several major technology players add a fourth "V": Veracity. According to IBM data scientists, it is simply defined as "*the uncertainty of data*" (IBM, 2012, p.1). Appendix 4 provides IBM's representation of the four characteristics of Big Data. According to the Gartner research, poor data quality has a major impact on the way businesses operate. On the one hand, the quality of data influences roughly 20% of productivity. On the other hand, when you do not meet your business objectives, 40% of it is due to the lack of data quality. Finally, they estimate that in 2/3 of organizations, more time is spent on data processing than on data analysis; poor data quality costs an average of \$15 million per year, each, according to Gartner's 2017 Data Quality Market Research (Gartner, 2017). IBM, meanwhile, has estimated that in the United States alone, poor data quality is costing companies \$3100 billion, as shown in Appendix 5. These few figures show how important it is to have quality data.

Another question therefore comes to mind: why can the data be of poor quality? The TechAmerica Foundation explains that the quality of information can be biased because of data inconsistency, incompleteness, ambiguities, deception, or model approximations. Therefore the veracity of data is an issue and decisions that are taken on the basis of data must be solidly justified. Finally, with the proliferation of social networks such as Facebook and Twitter, information from these Big Data sources needs to be scrutinized for veracity before it is used to make decisions (Liebowitz, 2016).

2.2.2 Big Data Value

It is difficult to know who was behind each of these characteristics. Nevertheless, major institutions such as the Harvard Business Review add a fifth "V" to describe the Value of Big Data. In their article

"*The Value of Big Data isn't the Data*" (Hammond, 2013, p.1), Hammond explains that it is not the data itself that constitutes the value of Big Data, but rather the ability of that data to generate profit. This is about the intelligence that data can produce. For organizations, this value manifests itself as much through the improvement of existing services as through the creation of new services in order to continually find new sources of revenue (Hamel & Marguerit, 2013).

The McKinsey Global Institute (2011) confirms the importance of Value when talking about Big Data. McKinsey's research also shows that data can create significant value for the global economy by improving productivity and competitiveness. This applies to both business and the public sector. Moreover, it creates a substantial economic surplus for consumers (MGI, 2011, pp.1-2). On this subject, Mr. Faulkner explained that, in his opinion, the value of data is the real issue today. He believes that as long as we have valuable data, and also people who can make it usable, Big Data could help us make better decisions in all areas. So we need people who can control the data at the source but also afterwards, once it has been reworked, reformatted for analysis. Once you have this added value of expertise, all sectors should benefit from additional information from a data environment.

2.2.3 Big Data Visualization

The SAS (Statistical Analysis System) institute, an American software publisher specializing in business intelligence, adds this sixth dimension. According to them, visualization is defined as "*the visual and interactive exploration of data of any volume, nature (structured or unstructured) and origin, and their graphical representation*" (SAS, 2015, p.3). Indeed, the knowledge resulting from data analysis must be shared in an efficient and comprehensible manner, especially to the company's shareholders. Given the astronomical amount of data, it is essential to be able to present the information in a clear and legible manner, which is a real challenge for data workers.

2.2.4 Viscosity, Variability, and Volatility of Big Data

In his study "*Realizing the Promise of Big Data*", Desouza chose to define Big Data according to 7 criteria. In addition to Volume, Variety, Velocity and Veracity, already discussed above, he proposes three other "Vs" (Desouza, 2014):

- **Viscosity**, which measures the resistance released by the Big Data when navigating through the data. This resistance is due to the fact that the data come from different sources, that there are different rates of data flow, or that the processes required to transform the data into useful information are very complex.
- **Variability**, which takes into account changes in data flow rates.
- **Volatility**, which indicates how long the data is valid and for how long it should be stored.

2.2.5 Big Data Vulnerability

Experian, a company specializing in credit risk, talked about a feature of Big Data that was not mentioned anywhere else. Indeed, when the other characteristics speak of the Big Data's capabilities, among other things, they evoke its fragility, raising an important aspect (Marr, 2016). Indeed, of all the data generated every second, some can be made public, but others have a confidential nature that is crucial for the organization that holds it. Marr defines Vulnerability as follows: "*Vulnerability addresses the fact that a growing number of people are becoming switched on to the fact that their personal data, the lifeblood of many commercial Big Data initiatives, is being gobbled up by the gigabyte, used to pry into their behavior and, ultimately, sell them things*" (Marr, 2016, p.1). We can ask ourselves how it is possible to protect such a large amount of data. As the platforms that contain this data move in hybrid technology, managing their security and availability becomes another challenge of Big Data, which requires ongoing diagnosis and monitoring (Diez, 2016).

We now have a good overview of all dimensions of Big Data. Some characteristics define what it is (Volume, Variety, Velocity), another what it can produce (Value), while others reflect the challenges (Veracity, Visualization, Viscosity, Variability, Volatility, Vulnerability). However, Big Data, by itself, does not create value (Hammond, 2013). This will be discussed in the next section.

3. Big Data analytics

3.1 Defining Big Data analytics

The aim is to study how data are used: to make sense of available data. We will thus address algorithms. If we could make an analogy between Big Data and cars, we would say that data is the essence of Big Data, but the algorithms are the engine (Bizet, 2018, p.165).

IBM gives the following definition of Big Data analytics: "*Big Data analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes*" (IBM, 2017, p.2). As mentioned earlier, Big Data applies to data sets whose size or type exceeds the capacity of traditional databases, which manage data with low latency. Artificial intelligence (AI), mobile communication, social networks and the Internet of Things (IoT) are creating complexity due to new forms and sources of data. Big Data comes for example from sensors, video/audio files, transactional applications, the web and social networks. Much of this data is generated on a very large scale and in real time (IBM, 2017). Big data analysis enables businesses to improve decision making and make it faster using data that was previously inaccessible or unusable. There are several types of analysis techniques: text analysis, machine learning, predictive analysis, statistics, natural language processing,

etc. These new techniques make it possible to obtain new information from previously untapped data sources. Moreover, it allows these data to be cross-referenced with data that companies already possess, which is a source of enormous value creation.

In 2015 Gartner published an article entitled "*It's not just about Big Data, it's about what you do with it: welcome to the algorithmic economy*" (Gartner, 2015, p.1). Indeed, in addition to having access to large volumes of data of various kinds, powerful data analysis capabilities are critical for organizations that want to create value with Big Data. These capabilities include not only technical tools, but more importantly, the skills to analyze them. That is why we talk about "Big Data analytics", a term that goes hand in hand with the Big Data revolution (Diez, 2016).

About Data-driven governance, Mr. Faulkner, who teaches Big Data at the University of Namur, explains that he doesn't really like the term. He explains that more and more decisions in the business world are based on data. However, he explains that the final decision-maker is always human being. Therefore, algorithmic governance, where decisions are based solely on data, should not be confused with data-driven governance, where data is used to make better decisions. This is a real change that he sees in everyday life, compared to before. He defines Data-driven governance as follows: "*There is a technological evolution that makes the decision, or at least the spectrum on which the decision is based, different. And I find somewhere that when there is a decision, there is governance, so today governance can be data-driven*".

We shall now review what an algorithm is and why it is so important in Big Data, as well as the possible problems and challenges they face in the public sector particularly.

3.2 The Power of Algorithms

Before starting to address this issue, it is important to define it. As with Big Data, there are several definitions of an algorithm. The one we will use is the Cambridge Dictionary definition: "*A set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem*" (Cambridge Dictionary, 2020, p.1).

Until a few years ago, only computer scientists and mathematicians talked about algorithms. Today, algorithms are in the public debate and are often perceived by people as a kind of occult power that would allow big companies like GAFAM to dominate the world (Cardon, 2018). But, let us also underline the fact that we have to "*relate the new popularity of algorithms to a shift in the form taken by computability in the construction of digital services*" (Cardon, 2018, p.2). The relevance of Web services today depends more and more on their ability to sort, prioritize or personalize an ever-increasing

number of data. In 2017, The Economist published an article "*The World's most valuable resource is no longer oil, but data*" (Economist, 2017, p.1). In some ways, this may seem true, because the race for data does exist. But in another way, the main challenge of Big Data is not simply to collect as much numerical data as possible, but to increase the strength and accuracy of their models' calculations. This explains the emergence of algorithms as a new form of power. They are the indispensable instrument for classifying information (search engine), customizing advertising displays (behavioral targeting), or making recommendations (cultural predictions) (Cardon, 2018). As we have already mentioned, it is therefore the ability to use the data that takes precedence over the ability to collect it. However, these algorithms used by Big Data are increasingly criticized, and we will see why in the next section.

3.3 The major challenges of algorithms faced by the public sector

In 1751, Diderot was already talking about a world governed by numbers: he was a visionary because today it is practically feasible. Indeed, due to the extraordinary evolution of technology and the incredible evolution of the amount of data available, algorithms can provide us with an almost scientific approach to reality. The dream of the 17th century political algorithm would almost becoming true (Bizet, 2018). In this section, we will mainly oppose three authors, Peugeot, Bizet and Cardon. Bizet and Peugeot puts forward the problems of algorithms, and Cardon, in the article "*Le pouvoir des algorithmes*" (Cardon, 2018, p.1), tries rather to answer Bizet's questions. The objective will therefore be to present the 3 major challenges facing the use of algorithms in the public sector, a practice that deserves a critical approach to avoid abuse and errors (Bizet, 2018).

3.3.1 Verifying the source of the data

The first question to ask according to Bizet is: where does the data come from? In the context of public services, most of the data are collected by the administrations themselves. These data are also data collected over a long period of time, to measure and manage public actions but also to constitute public statistics. These data are also of the "Open Data" type, which we will define in the next section. Administrations publish the data in a "transparent" way, and the data circulate from the public to the private sector. Once these data are released, everyone is free to use them and do what they want with it (Peugeot, 2011). Nevertheless, there are of course laws that regulate the use of these data, such as the CADA law (1978) or the "Lemaire Law" (2016), whose purpose is, among other things, to anonymize data as much as possible to avoid the publication of personal data (Bizet, 2018). We will not discuss these laws in detail because this is not the subject of the brief. Despite the fact that Mr. Faulkner is more in favor of governance based on more and more data, he confirmed that the legislative framework, despite the RGPD, still allows a lot of interpretation. While being in favor of the use of Big Data in the

public sector, he stresses the importance of certain ethical rules and the protection of privacy. All this needs to be better framed by legislative rules.

Moreover, these laws are quite difficult to implement in a "Big Data" context where the proportionality of the data is not clearly defined and is almost impossible to define as the volume of data is rapidly increasing. The problem raised by Bizet is not actually at the level of the administration, which complies with these rules, most of the time! Indeed, in the data published by the administrations, little public information contains personal data. The real problem is that once this data is released, private companies can process this data, cross-referencing it with other data sources, such as data available on social networks. After numerous processing operations and via the computing power of algorithms, the initial "anonymized" data can be transformed into personal data (Peugeot, 2011). Anonymization of data is therefore clearly not effective (Montjoye et al., 2015). As Mr. Maes confirmed, the data he works with in the manufacturing sector is not anonymized at all. This confirms the hypothesis that, once reworked, the data, anonymous at the outset, can become personal after being cross-referenced with other databases, at least in the private sector. Mr. Dombret, with the example of car plates, explains that it is very easy, from a plate number, to find the person's name, address, etc. by cross-referencing the data, and that this becomes completely private data.

The problem underlying this analysis is the openness of public data in the context of "open government". We are entitled to ask whether it is really necessary to publish a lot of government data in Open Data. Peugeot (2011), in its article entitled: *"L'ouverture des données publiques: convergence ou malentendu politique ?"* (Peugeot, 2011, p.1) also addresses this question. We can summarize the two "official" arguments raised by those defending the openness of public data. On the one hand, the first argument systematically put forward is that of "democratic deepening", which translates into a desire for transparency and efficiency and raises the question of the place granted to the citizen in a democracy. On the other hand, they often refer to the stimulation of economic and technological innovation, based on the fact that the opening up of these data would be supposed to allow the development of new services and thus an improvement in the daily lives of citizens. We shall also refer to the virtuous circle proposed by the advocates of the release of public data. By releasing information, the State will position itself as a generator of economic activity and therefore of employment and growth, from which it receives a benefit in return via the taxes levied. It is therefore a kind of "return on investment" (Peugeot, 2011).

The pretext of democracy for opening up public data does not actually come from a real citizen demand, but from a very powerful lobby, that of the major technology players, who have understood the value of this data. Indeed, the real question we are asking ourselves is: why is it that the large multinationals that are demanding the opening up of public data do not apply this principle to themselves? We noticed in the Netflix documentary *"The Great Hack"* (2020), that when Professor Carroll demanded that

Cambridge Analytica give him back the "5000 data points" it had on him, they refused. The hidden issue behind opening up public data is therefore their reuse. If these data were released to allow the citizen more transparency, it would not be a problem. But, to the readers of this brief, we will ask the following question: how many ordinary citizens in your entourage will use or consult this public data on a regular basis?

3.3.2 Trusting algorithms

Contrary to what many people think, the purpose of Big Data is not to collect as much data as possible, but rather to develop models, i.e. algorithms, which make it possible to improve a situation and provide answers (Bizet, 2018). To counter the authority argument of "*proclaimed true, because mathematical*" (Kybernetes, 2015, p.1004), Loukdies proposes to open up models, not only data (Loukdies, 2014). Indeed, openness would allow everyone to understand the models applied and would allow outside communities to analyze, criticize, and even improve the models used. Because in the end, we can have all the data we want, whether for health, crime or other purposes, if we don't know which models are used to collect them, we won't get many answers. In this regard, mathematician O'Neil said: "*you don't really know what a model does until you can interact with it. You don't know if a model is robust unless you can fiddle with its parameters and check. And finally, you don't know if a model is best possible unless you've let people try to improve it.*" (O'Neil, 2013, p.2). Mr. Dombret gives us a concrete example of an application where he would like to open up the algorithmic model. For some years now, we have been forced to introduce our number plates when we park. But no information is given by the State on what is done with this data. Is it only for statistical purposes? Or is it really to track the movements of people? We don't know, and this is often the problem with the data we are forced to provide.

But the openness of models is an annoying subject in administrations because these models concentrate the knowers' power (Bizet, 2018). These models, which may seem objectively neutral and fair because they are mathematical, are strongly criticized. It is their neutrality that is questioned. If the models could be considered as neutral, their creators may not be (McGinty, 2016). Indeed, even if the technological process of algorithms generates fascination, human discretion still plays a role. Algorithm creators can leave their ideological imprint, for example, by choosing which databases they will use, the perimeter, etc. (McGinty, 2016). But models can also be biased by other factors, such as their subsequent use. Indeed, many "predictive" algorithms can suffer from self-fulfilling predictions by validating themselves as the predictions are made. To sum up, we can therefore say that, hidden behind the supposed mathematical neutrality, models carry with them the biases of their creators (Bizet, 2018). Therefore, public models should be open in order to be controlled, especially since the control of external entities can be more effective because non-governmental agencies often have more time, skills and means at their disposal (Peugeot, 2011). Communities could also make these algorithms understandable

to ordinary people. Mr. Dombret, who works at Ethias (insurance), explains that the algorithms are partly based on intuition. He says: *"This intuition, in some cases, will also be used to guide us to create a predictive model rather than plotting millions of raw variables"*. According to him, a mix of intuition and data is therefore necessary.

But even if the models were open, the trust we could place in them remains limited, because the control methods do not allow, *ex ante* or *ex post*, checks since the very structure of the model changes with the results. With Big Data, traceability becomes more and more difficult. The race for transparency behind the evolution of data systems therefore seems a real challenge, for which we are not well prepared (Bizet, 2018). To this problem raised by Bizet, Cardon provides the following answer: *"algorithms are not neutral, but they must be loyal"* (Cardon, 2018, p.5). Indeed, we cannot ban algorithms for not being neutral, even though they are designed to select, sort, filter, or order information according to certain principles. Algorithmic biases do exist, but we often find it impossible to oppose them with a neutral representation of the information, from which the biases could be observed. The solution would therefore be to study the alignment between the service that the platform claims to provide and the reality of what it offers. This clarification is necessary to avoid the imputations of intentionality that we sometimes address to algorithms when we contest their result: "Google is racist", "Facebook publishes fake news", etc. This way of talking to algorithms as if their decisions were human, forgets that their calculations proceed from a cold procedural rationality (Cardon, 2018).

3.3.3 Continuing to look for causes

Bizet's third challenge is to keep looking behind the causes (Bizet, 2018). Big Data technologies look for similarities and correlations, to predict outcomes and correct consequences. This is probably a way of providing concrete solutions to problems that have been identified, but it does not allow the cause of the problem to be identified. In their book, Cukier & Schönberger state: *"we're entering a world of constant data-driven predictions where we may not be able to explain the reasons behinds our decisions"* (Mayer-Schönberger & Cukier, 2013, p.17). The pursuit of a data-driven model should therefore not make us forget the treatment of the cause, which is much more subtle in terms of numbers, and which fades away at the same time as the cause (Bizet, 2018). As the mathematician O'Neil states, *"A formula, whether it's a diet or a tax code, might be perfectly innocuous in theory. But if it grows to become a national or global standard, it creates its own distorted and dystopian economy"* (O'Neil, 2016, p.49). In an imaginary future governed by automated decisions, there is a risk that these could become standards without us being able to explain why. This governmentality must therefore be accompanied and critically debated (Bizet, 2018).

On this subject, Cardon (2018) explains that algorithms are blamed for things for which they were not originally designed. Indeed, current algorithmic techniques increasingly use learning techniques (machine learning) to produce models from data. Let's take image recognition that has made spectacular progress recently. Initially, to recognize a cat for example, humans did not give the algorithm the shape of a cat, but thousands of pictures. The algorithm is able to detect cat pictures where no previous information is given. Some techniques, especially deep learning, have the double characteristic of learning through data, yet not even being understood by those who have access to it or created it. Back to the cat, we access a matrix of coefficients, but no semantics giving the cat shape. This illegibility of models can be interpreted as an "*the end of theory*" (Cardon, 2018, p.7). The real strength of algorithms is they can produce efficient results, even though one cannot know how they produce them. By learning from the data, computational techniques can very well produce unexpected, curious, or shocking results without being anticipated by those who program them. Bizet's criticism of algorithms is therefore legitimate, but, in a way, he blames the algorithms for the shortcomings of their qualities.

Now we know that Big Data itself does not create value. Indeed, it is primarily a question of what Big Data has done with it. Without advanced technologies and techniques for data usage and analysis, it is not possible to produce value (Diez, 2016). Let us now look at the different sources of Big Data that exist and that allow huge volumes of data to be generated.

4. The sources of the Big Data

When we talk about the sources of the Big Data, we are essentially referring to the third "V" of Big Data, i.e. **Variety**. Having talked about Variety, we now need to delve deeper into what this variety is really made of. On the theoretical level, first, we can see that these data are divided into 3 big groups, as shown in Appendix 6. BrightPlanet (2020), calls the 3 categories of data type :

- **Structured data:** *"structured data refers to any data that resides in a fixed field within a record or file. This includes data contained in relational databases and spreadsheets"* (Beal, n.d, p.1). This means that we are talking here about data that has a high level of organization, that is easy to include in a database and that is easy to process, thanks to simple algorithms. It is with this type of data that we often work with in traditional data analysis, and they are mainly derived from standardized forms such as surveys or membership forms (to a political party, bank, university, etc.)
- **Semi-structured data:** this data does not fully fit into the formal structure of a relational database. But they do have labels or other markers that will eventually allow us to class them in the necessary fields.
- **Unstructured data:** Gartner defines: *" unstructured data as content that does not conform to a specific, pre-defined data model. It tends to be the human-generated and people-oriented*

content that does not fit neatly into database tables" (Gartner, 2013, p.5). This is by far the most represented type of data in the Big Data, with 80 to 90% depending on the different authors. Information is found here in a variety of different formats such as text, audio, video, web pages, long files, etc. (Gartner, 2013).

Now we will look at the different categories of data sources, which are not to be confused with data type categories, as explained below.

4.1 The Three Categories of Sources

In his article "*The real-time city? Big Data and smart urbanism*" (Kitchin, 2013, p.1), Kitchin explains that generally, Big Data sources can be divided into 3 groups, described below. We thought this categorization was relevant to our research question, because the Smart Cities he talks about are kinds of smaller-scale, data-driven government experiments. He says: "*The sources of Big Data can be broadly divided into three categories: directed, automated and volunteered*" (Kitchin 2013, p.4).

4.1.1 Directed Sources

Directed Data are defined as follows: "*Directed data is a type of data which is generated in offices by employees who are usually directed by managers. This is also the type of data which is generated in laboratories. In addition, this type of data is stored automatically and in a suitable format*" (Skourletopoulos et al, 2017, p.264). These are traditional monitoring technologies, which are directed to a person or place by an operator, for example passport control systems. They collect and verify passenger information which is stored in real-time databases. From these databases, new data can be generated such as passport photos, fingerprints, or retinal prints. This is also the case of electromagnetic or thermal environmental sensors, which allow real-time mobile 2 and 3D mapping (Kitchin, 2013).

4.1.2 Automated sources

Skourletopoulos et al. simply define automated sources as: "*data generated by automated devices*" (Skourletopoulos et al, 2017, p.264). However, they add that this type of sources should be divided into several subtypes: Data generated by automated surveillance, by digital devices, sensed data, scan data and data interaction (Skourletopoulos et al, 2017).

All of these types of data, referred to as automated, constitute the largest share of data generation. Moreover, it is interesting to note that many of these automated data originate from "connected" physical objects. This is closely related to what Kevin Ashton has called "*The Internet of Things (IoT)*", a concept

he launched in 1999 when he worked at Procter & Gamble (Lueth, 2014). Gartner, in his dictionary, defines it as follows: "*The Internet of Things (IoT) is the network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment*" (Gartner, 2020, p.1). These physical objects can take many forms, such as mobile devices, sensors that can sense the environment, such as geolocation, temperature or movement. Once these objects become capable of communicating with other objects or databases, they are referred to as connected objects. On a global scale, the network formed by these connected objects is called the Internet of Things (IoT) (Diez, 2016).

The Internet of Things is a complementary trend to the Big Data trend and will therefore also revolutionize the world. We can observe that today almost all objects are becoming connected (watches, cars, refrigerators, heating and lighting systems, clothing, etc.). This is why the Internet of Things is one of the most important sources for Big Data.

4.1.3 Volunteered sources

Kitchin provides the following definition of volunteered data sources: "*In contrast, volunteered data are gifted by users*" (Kitchin, 2014, p.4). Example, interactions on social networks, such as posting comments, observations and uploading photos to sites such as Facebook, Twitter and Instagram. They also come from data crowdsourcing, where users generate data and then transmit it to a common system, such as the generation of GPS tracks downloaded by the Waze application. Ultimately, this creates a common and open mapping system. Wikipedia is also a good example of a voluntary source. It is the users themselves who contribute to the content of the Web. Voluntary sources are therefore sources in which people contribute.

As said in the introduction of this section, the 3 different sources we have just mentioned therefore reflect the Variety dimension of Big Data. And this creates a major strength of the Big Data, which has to bring together this data from many different sources, from many different backgrounds. Then, we must cross them in order to bring out correlations that we could not discover otherwise (Diez, 2016).

4.2 Open Data

Another important source of Big Data is Open Data. Before addressing this topic, we can quote Gartner's definition: "*Open Data is information or content made freely available to use and redistribute, subject only to the requirement to attribute it to the source. The term also may be used more casually to describe any data that is shared outside the organization and beyond its original intended use, for example, with*

business partners, customers, or industry associations. Formally, data designated as "open" is subject to several conditions and licensing that can be found at opendefinition.org." (Gartner, 2020, p.1).

As explained by Peugeot, open data are often associated with government data and are made public by governments. The Open Data movement, i.e., the opening of data by governments, originated in the United States. But this phenomenon is quite unique in the history of our democracies. It is the first time that a political transformation has been adopted in such a short period of time by local governments, national governments and even international institutions. It even speaks of a "*flash political consensus*" (Peugeot, 2011, p.1).

In its report "*OpenData Barometer*", the World Wide Web Foundation explains that Open Data already covered 115 countries in 2017 (World Wide Web Foundation, 2017). Compared to the previous edition in 2016, this was already an increase of 25%. It is estimated that there are more than one million databases that are made available by governments worldwide. Nevertheless, it is still important to remember that governments are not the only source of Open Data. The Open Knowledge Foundation gives the following 3 criteria for Open Data (Open Knowledge Foundation, n.d):

1. **Availability and access:** data must be available in its entirety, i.e. data that is only partially published is not considered Open Data. They must also be available at a reasonable reproduction cost (e.g. downloadable from the Internet). Finally, they must be available in a convenient and editable form.
2. **Reuse and redistribution:** Open Data are provided in a way and under conditions that allow for their reuse and redistribution. This includes being able to integrate them with other databases. The data must therefore be readable by computers and other machines.
3. **Universal participation:** There can be no discrimination, whether between fields of activity, persons or groups of persons. An example is restrictions that would prevent the commercial re-use of data. This type of data would then not be fully open (Open Knowledge Foundation, n.d).

This is why McKinsey, in its 2013 report, referred to these data as "*liquid information*" (MGI, 2013: p.1). They also explain that Open Data is not to be confused with Big Data. Indeed, Open Data is one source of data among others, as is well illustrated in Appendix 7. Even open data from governments is only part of the Open Data. But we talk about it more because it is more relevant to our research topic, and particularly to part 3, where we're going to discuss the possibility of data-driven governance. After a review of the available data sources, in the next section, we will explain the main tools used to handle these data.

5. The tools of Big Data

Handling Big Data is certainly not easy. As stated in the previous sections, Big Data is such a large volume of data that it cannot be handled with traditional data management tools (EOD, 2020). Therefore, companies that have made inroads into Big Data have needed to have ultra-powerful IT tools at their disposal. Here we outline the key technologies that have supported and continue to support the Big Data industry. As this does not influence the rest of our work, and as it is long and very complex to define, we will not differentiate here between application tools, computing tools and storage tools.

5.1 The Apache Hadoop Platform

Currently, Hadoop is the main platform for Big Data, used for storage and processing; it is used by many companies to carry out Big Data projects. *"The Apache Hadoop projects develops open-source software for reliable, scalable, distributed computing. The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a high-available service on top of a cluster of computers, each of which may be prone to failures"* (Apache Software Foundation, 2020, p.1). Based on Java, the framework is part the Apache Software Foundation.

This platform allows, for example, thanks to its MapReduce framework, to move the processing software directly to the data, rather than having to move the data to a network for processing (Bastien, 2018). Another strength of the platform is to allow fast data processing, thanks to the node system. This system provides protection against hardware failures. If a node fails, tasks are directly redirected to other nodes to ensure that distributed computing does not fail. In addition, multiple copies of all data are stored automatically. The platform also allows unlimited storage of data before processing. It is possible to store any type of data and decide later how to use it. In the same way, this platform makes it possible to archive a lot of data. But going into details would require extensive work not necessary here.

Nevertheless, it is still interesting to translate the importance of Hadoop by a few figures, as well as to give examples of firms that use it to understand how important this tool is. While Hadoop was valued at \$7.69 billion in 2015, it is estimated that its value will reach \$87 billion in 2022 (Fontaine, 2018). As illustrated in Appendix 8, Hadoop's growth is impressive with 51% annually between 2016 and 2021. We can also see how important Hadoop is by looking at clients: Facebook uses it to suggest people you might know, LinkedIn offers jobs that might interest you, Netflix and eBay recommend content (Bastien, 2018).

5.2 Non-relational databases

Unlike traditional relational databases, which present data according to a system of relationships in a two-dimensional table and use the SQL (Structured Query Language) programming language, non-relational databases, renamed "NoSQL" in 2009 (Not Only SQL), are databases that make it possible to face two major challenges of Big Data: its Volume and its Variety. MongoDB gives the following definition: *"NoSQL databases (aka "Not Only SQL") are non-tabular, and store data differently than relational tables. NoSQL databases come in a variety of types based on their data mode. The main types are document, key-value, wide-column and graph. They provide flexible schemas and scale easily with large amounts of data and high user loads"* (MongoDB, 2020, p.1). Appendix 9 provides us a good representation of the difference between an SQL and a NoSQL database.

As shown in Appendix 10, it was the declining cost of data storage that allowed the emergence of non-relational databases in the early 2000s. Indeed, computer technology has developed so rapidly that the power of servers has multiplied. This allowed a drop in the cost of data storage, and computer scientists developed much more flexible databases where it was possible to store all types of data. NoSQL data models allow nesting of related data in a single data structure (MongoDB, 2020). Once built, most of these databases are then incorporated into websites such as Google, Amazon, and Facebook.

5.3 Cloud Computing

Cloud Computing is defined by IBM as follows: *"Cloud computing, sometimes referred to simply as 'cloud', is the use of computing resources - servers, database management, data storage, networking, software application, and special capabilities such as blockchain and artificial intelligence (AI) - over the internet, as opposed to owning and operating those resources yourself, on premises"* (IBM, 2020, p.1). This technology therefore allows for large data storage capacities but also for the provision of computing resources, available online, at fixed prices and with high levels of flexibility and reliability. Thanks to Cloud Computing, companies active in the Big Data industry have an infinite quantity of quality IT resources at their disposal in just a few clicks (Diez, 2016).

5.4 Data mining systems

They constitute a set of methods, algorithms and techniques that allow the use of Big Data by providing effective solutions to the challenges of Volume, Variety and Speed. PageRank and MapReduce (which belongs to Apache Hadoop platform) are computer models of this type used by Google and are the symbols of an important advance in the field of usage and analysis of large volumes of data. Data mining is therefore a process, which allows you to identify patterns and that subsequently allow you to judge certain events or support decisions (IBM, 2012). The complete process is detailed in Appendix 11.

6. From Big Brother to Big Data

In his novel "1984" (published in 1949), George Orwell already anticipated the fact that we were moving towards a surveillance society. The main figure in the novel, Big Brother, became a metaphorical figure of the police and totalitarian regime of the surveillance society and the reduction of freedoms. Are we today, having gone from Big Brother to Big Data? (Law, 2019). But in their book, Mayer-Schönberger and Cukier evoke the irony that there were more than 30 surveillance cameras within a radius of 200 meters around George Orwell's apartment when he wrote the book "1984" (Mayer-Schönberger and Cukier, 2014).

In their book, they discuss the risk and dangers of Big Data revolution. Never before in history has so much data been collected and stored about all of us. We are constantly watched, whether we use credit cards to pay, phones to communicate, or social security numbers for identity. Even before the advent of the Internet, companies like Experian or Acxiom were collecting and tabulating personal information on hundreds of millions of people around the world, for resale, of course. The Internet has made tracking easier and, above all, cheaper. But government agencies, like the CIA, are not the only ones spying on us. Amazon monitors our shopping preferences, Google our browsing habits, while Twitter keeps us entertained. Facebook can capture all this information at the same time, in addition to knowing our social relationships in detail. Mobile operators not only know who we are talking to, but also who is close to us, thanks to geolocation for example (Mayer-Schönberger and Cukier, 2014).

While Big Data promises a bright future for those who analyze it, from the citizens' point of view, all signs seem to point to an increase in the collection, storage, and re-use of our personal data. As long as storage costs continue to fall and analytical tools become increasingly powerful, the size and scale of data collections are likely to continue to grow by leaps and bounds. Mayer-Schönberger and Cukier therefore ask the question: *"If the Internet age has threatened our privacy, is the Big Data putting it at even greater risk? Is this the dark side of Big Data?"* (Mayer-Schönberger and Cukier, 2014, p.151).

Apart from the violation of individual freedom, which Mayer-Schönberger and Cukier see as a risk in Big Data that endangers our privacy, they mention two other important risks:

- The tendency to punish and judge an individual before he has even acted, based on the mere presumption of what he could have done ("Fichés S" in France). Since most attacks are carried out by people who are already on the S files, including potential terrorists, would it be necessary to punish them before they have even acted?
- The dictatorship of data for those who exploit it. The risk is to consider only the promises offered by Big Data without considering their limits. This is the question that Bizet asks in his article *"Le Big Data est-il l'Eldorado des gouvernements ouverts?"* (Bizet, 2018, p.1).

Finally, it is necessary to specify that companies are not the only ones targeted. Cardon and Bizet point out that States, too, are likely to commit abuses. As for the future of data-driven governance, Mr. Faulkner is very clear. He believes that this trend will become more pronounced. The next generation will, in his opinion, be increasingly confronted with the exploitation of their data. He believes that in everyday life, we will find recommendation systems that will only be able to function once we capture personal data: *"So I think we're just at the beginning of this process, that it's going to be even more pronounced in everyday life"*. Mr. Dombret has a less clear-cut opinion on the matter. It's certain that the use of data is not going to disappear, but he believes that we are moving towards *"cruising speed"*.

To conclude this section, we will, following Kant's logic, explain that if we have not managed to identify two distinct camps, we have not done the job to the end. A good example is sports: there is no sport with three-team matches, or else it creates alliances. The two camps we have established are, on one hand, the potential victims of Big Data, and on the other, the potential "criminals". If this is a strong word, we will see next (cf. Part 2: Big Data in the election sector) that it is not always exaggerated. Potential victims are ordinary citizens, the consumers, whereas the potential criminals seem to be either big companies such as GAFAM or States. Mayer-Schönberger and Cukier make an alarming assessment of the Big Data revolution, which worries us about the potential threat to our society and the most fundamental human rights.

Are these fears realistic or disproportionate? Is humanity at risk? What limits should be imposed on Big Data to prevent it from degenerating? Is it possible to control Big Data revolution? These are important questions that arise and will also be addressed in our analysis of the impact of Big Data on the electoral sector (cf. Part 2) as well as in the area of public governance (cf. Part 3).

7. Our own definition of Big Data

Based on all that has just been said, we offer a synthetic view:

- on one hand, by its **Volume, Variety and Velocity**, it has enormous potential for value creation, both for the benefit of the organization and the individual.
- on the other hand, it draws major challenges in terms of data **Veracity, Visualization, Viscosity, Variability, Volatility, and Vulnerability**.

Big Data analytics, i.e. the usage and analysis of this data set by algorithms, make it possible to generate knowledge and techniques that will change the way the world is today. But the question is: how? The future will tell us whether it is a question of creating countless possibilities to improve daily life and create value for business, or whether it is in fact a way of widening the gap between the silent majority and the dominant minority. The new risks, if left unchecked, could endanger the fundamental civilization and democracy principles. Let us move on to a specific sector that has amplified the importance of Big Data in recent years!

PART 2: The 4 major trends identified from Big Data in the Electoral sector

In this second part, we examine the Big Data techniques observed in election campaigns. We will see that these techniques can be summarized by four main trends. This second part is a sector study, which is based on all the techniques, definitions and sources of Big Data developed in Part 1. Then, in Part 3, these four major trends will be applied to other sectors. Sectors where the government could use them to apply Data-driven governance.

First, we will explain where electoral Big Data arose, a trend that came from the United States, and spreading more and more in Europe. It will also allow us to highlight the differences between the use of data in the US and in Europe. Next, we will focus on the electoral Big Data in France. Why the election sector? Because the electoral sector is in a way the first step to access to power, and as we discuss the possibility of Data-driven Governance in Part 3, we considered it relevant to study this sector. The area of elections is the necessary first step to understand government, and governance, studied thereafter, is a logical continuation. Why in France? One, because literature on the topic is in-depth and has received more media coverage. Two, because France is a neighboring country to Belgium, quite close from a socio-demographic and cultural point of view. Finally, because the same laws concerning data protection are in force (RGPD in particular). The four major trends identified for electoral Big Data in France will therefore be transposable to Data-driven governance in Belgium.

The three main axes of this part will be the origin of electoral Big Data, i.e. a focus on the US. Hence, a focus on France. Then, we will see that there are two very distinct parts to the electoral Big Data. On one hand, political parties and elections, on the other, commercial actors. These market actors will be very interesting to study as they are sometimes unknown to the public but very important in real life. Afterwards, we will study, one by one, the different applications of Big Data in politics, especially those used during election campaigns. Finally, we will try to establish a general process of the electoral Big Data sector and explain the future prospects of electoral Big Data in France. To conclude, we will summarize everything that has been written by four major trends, which we will apply to each use of Big Data in the electoral sector via a summary table.

1. The Origin of Electoral Big Data: American Campaigns

1.1 The Beginnings of Electoral Big Data in the United States

As with anything else, it is always interesting to go back to the origin of a phenomenon to better understand it. This is what we will do in this first section, where we will focus on the first presidential campaigns to have used Big Data during elections: Barack Obama's campaigns (Pène, 2013). If Trump's

use of Big Data was highly publicized, notably through the “*Cambridge Analytica*” case, it is indeed Obama who was the first to benefit from Big Data.

The trend towards the specialization of political auxiliaries is particularly evident in the United States, especially for fundraising activities. The profession of political communicator and the figure of the “campaigner” are presented as being of American origin. We find this American origin in the discourse of European web-strategies that consider Obama's campaign as the starting point for data analytics in the political sector. This American origin is supported by the politician Kreiss, who proposes a socio-historical analysis of the rise in the political use of data in the United States from 2004 onwards (Kreiss, 2012). Kreiss stresses that it is necessary to go back to 2004 and Howard Dean's campaign to win the Democratic Party leadership in order to capture the construction of a national database in the US. Indeed, Dean undertook two important projects for this election: the creation of a national electoral register and the implementation of a new online electoral platform. The goal was to centralize the databases held locally by the Democratic Party Federations. Once this database was set up, it was necessary to connect it with all the federations in order to centralize the data held by local entities (Théviot, 2019).

The innovations launched by Dean were thus carried and developed by his teams between 2004 and 2007, then put at the service of candidate Obama in 2008, with attention paid to rigorous data management. By setting up in 2008 an email database, called Catalist, which contained information on nearly 220 million Americans, with up to 600 pieces of information per person (Terra Nova, 2009), Obama's digital team marked a turning point in the use of the web in politics and fulfilled Dean's promise: to massively recruit and organize supporters through the internet to send them, in a coordinated manner, to militate on the ground (Kreiss, 2012). To take part in a traditional campaign, one had to go to the section attached to one's home or contact the campaign headquarters. It was a rather cumbersome process and many people gave up. Obama's campaign strategy reverses the logic: you don't have to wait for supporters to come to the campaign, you have to go to the supporters. We can summarize the campaign strategy in one sentence: “go where the people are”. Digital networks such as BlackPlanet have thus given a major place to the recruitment of volunteers (Pène, 2013). Imbued with the academic literature on the subject, Obama's web campaign was able to combine mass targeted emailing's and investments in social networks to mobilize the Democratic camp and the candidate's potential voters, but also to the former abstentionists (Théviot, 2019).

In 2008, during Obama's campaign, 1.2 million people went out to meet voters and present the Democratic program to try and convince abstentionists to vote, through a massive door-to-door campaign (Pène, 2013). This operation, carried out throughout the country, required a strong coordination of actions and optimal coverage of the territory. These conditions were met by the creation of the ‘my-barackobama.com’ platform. This platform did centralize the organization of the door-to-

door campaign with creation of two million profiles, 35,000 groups of volunteers, 400,000 blogposts and 200,000 events organized (Théviot, 2019). The web capacities (communication and coordination) were put at good use in an articulation between online and offline. This is how Obama achieved the Orwellian dream of any American candidate: to put the entire country on file. The technique was based on micro-targeting: it involves consolidating as many existing databases as possible (electoral, commercial, political databases) in order to obtain individual data on all voters (Pène, 2013).

Following this success, campaign teams around the world, particularly in France, were inspired by Obama's campaign techniques. Claiming American influence is part of the staging of a campaign today. The strategists of Obama's web team, as we will see in section 3 (cf. section 3.1.1 The emergence of the market under US influence), came to advise the web teams of French politicians. Having recourse to campaigners involved in the Obama campaign is now a resource legitimizing digital campaigns. It is thus a question of staging media interest in American expertise, which is at the cutting edge of innovation in the field of digital communication. This appears to be a guarantee of success and innovation: "*the foreigner is above all a first-rate resource when it comes to looking for models or recipes for public action that are supposedly effective*" (Frinault and Le Bart, 2009, p.629).

1.2 Major differences between campaigns in France and the United States

Although digital strategists explicitly refer to the American model, it is not an exact copy. Indeed, one must include the particularities of the French and European contexts: the legislative framework, the way of campaigning, the partisan cultures, the budget allocated to digital and to campaigns, the internal routines of political parties, etc. (Théviot, 2016). We noted three major differences between a presidential election in France and in the United States.

1.2.1 The difference in financial resources

The first difference in scale is primarily economic (Théviot, 2019). The use of data in politics remains very dependent on budgets and the number of people able to implement its technical development. Even if the budget allocated to digital and more recently to the use of data increases with each national campaign in France, it is still much lower than that allocated in the United States. According to the NGO Center for Responsive Politics (CRP), which specializes in the transparency of political financing, the American election campaign will have cost nearly 7 billion dollars. Trump's campaign budget is estimated to have been around 600 million dollars. This is far more than what French political parties can afford. Indeed, election expenses are capped at 16 million euros for candidates in the first round and 22 million euros for candidates in the second round of the presidential election (Lambert, 2012).

1.2.2 The difference in legislation

The second major difference lies in the French and European legislative framework (Théviot, 2019). At a conference in 2009, Dean conveyed the secret of campaign strategy: targeting your voter by "*reinvesting abandoned electoral targets and targeting new ones*". He thus deciphered how democrats, through opinion research and extensive database analysis, adjusted campaign efforts to those who they felt were in line with values they wanted to defend. This work made them focus on middle classes and ethnic communities, but also on categories acquired from the Republicans, such as Evangelical Christians. His discourse considered national disparities: it seems difficult to file by ethnicity and religion in France. In most European countries, individual data protection is much more important than in the US, much more relaxed with data. They see it as something transactional, whereas in Europe, we have a much closer relationship with our data in the sense that we feel that we own it. That is why people in Europe want all data to be anonymized, which is not the case in the United States. This is changing with the new European RGPD legalization, but there is still a big gap with the US.

For example, in France, the CNIL has created an election observatory and put in place legal barriers that prohibit the acquisition and cross-checking of databases without the prior consent of the people listed in the files (Rallet and Rochelandet, 2015). In 2016, the CNIL carried out, on its website, a severe diagnosis of the Knockin (knock on the door) application used in particular by Sarkozy's campaign team, stressing data protection and massive collection of data from social networks, with information void for the concerned (CNIL, 2016). In France, you do not have the right to buy a commercial database from Carrefour, then take the personal data from Facebook, and cross-check it with the data available to create a database. To avoid using personal data, some agencies specializing in electoral strategy collect data, not on an individual level, but on the scale of a neighborhood or a polling station in order to map the territory and then sell this database to political parties. In France, for example, the results per polling station of an increasing number of past elections are made available to the public by the Ministry of the Interior. The online inclusion of public data, or "Open Data", by the Etalab government service, thus contributes to the qualification of the databases.

1.2.3 Differences in organizational methods

A third difference in scale is visible. The low weight of partisan organization in the United States facilitates technical innovations without internal resistance. In France, political data professionals confront the mistrust of partisan executives, traditional ways of doing things and competitive games with the party's communication professionals. Qualified database targeting was not, before 2012, really taken seriously by political leaders, who saw it as an additional gimmick. Campaign managers questioned these practices and thought they were "overvalued" (Pène, 2013).

As a result, the databases of political parties are quite thin compared to the long-term work done in the United States. But even timid compared to the United States, the arrival of these new technologies in the French political space was not neutral. The 2017 presidential election was also disruptive in terms of web strategies for electoral mobilization (Théviot, 2019). The election was strongly oriented towards digital optimization. This technical test bed allowed independent data workers as well as specialized analytical firms, both French and foreign, to make themselves even more indispensable to parties and candidates (Pène, 2013). Finally, despite scandals such as the Cambridge Analytica scandal, the signals of modernity seemed to outweigh the risks of developing an unprecedented form of algorithmic governmentality, thus underlining the now major role played by this new predictive martingale to beat the campaign and win the elections.

2. The interaction between political parties and the market players of the electoral Big Data

This section introduces the next two. If we talk about electoral Big Data, we often think of political parties and candidates (Trump, Macron, Obama), but we do not often talk about the "hidden" actors of electoral Big Data, which are private companies. If there are the political parties, which campaign, appear in the major media, and their candidates who are potentially our future leaders, the commercial actors are companies that political parties use to improve their use of Big Data, refine their databases and improve their digital communication (Ehrhard et al., 2019).

The link between political parties and electoral market players is, among other things, a commercial one. Companies such as NationBuilder, Fédéravox, LiegeyPonsMuller and others, which we will study in the following section, are paid by political parties to manage their digital campaigns, notably by offering the technological services of Elections 2.0 (targeting priority areas, data analysis, etc.). This is a way of subcontracting part of the campaign to private actors (Théviot, 2019).

In addition to the rather technical Big Data services, many companies offer other very varied and less complex services (database creation, CRM, etc.). In fact, the majority do not act as election 2.0 specialists, neither as experts in communication or digital organization. The core technical activities (mailing, activist databases, communication within campaign teams, feedback, etc.) is composed of simple functionalities, present in most business management software, with no specificity to the electoral or political field (Ehrhard et al, 2019). Organizations rely in part on recent technological innovations but include more traditional management aspects. This dynamic of services offered responds to the need to digitize the entire campaign organization.

The link between clients, i.e. political parties, and market players is therefore twofold. On one hand, it is technical, quantity and type of data integrated into software. Indeed, most political parties are unable

to make full use of the databases. On the other hand, the link is strategic, because finesse is not key, companies must therefore find other ways to differentiate themselves: more consulting, more advanced software edition or a combination of the two.

Now that we have made a clear distinction between the two major players in election, political parties and market players, we shall detail role, actions and strategies in the electoral Big Data.

3. The electoral Big Data market players

In this section 3, we will focus on the commercial actors of the electoral Big Data. This will allow us to make the link with section 4: Big Data, political parties and elections. Indeed, a new market has emerged in recent years. While these companies are increasingly being put forward on the media scene, we do not know much about them after all. On the one hand, the objective of this section is to develop empirical knowledge on this subject, and on the other hand, to pursue some questions related to the study of data workers. Our study of this section will be based essentially on the work of Ehrhard et al. (2019), and on their article "*A la conquête des élections 2.0*" (Ehrhard et al, 2019, p.1). This will allow us to answer the following questions: who are these actors? What do they do? How do their strategies influence a market with strong economic constraints? In addition, it will allow us to know who is behind the Big Data tools used in elections, developed in section 4. Objective is to target players present in France.

3.1 Who are the commercial actors of Elections 2.0?

We can attribute two characteristics to this new group of professionals. On one hand, they are "new", having been in business for maximum ten years. The oldest are foreign companies such as NationBuilder, created in 2009. On the other hand, this is a small group of professionals. On the French market, there are less than ten identifiable players (Ehrhard et al, 2019).

3.1.1 The emergence of the market under US influence

This concomitance can be explained by the fact that players created a market that did not previously exist: they emerged together. They developed a need for data-driven modernization and rationalization of electoral campaigns, for which they offer services to political actors. Apart from foreigners, these professionals have homogeneous background and inter-knowledge (Théviot, 2019). They studied in the US (Harvard) and participated in American electoral campaigns, for example the "three Bostonians" who created LMP (LiegeyPonsMuller). These companies were founded with entrepreneurial logic. Similarity of individual backgrounds and shared influences does not prevent strong differences at the level of the companies.

3.1.2 Heterogeneity of entrepreneurial forms and logics

The first difference we found between the companies is their size. Indeed, it varies from very small start-ups like Fédéravox (3 employees), through small SMEs (LMP, about 20 employees) to larger multinational companies like NationBuilder (135 employees). This initial diversity in size is then reflected in several sociological aspects, but also in the services offered and the strategies of the players. Thus, NationBuilder, the most developed and structured internal organization, has more than thirty business specialties related to data, digital, but also commercial and territorial development. For example, at LMP, there are "data scientists", "developers" and a sales team. This distinction is therefore functional (Ehrhard et al, 2019).

But a more interesting feature is the presence of "political specialists" in the electoral Big Data companies. For example, at DigitaleBox, there are only coders and developers. This says a lot about the relationship that companies have with politics. This absence exposes the primacy of technology over politics in the field of poltech (Cardon, 2018). It also shows that the mathematical and statistical exploitation of data is more important than the nature of the data. In other words, there is no mathematical particularity in the mathematical exploitation of political data. This corresponds to the paradigm of Elections 2.0, according to which data allows for the rationalization of the way politics is done, with more efficient tools that allow us to move away from intuition and illusion. Nevertheless, if the political dimension is absent internally, it can be present externally with the use of external partnerships. This is for example the policy of NationBuilder, which focuses on technology, but which has an ecosystem around its platform enabling it to offer policy advice to its clients. The company then calls on external "policy experts". The political dimension is therefore outsourced to external service providers but via the NationBuilder platform (Ehrhard et al, 2019). Other companies are in the opposite logic, such as Fédéravox, and subcontract the technology and focus on the political aspect.

From individual expert to international company, start-ups or older company, there is no single type of player in the political technology sector in France. Despite homogeneity of backgrounds in French players, study shows a strong heterogeneity of the workforce within companies and disparity between entrepreneurial logic (Théviot, 2019). Differences are found in the services offered by the actors as well as in the way of doing political technology in the field of elections in France.

3.2 What do market actors do in elections 2.0?

To study the services offered by the commercial actors of electoral Big Data, Ehrhard et al. have established a typology of the actors of elections 2.0 in France, presented in the table in Appendix 12. Firms are classified according to their degree of specialization (how many different types of general

services do they offer) and the different parameters that characterize them. This allowed them to identify 3 types of actors: actors specialized in a field of expertise, integrated actors and relay actors. The table indicates the existence of a multiplicity of means (parameters), declined according to the actors, and series in a duality of services. Precisely, the activity of actors in elections 2.0 is divided into four classes of general parameters that cover the scope of services offered including from a technical point of view: software services, support services, type of clients, data used (Ehrhard et al, 2019). We'll detail the parameters used before presenting the activities of the actors through the groups to which they belong.

3.2.1 A multiplicity of means for a duality of services

There are two types of services: software and support of customers with political advice. The first consists of offering software based on data and requires computer and/or statistical skills. Several types of software exist, based on defined targets with different levels of granularity (Erhard et al, 2019). These result from different strategies and constraints; targeting can be both nominative and immaterial. This is called Customer Relationship Management (CRM); for a political actor, it is a software that manages the members of its databases (activists, potential voters, etc.). It offers customers, i.e. political parties, the possibility to directly contact the members of its database to mobilize them, for example via e-mail. Depending sophistication and, above all, data, it can cross-check different types of information to refine its mobilization according to its strategic choices (GPS coordinates, client preferences, phone numbers, socio-demographic data, etc.). In general, the software also provides a visualization of all the data on a map. This facilitates the organization of field campaigns, which is often referred to as electoral targeting (Théviot, 2019). Even when the data are not sufficiently accurate, some data can be indirectly retrieved via a prediction model. There is also a difference between companies that publish their own software (the software is then designed by the company itself) and companies that do not publish it.

The second type of service is the provision of policy advice based on a knowledge of the digital tools previously proposed, with "software training". This classic service is modernized in that the consultation is based on the interpretation of the software it offers its client. Finally, some clients also offer training for field actors to set up campaign actions in a more efficient way, such as door-to-door (e.g. LMP). These two types of offer (software publishing and consulting) are the determining variables to distinguish the three types of actors (Ehrhard et al, 2019).

3.2.2 The Three Types of Actors in Elections 2.0

We identified three types of actors in elections 2.0: actors specialized in a particular field (consulting, software publishing, data science), integrated actors (software publishing and consulting) and relay actors who are located at a junction point between the other two (Ehrhard et al, 2019).

3.2.2.1 Specialized actors

Specialization concerns: digital consultant, data scientist and software publisher (DigitaleBox, Ecanvasser). For example, DigitaleBox's job is really to develop the best possible software at the best price, to reach the largest possible number of elected officials. The specialization does not exclude the possibility of an accompanying after-sales service. But this service must nevertheless be distinguished from a political consulting service, which is a fully-fledged service of interpreting the data in the software to build a campaign from this digital tool.

The actor specialized in software does not provide this service, just as the consultant specialized in campaign construction does not provide software (Théviot, 2019). The functionalities of the different tools proposed then obviously depend on choices, leading to new distinctions. Thus, DigitaleBox offers its tool to many partners who sell services using it mainly for consulting. Conversely, Ecanvasser is a partner of the NationBuilder network. The data used comes from the customer through NationBuilder or directly from the customer himself. The relationship in the value chain of the finished product is therefore different. One is self-sufficient as software, while the other interfaces to multiple tools in the platform created by NationBuilder.

3.2.2.2 The integrated actors

The second type are integrated actors who combine digital campaign tool and electoral campaign advice. Integration enables them to provide a self-sufficient (standalone) service to their clients (Ehrhard et al, 2019). This is the case of LMP, Fédéravox and Spallian, which offer software with more or less similar functionalities relating, in particular, to the elaboration of a map (more or less interactive), the presence or not of a CRM and a targeting functionality.

However, their services also include a strategic aspect in the development of the electoral campaign based on software's data. The software is used as a decision support tool and they act as its own interpreters by providing advice (Théviot, 2019). However, not all actors operate in the same way. Several types software and targeting results exist. For example, LMP organizes the campaign on the ground, Fédéravox's provides an argument to the voters, and Spallian shows the candidates what the voters expect from them.

3.2.2.3 The relay actors

Only one actor (NationBuilder), a specialist in software, proposing to clients, as Cowan: "*We are not experts in campaign strategy, we are experts in how NationBuilder works and how your campaign infrastructure is created. So we just provide the technology and training around the technology of course*" (Ehrhard et al., 2019, p.115).

So, it is not a political consulting service either, but more of a consulting service to organize a community. It does edit site creation but does not interpret the data nor creates a campaign. For this purpose, NationBuilder has a particular profile of specialization to relay to its partners.

3.3 The partisan dimension of market players

Key element of the strategy, is it exclusively partisan? Do market actors have an unofficial partisan approach and/or affiliation? In France, the diversity of clienteles was initially a notable difference between entrepreneurial strategies, with, on one hand, a limitation to political actors (DigitaleBox, Fédéravox, LMP) and, on the other, an opening towards other clienteles (NationBuilder, Ecanvasser, Spallian). Not the nationality but rather experience and seniority. NationBuilder has been in existence for many years, but companies such as LMP, DigitaleBox or Fédéravox emerged at the same time in the electoral Big Data market (Ehrhard et al, 2019).

Initially, the question of the partisan dimension is therefore relatively clear, despite a discourse of the actors sometimes out of step with reality. The French market is split between market players whose positioning is notoriously "right-wing", "left-wing" or "non-partisan", regardless of statements denying any partisan dimension. When it exists, the partisan dimension is chosen because of ideological proximity and/or the political commitment of the leaders (Théviot, 2019). In France, for example, a clear distinction can be made between Fédéravox, which has a predominantly right-wing clientele, and LMP, which has a predominantly left-wing clientele. But we will see in section 6 of this part, which deals with prospects, that more and more companies are trying to get rid of this partisan label, which is a real handicap for their development.

4. Big Data, political parties, and elections

In this section, we detail the uses of Big Data in politics, specifically techniques and models used during political campaigns in France. This will allow us, in section 7, to identify 4 major trends in the use of Big Data, which we will then apply in other sectors (cf. Part 3: Data Driven Governance). We will begin by explaining how campaign-specific data is collected and how electoral files are created. Then we will see how this data is used during campaigns, how it can improve or completely renew certain campaign techniques.

4.1 Campaign team data

4.1.1 Data collection

In terms of data collection, campaign teams generally rely on three sources: socio-demographic data from the population, data from opinion poll studies and data from voting histories (Théviot, 2019). Other data from the aspiration of digital traces left online can also be captured to enrich the database.

In France, as early as 2007, political parties began collecting data. In the beginning, this consisted essentially of collecting as many emails as possible, without knowing how to use them effectively (Le Monde, 2012). These emails were collected to send a maximum number of partisan messages, in a very imprecise manner. Then, political parties started buying keywords, in order to increase the visibility of the campaign site and the candidate, and thus recover their email address again. Another way to collect data in 2012 for the Socialist Party was the primaries (the UMP did not organize primaries at that time). More than 50% of people left their email addresses in order to be informed about the campaign's outcome (Théviot, 2019). But one problem that arose, particularly in the PS, was the aggregation of these databases on a national scale. Regarded as a war treasure, these databases never merged perfectly because of internal struggles and different currents within the party itself. It was difficult to provide yesterday's opponent with the valuable data collected during the primary.

Nevertheless, this operation of massive data collection made Nicolas Sarkozy's UMP, which qualified this operation as "voter registration", jealous. Indeed, in the absence of primary, the candidate could only rely on social networks and his campaign site to collect data. Another key element in the collection of databases for electoral campaigns is temporality. Database building, presence on social networks, visibility, audience, traffic are long term phenomena. Candidates who launch their presidential campaigns rather late therefore find themselves at a disadvantage, which was the case for Sarkozy. To make up for this delay, the UMP bought commercial databases and used specialized agencies. This is therefore part of an entrepreneurial rather than political approach. Database purchases and mass mailings resulted in strong criticism from Internet users who have felt "spammed" (Europe1, 2012).

Finally, another means used by local officials to collect data was field operations. Politicians encourage their activists to collect as much data as possible on the people present during traditional field operations, particularly during meetings, towing or door-to-door visits to increase their databases. According to some politicians, this is even the surest way to obtain quality data, as people who are solicited afterwards are much less likely to feel spammed.

However, it is essential to note that collection of such data is only part of data analytics modeling if accompanied by database qualification. Indeed, files containing only e-mail addresses are not sufficient to develop electoral Big Data strategies. It is then necessary to try to aggregate other data, such as place of residence, centers of interest. Data science requires very specific data that is not limited to an e-mail address. Since 2012, the political parties have better understood the importance of working on their databases over time so as not to start from scratch at each election. In this objective of database efficiency, the exploitation of data is proving to be at the heart of electoral strategies (Théviot, 2019).

From a media point of view, talking about the work on databases, the number of emails collected, on the click rate, on the transformations into databases and others, is not very sexy. Beyond the difficulty of a media staging, this poor communication around the use of databases is also explained by a mistrust of the campaign teams, who do not want to be blamed for the use of personal data. This willingness of campaign teams not to mention, in their speeches, the work done on databases also explains, in part, the lack of academic work on the use of data in politics (Ehrhard et al, 2019).

4.1.2 The creation of electoral software in France

The Electoral Big Data market players create software. But here, we will be more interested in the software itself and the source of its data, as well as its use during campaigns, than in the companies that make it. Moving from a form of amateurism in 2012, French electoral Big Data experts have since become well structured. Each agency offers different services and software, while understanding that their strength would be to ride the wave of algorithmic innovation and the staging of a group of skills labelled Data scientist (Ehrhard et al., 2019). As we have seen, two types of software can be distinguished, with a focus on politics on the one hand (carried by LMP, Fédéravox) and technology on the other (with DigitaleBox, NationBuilder) (Théviot, 2019).

For the first type of software, the two main sources of data are, on the one hand, the State, which discloses electoral data from recent years in Open Data, and, on the other, the Institut National de la Statistique et des Études Économiques (INSEE), which provides socio-demographic data (INSEE, n.d.). All these open access data are then cross-referenced and compiled in software programs. A major challenge is to constantly update these data collected in open data to improve targeting. As time goes by and the data increase, the software becomes more and more accurate. This really reflects the Velocity aspect of Big Data; one should never miss an update and be as reactive as possible. This software made available by service providers (market players) to political parties can then be supplemented by other, more individual types of data collected by the campaign teams themselves (Théviot, 2019). The data of sympathizers who have voluntarily agreed to provide their contact information, for example on the candidate's campaign site, are then added to the software, which makes it possible to refine and increase

the efficiency of the algorithms. The membership files, or those coming from door-to-door for example, are thus complementary data to the open-data. All of this allows, in the end, to map (as we will explain in detail in section 4.2.1) the country in order to better target and rationalize activism.

The second type of software is more akin to a multifunctional platform that can automatically organize and update supporter data. The idea is to get to know and manage one's community well in order to start from this very involved core group and to reach out to other voters through a snowball effect. This approach corresponds to the one proposed by NationBuilder and DigitaleBox. This software is based on an extremely sophisticated system of algorithms, improved and built over the years and based on American strategies, as discussed in section 3.1.1 (Ehrhard et al, 2019). Unlike the other type of software, this software does not provide any data and does not look for data in the open-data. This may seem trivial, but it is a completely different approach. These are "all-in-one" software programs that can orchestrate an election campaign in all aspects (websites, organizing offline and online, profiling different voters according to the degree of support, etc.). If we were to sum up the objective of this second type of software in one sentence, it is to federate a community.

The idea is to make politicians aware that they have a lot of data and that it is interesting to cross-check, clean and organize it to make their campaign more effective in many aspects (not only digital communication, but also calling for donations, organizing activism, etc.). What they do with their clients is really an aggregation of all types of data (Théviot, 2019). While they will not provide the data themselves, they can, however, ask political parties for some of the valuable data they have in their possession without them using it properly. For example, data from candidate files can be aggregated with data from electoral lists. These lists provide a lot of data such as: gender, date of birth, identification of abstainers, etc. It is an essential source for software (Girard, 2017).

To conclude this section, we can now see that these two types of software are not opposed but complementary. For example, during the 2017 election, François Fillon used both Fédéravox and NationBuilder. Contrary to popular belief, we can see that these software do not quite seek to predict the outcome of an election, but rather to optimize the campaign. Even if there are techniques that can predict voter behavior, this is not the essence of what they do. The system is to understand how voters behave, how they have behaved in previous years and how in the territory, there are opportunities, priority areas, variations in electoral behavior in certain socio-professional categories, to do what marketers will call targeting (Théviot, 2019).

4.2 The use of data during electoral campaigns

We will now look at concrete uses of data during election campaigns. To study this, we base ourselves mainly on the 2017 presidential campaign in France: professionalization, standardization of data work within political parties, coupled with the construction of a new professional market outside political parties, associated with an politics awareness of strategic stakes. These are the factors that make us particularly interested in this campaign.

4.2.1 Mapping

Cartography is a software program that cross-references electoral results data with INSEE socio-demographic data to obtain a very precise map, polling station by polling station. It is an approach that proves to be strategic for targeting territorial areas of voters (LesEchos, 2017). It does not consider individual data, but cross-referencing a maximum of data, it allows experts to organize the campaign in the most targeted way (Théviot, 2019).

In concrete terms, France is divided into 60,000 small squares, each of which corresponds to a polling station. For each of these polling stations, data workers have in their possession both the election results and the INSEE census. All these data were integrated into the software as they were available. Even if they are not allowed to have individual data, it still allows them to obtain aggregated data on a fairly fine scale (Priol, 2018). Over time, with the increase in the amount of data, especially published by the government itself in the form of open-data, this mapping is becoming more and more accurate.

Ultimately, this allows a candidate's campaign team to superimpose several types of data, neighborhood by neighborhood, and to answer a number of questions: Who did the inhabitants vote for in the last elections? How many abstained? What is their average age? Are we dealing with young families or retirees? And, according to the polls, are they attracted to the candidate in question? (Théviot, 2019). But the advantage of this mapping, compared to polling institutes for example, is that it works with objective data, certainly partly from the past, but objective. Polls are often biased because in France, for example, few people dare to say that they will vote Front National or that they will not vote at all. Based on this data, each zone receives a score, and depending on the score, other political techniques will be deployed, which we will detail in the following sections.

4.2.2 Co-constructing the programs

Another application of Big Data is to co-construct programs with candidates. The first one to attempt this was Ségolène Royale in 2007, with *Désir D'avenirs* (Gauthier, 2007). But in 2007, it was a failure and we cannot yet talk about Big Data to co-construct a program.

It is indeed in 2017, with Emmanuel Macron and La République En Marche (LREM), that we really observed a program construction using Big Data. *Le Monde* wrote an article in 2017 entitled : “*Macron, le candidat des médias*” (LeMonde, 2017, p.1). But the real question we are asking ourselves is this: Wasn't Macron the data candidate rather than the media candidate? (De Munk, 2017). Indeed, Macron built its political program, in part, through Big Data and by relying on the French population, which was a completely new method in France. The fact that he was able to put this new system in place can be explained in particular by the fact that he did not have to face internal resistance, because a year before the presidential election, the LREM party did not even exist (Théviot, 2019).

It all began in 2016 with "La Grande Marche", a sort of major national opinion survey carried out by more than 4,000 volunteers and aimed at making a "*diagnosis of the country*" (Théviot, 2019, p.166). It is a sort of huge survey conducted by data, which collected more than 100,000 conversations and 25,000 questionnaires in the field. La Grande Marche thus made it possible to highlight the main issues and respond to them, according to the different profiles of citizens. However, behind the image of a president who listens to people, this way of shaping a program has been the subject of much criticism. People have accused him of not having ideas, of having to go and get them from the French population. But the main criticism was that he was using demagoguery to please citizens by co-constructing a program. In an interview with Anaïs Théviot, the founder of the Civitéo agency said "*Macron, he made his program using algorithms, which is true*" (Théviot, 2019, p.168). But he also explains that what we reproach today to Big Data is what we reproached yesterday to the polls. If wanting to best meet the expectations of the French is considered demagoguery, it is not a problem according to him.

The real problem with the construction of a program via the Big Data is that it actually allows us to map the real concerns of the French, certainly, but then to build a program that only selectively meets their aspirations (De Munk, 2017). When we observe in particular the anger of the “Gilets Jaunes” and the very low popularity of Macron among public opinion (only a third of French people are satisfied with their president), we are entitled to wonder whether the construction of the program was really just a cloud of smoke to get elected, instead of a program that really wants to respond to the issues of the French people (JDD, 2020).

4.2.3 Candidates' websites, which have become interactive Big Data platforms

As already discussed in Section 1, mybarackobama.com has pushed candidates in France and around the world to focus on developing a high-performance website (Pène, 2013).

It is NationBuilder that best explains how to create an efficient and attractive website. Instead of bringing people from social networks to the site, it is social networks that are strongly promoted on the platform. This makes it possible to build a qualified database by cross-referencing the site's data with the data from the social networks. NationBuilder also offers in the back office (part of the website that is not visible to Internet users) several tools for secure donations and fundraising monitoring, integrated with the rest of the community ecosystem (Durand, 2016). Previously, data such as the email addresses of supporters registering on the site were not sufficiently processed and cross-referenced with the personal information available on social networks. Now, thanks to the site, it is possible to invite the voter to a meeting in his or her region, to indicate the nearest voting location, to send thematic messages related to areas of interest, etc.

Macron applied this very well with its campaign site, where it was possible to join online for free. This contributed to the attractiveness of LREM membership, but above all it made it possible to centralize the data of the registrants. It also only took a few minutes to register. The new party did not therefore mark any major differences between members and supporters. They responded to the desire to make a specific commitment on a specific subject for the duration of an election. It is no longer a body and soul militancy that was practiced before (Ion, 1997). The use of data has made it possible to take a step forward in adapting to the decompartmentalization of the classical boundaries of militancy (Théviot, 2019).

But data science has not been limited to the use of this technical instrument, however developed it may be. For a simple reason: relying on a platform for optimizing contact collection does not in any way replace the strategy of defining the electorate to be targeted as a priority and the development of solutions for establishing political relations. To conclude, we will quote Jacques Priol: *"Everyone can afford NationBuilder, but it requires a data strategy that goes along with it"* (Priol, 2017, p.163). In other words, even the best database cannot do without strategic thinking.

4.2.4 Mailing

The case of e-mail addresses is an interesting case to study because it is both important upstream and downstream. Upstream, the acquisition of e-mail addresses is the essential starting point to be able to create a qualified database. A simple e-mail can allow to find name, address, and social network of a

person. Before being able to qualify a database, it is essential to obtain an e-mail address. This also applies in politics, because as Théviot explains: "*the science of data requires very precise data that is not limited to an email address*" (Théviot, 2019, p.103). But the email address is indeed the starting point for creating a database, and that is why politicians put a lot of effort into collecting it, whether in the case of physical meetings (meetings) or on the internet (website).

Downstream, the mailing is one of the ways to contact and appeal to voters. Today, the political parties are sending emails about the candidate's proposals concerning security to a group of voters who claim to be sensitive to this issue. In other words, it is a question of segmenting the electorate according to the message disseminated. For example, a 19-year-old voter should not receive e-mails about the program of nurseries or old people's homes (Jirbandey, 2015). All this has to do with targeting, made possible by the tools mentioned above. There is a real need to target the population contacted. The risk, otherwise, is to create tension and rejection on the part of voters (Théviot, 2019).

However, it should not be forgotten that data tools are designed, among other things, to support fieldwork in a very pragmatic and rationalized logic of partisan activism. According to Eve Zuckermann, at the time director of Juppé's digital campaign, the main objective of data science is "*a logic of incitement to action*" (Zuckermann, 2016, p.156). Nowadays, politicians ask themselves the following questions before sending an email: What are we doing this mail for? Who do we send it to? Does it correspond to the target? The mail strategy is really starting to be a strategy guided by statistics. Traditional practices considered archaic, which do not rely on the skills of data policy workers, are being devalued. The scientificity of these new methods is based on statistics that estimate the reach of an email, for example, by analyzing the rate of opening, clicking, but also negative metrics such as the unsubscription rate or the number of people who mark the email as "spam" (Théviot, 2019).

The ambition is to succeed in sending so-called "optimized" emails, thanks to an A/B test system. Two different versions of emails are sent to two panels of supporters and an assessment is carried out to find out which of the two versions is the most effective (open emails, clicks on the attachment, response, etc.). Before sending an email, the first email is sent to 10% and the second to another 10% of the database. Then, you have to take the best result to send it to the remaining 80%. These are now large-scale tests that are carried out and this is the real contribution of Big Data, we no longer rely on the political guesstimate (Pilon, 2016). However, these testing and mailing techniques are ultimately carried out to encourage action, whether in the field or online.

4.2.5 Phoning

Phoning is a technique very comparable to mailing, in that it also results from the targeting and personalization of messages sent to voters. The three Bostonians claim that phoning makes it possible to convince one voter out of 38 (Théviot, 2019). It is also a function offered by NationBuilder, in the continuity of emailing, they also leave voice messages in the mailboxes of all the members of their database. One of the founders of LMP also explains that they had an application that allowed them to track the conversations they had had with people. Like mailing, phoning is therefore also used upstream, to qualify databases. However, this technique is not very well explained because it is a bit scary and on the edge of legality.

Macron, during its election campaign in 2017, opted for targeted phoning operations in order to go beyond the doors that don't open. When door-to-door, a technique that we explain in section 4.2.6, did not work, the idea was to contact people directly on their landline using the numbers provided in the telephone directory. Macron's teams contacted more than 6 million households through a targeted telephone campaign in which the candidate explained why he or she should be chosen in the first round of the election, using automatic and recorded messages.

Many voters had a bad experience with these telephone calls and reported it to the CNIL, which stated that from a legal point of view, it was possible, because there is always the possibility for the voter not to listen to the message. The CNIL nevertheless said that this did not correspond to their recommendations, to which Macron reportedly replied: *"you are right, but that does not mean that it is illegal"* (Macron, 2017, quoted by Théviot, 2019, p.173).

4.2.6 Door-to-door contact

Door-to-door campaigning is a very old militant technique, but in our opinion, it is the beginning of the electoral Big Data, since the dazzling success of Obama's door-to-door campaigns in 2008 and 2012. It has existed since the beginning of politics. However, the Big Data has greatly evolved it. We will see how door-to-door is practiced today in an election.

As said before, the mapping and the algorithms used allow to deploy multiple actions of electoral mobilization. Door-to-door is one of them. What has changed dramatically today is that door-to-door has become very targeted. You can, for example, go door-to-door in a certain neighborhood, where most of residents are supporters who abstain. What has also changed is that now we report after going door-to-door, to be able to qualify the databases. The membership files from the door-to-door meetings therefore complement the open-data provided by the service providers. Clearly, the approach has always

existed but the tools are changing. Today, the door-to-door process is supported by software and databases, with a view to rationalizing it (Moysan, n.d.).

As detailed in Appendix 13, the door-to-door scheme during the François Hollande 2012 campaign consisted of three stages: recruiting volunteers, training them and then setting up an efficient organization to coordinate their actions (Théviot, 2018). But before recruiting volunteers, it was necessary to target high-potential neighborhoods (left-wing abstainers). Then, it was a matter of relying on the hard core of activists already in existence to promote this system among non-members. The new door-to-door campaign is based on a search for efficiency; it is no longer a question of walking the streets and ringing every doorbell. There were very strict instructions, such as never spending more than 5 minutes per door. Questions were also written in advance. In addition, if the voter was rather reluctant, the aim was not to convince him or her, but to end the conversation. After each door-to-door session, a follow-up sheet had to be filled in, in order to fill in the database and centralize the information on the *toushollande.fr* website. Now, what is done in the field is recorded via a web platform. It is a continuous cycle. The difference between Hollande and Obama is that Obama has managed to bring in people who were not initially party supporters. This is the major reason for the failure of the door-to-door campaign of Hollande in 2012 according to the 3 founders of LMP (Moysan, n.d.).

In 2016, during "La Grande Marche", a few thousand volunteers also devoted themselves to this practice. However, the objective was not the same because Macron was not officially in the campaign yet. As we have already explained, the objective of this great march was to go door-to-door to identify areas representative of the French population. The work that LMP did in 2017 is therefore very different from that done with Hollande in 2012 (Théviot, 2019). In 2017, they were really looking to create a map that was the most representative of the French population. It was therefore a work upstream of the campaign, and it was a real success. According to the three Bostonians, door-to-door is a more efficient technique than phoning, for example, and would make it possible to convince 1 voter out of 14 (against 1 for 38 for phoning).

4.2.7 Predicting voter behavior

Among the many fantasies surrounding electoral Big Data, some people believe that it would be possible to predict voter behavior (LePoint, 2017). Data workers claim to be able to act scientifically on the outcome of an election. The use of data by political parties appears to be the most effective and recent technique for reducing electoral uncertainty. In the scientific literature and articles, there is much more talk of "reducing uncertainty" than of "predicting voter behavior". However, reducing uncertainty is a way of "maximizing certainty," and therefore still trying to predict voter behavior as best as possible.

Obama's teams had implemented a microtargeting system that used the electronic trail left by voters on the Internet to give them scores based on the likelihood that they would or would not vote for him. The goal was simple: to avoid dispersal by chasing away "hidden" voters and to better target the candidates' persuasion efforts (Pène, 2013). Sarkozy in 2012 had also asked members of his campaign team, both national and local, to fill out forms to find out the level of commitment of his contacts (Théviot, 2019).

What is called 'predicting voter behavior' is, in our view, the ultimate step in qualifying a database. This allows, in fine, to classify voters according to many criteria. It allows, for example, to estimate whether or not Mr. X will agree to participate in door-to-door canvassing. But we can still consider that these prediction models are very efficient at least for estimating the people who will certainly not vote for the candidate in question. Thus, voter behavior is predicted more in the negative, and as mentioned earlier, this is more often "reducing uncertainty".

4.2.8 Improving Internal Party Management

Political parties have also started to use marketing techniques to strengthen the customer relationship (the voter is then seen as a customer) that has long been known in business and which now applies to "political companies". Campaign teams help organize more effective campaigns mainly in two aspects (Théviot, 2019). The first is the strategic aspect. They help the parties to know the territory better, to know the electorate, to understand who they need to talk to, in which neighborhood they need to organize their campaign as a priority. It is therefore the first dimension of helping the strategy from the data. Then there is the second dimension, which is help in organizing the campaign (Pilon, 2016). This is done, for example, using Customer Relationship Management tools, which refers to computer software that helps manage the relationship of a company or political party with its voters or customers. There is also help in creating events. These are more operational actions to help organize traditional militant actions.

Today, we can no longer say that there are computer scientists on one side and politicians on the other. This change in professional qualifications has led to a reorganization of the militant work of campaign teams where online, but also offline mobilization is thought of in direct connection with the digital poles. Digital is no longer an extension of the communication sector, but has become the central gateway for supporters. It is therefore no longer just a question of thinking of the web as a communication tool, but also as a strategic tool for mobilizing and organizing the field offline (Georis, 2017). Electoral data and Big Data workers then present themselves as indispensable to support campaign teams in order to train them in the management of these tools (Ehrhard et al., 2019).

A good example of new internal party management is, of course, Macron's LREM party, which many consider as a start-up because of its functioning (De Munk, 2017). The rhetoric of this new party, around

the participatory dimension, which is supposed to be a break with the traditional ways of political parties, began and was built from "La Grande Marche". While the traditional and hierarchical structure of classical parties were increasingly criticized, the new LREM members emphasize the participatory side of the movement with a territorial focus, particularly thanks to the contribution of digital technology, which makes it possible to send data up more quickly and work more efficiently locally. One of LREM's strengths has been to share its database on a local scale, dividing it up by territory to support its candidates (Théviot, 2019). The reverse is also true: during the legislative elections, local candidates brought the information gathered in the field to the national level to enrich the global database that could be used for other future elections.

4.2.9 *Renewed activism*

A major change that has taken place in France since the years 2010 is the renewal of militancy. In France, until recently, there was a strong partisan culture, it was necessary to be "embedded" in a party, we spoke for example of the *"éléphants du PS"* (Boireau, 2017, p.1). The trigger for this renewal of militancy is the recruitment of online members, whether at the UMP, the PS (with the '20-euro members') or even more recently at LREM. Indeed, recruiting online favors a deterritorialization of activists. It allows new members to move away from local logics of allegiance to the currents. These Internet-members or "off-ground" members have been the subject of much criticism, rejection, from traditional supporters (Fillieule, 2005).

A second characteristic of the renewal of militancy is the willingness of activists to actually participate in politics. The "Désirs d'Avenir" platform created in 2007 by Ségolène Royale reflects this trend. Voters are increasingly eager to debate, give their opinions and submit their proposals. In 2010, the PS's Coopool (Political Cooperative), for example, allows activists to exchange among themselves, to carry out actions, demonstrations, working groups, think tanks, not only on their own territory but all over France.

Finally, the third and final characteristic of the new activists is their "punctual" engagement. This is also due to the fact that parties, for example during primaries, apply a form of ad hoc openness with a view to maximizing electoral performance. Many members and sympathizers are interested in politics only during the election campaign and do not show any sign of life afterwards. This "ephemeral" nature of activism is also something that characterizes electoral Big Data. When it comes to defining one's political strategy, it is therefore important to take this factor into account. There are therefore two possible strategies: either to base oneself on a hard core, as François Fillon did by basing himself on the "right-wing Catholic bourgeoisie". Or, as Macron did, to make the gamble of relying on data sets to encourage the creation of new groups (Théviot, 2019; De Munk, 2017).

4.2.10 Managing and Creating Community

We have noticed that the Big Data revolution in politics is strongly marked by the desire to federate and create a community. Of course, this has always been the case in politics, but the way of doing things has changed. Communities are now being created online.

One characteristic of these online communities is that they emerge very quickly and are quite difficult to control and manage. Often, they are created against the power in place, it's like a new form of contestation of political power. But we have noticed that having an active online community has become a must for political success. NationBuilder understands this very well and it's one of their main features (Pilon, 2016). The mentality, when they work with people who are starting a political project, is really to create a community that the candidate needs to succeed in the project they have in mind. The better this community is known, be it through their history of engagement (the disappointed on the left for example) or through information gathered on social networks, the better one can create a link with this community.

Once the community is created, it is possible to classify it, and this is where the use of mass data comes in. Obama, in 2008, had managed to bring together a lot of people who initially had nothing to do with his political movement. The use of the web appears to be a relevant tool to address this new audience and thus create a community of support. Once the "grassroots" circle has been created, the challenge is to broaden it as much as possible.

Macron, in 2017, has perfectly understood this (De Munk, 2017). He put a lot of emphasis on recruiting volunteers through the Internet, and digital networks then allowed him to promote his ideas and his program as close to the ground as possible. Depending on the profile of the Internet user, it was possible to propose specific actions. The use of data made it possible to propose a whole range of one-off commitments: simply retweeting a message from a candidate, going to drop leaflets in the mailboxes. The qualification of the data makes it possible to limit the risks of saturation and unsubscription.

4.2.11 Voter profiling

Software and algorithms now make it possible to parameterize the science of elections. It is really a question of segmenting the electorate according to the message being broadcast. The simplest segmentation is at least by age. There is no need to send e-mails about the university curriculum to people in nursing homes. This categorization of the electorate is made possible through the use of a qualified database (Pilon, 2016). The idea is to enrich the database to personalize it as much as possible by individual. This is what we have called voter "profiling". This profiling makes it possible first to

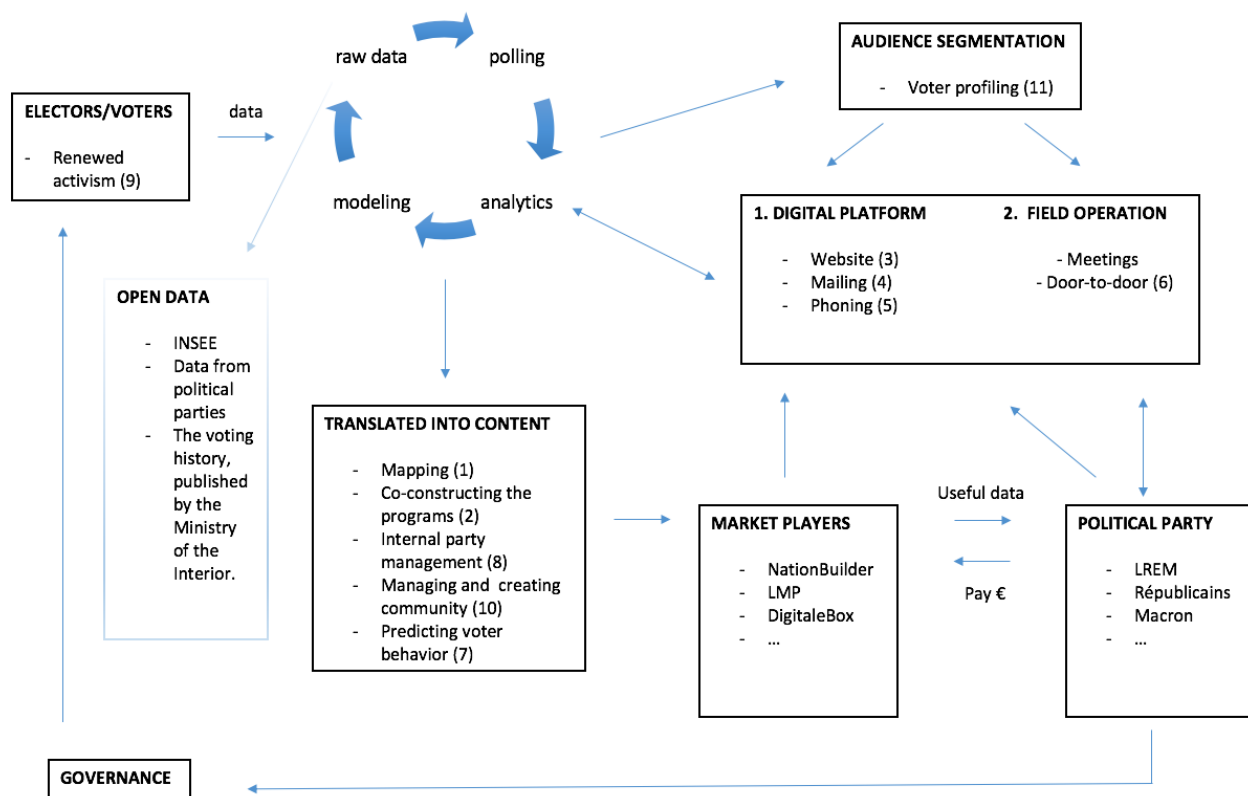
identify people who are likely to be interested in a candidate's political platform, and then to send the right message to the right people. It also makes it possible to create a profile of a "typical" elector, i.e. the elector most representative of the candidate's clientele.

During the door-to-door campaign, for example, the profile of voters was established as follows: "left-wing abstainer", "active left-wing voter", "other", etc. The profile of voters was also established as follows. The ready-made questions were then asked according to the voter profile, and how to answer them as well. Profiling is to the individual, in a way, what mapping is to geography.

5. General Process of the Electoral Big Data in France

The following process was partly inspired by Cambridge Analytica's leaked plan for Trump's victory in the 2016 elections. But it is, for the most part, the fruit of our own thinking and work. It is a complete electoral Big Data process in France, starting with "voters" and ending with "governance". It synthesizes everything we have discussed in section 2, as well as some of the tools discussed in the theoretical part.

“Complete process of the Electoral Big Data in France”



6. The prospects of the electoral Big Data

We want to make a clear distinction between the "future prospects" and the "risks and dangers" of its use in politics, which we will detail in Part 3 (cf. Part 3: Data-driven Governance).

6.1 Standardization of the commercial actors of elections

The first trend that we observe in the elections is a reconfiguration of the market of market players. This is not a question of questioning the effectiveness of their actions, but rather of studying their economic logic. Like all businesses, the market players in the electoral Big Data pursue objectives of profit. The "poltech" has a low economic potential and limited profitability according to the actors. Indeed, it is conditioned by the public funding used by candidates during election campaigns and the funding of political parties is limited by the law, which restricts their potential (Théviot, 2019). There are 3 solutions that actors implement to face economic difficulties: intra-political diversification, extra political diversification or a complete withdrawal from the market (Ehrhard et al, 2019). The problem is that many political parties call on actors during campaigns, but once elected, close down applications. This is what pushes market players to change their strategy. Moreover, the problem with elections is that they tend to depreciate the value of actors in the electoral Big Data. It's an environment where there's a lot of blackmail, a lot of hitting below the belt, but, in the end, it doesn't bring in much money.

Until now, the actors in Elections 2.0 have been heterogeneous in three ways: sociological (type of actor), professional (diversity of services and tools) and economic (in development strategies). However, the current transformation of the elections 2.0 sector corresponds to a process of professionalization (Ehrhard et al, 2019). The standardization of entrepreneurial logics, the services offered, and the implementation of new strategies open up the possibility of a reconfiguration of the poltech sector. The economic limitations of political actors in France are leading to an increased empowerment of economic actors vis-à-vis the political field. The three types of strategies of market players essentially respond to the economic limits of the market (profitability, development potential) in its current state (Pilon, 2016). Its extension outside of election campaign periods and its opening up by diversifying the customer base outside politics are responses to this "glass ceiling" (Ehrhard et al., 2019). These strategies outline a new market configuration from which some market players are withdrawing. For them, the solutions envisaged will not make it possible to go beyond the economic limits or would require too radical a change in the business model. The reconfiguration of the market therefore leads to a reduction in the number of players, in favor of larger, internationalized companies, whose development requires activities that are less focused on elections (Pilon, 2016). This is already the case for NationBuilder, which has more than 9,000 clients in 112 different countries (Ehrhard et al, 2019).

In this sense, the French players, in addition to seeking corporate clients, are trying to export themselves. For LMP: *"the political environment is not too small because, throughout Europe, the political market is fairly large, there is a lot to be done"* (Pons, 2019, p.130). For example, DigitaleBox has created a geographical presence in Italy and England, thanks to its ecosystem of some twenty companies that carry the DigitaleBox brand using their software. Foreign players consider that their presence in France will increase and that the European market will grow. Ecanvasser believes, for example, that the glass ceiling of data analysis has been reached in the United States, whereas much remains to be done in Europe, even if economic constraints remain strong for the time being (Ehrhard et al, 2019).

6.2 Towards an American-style two-party system

The second future perspective that we imagine is that of the recomposition of the French political system into a very right-wing bipartisan system and a center-left party, as in the US (Cossé, 2018). Without going into too much detail in French politics, we will focus on this future and potentially probable model of French politics copied from the Americans.

Several clues lead us to believe that the reconfiguration of French politics, which is in the process of being recomposed, would tend towards an American model. The Democrat party, center-left, already exists and is perfectly embodied by LREM, which has succeeded in merging the left and the center, leading to the almost certain disappearance of MODEM (center party) and PS (left party). The Republican party could be embodied by LR, which has already taken its name, and which could turn into a merger of former UMP and Front National. The disappearance of the FN does not yet appear to be on the agenda, but their failure in the 2017 legislative elections does not bode well (LeMonde, 2017).

If such a change takes place, the primaries would then have their rightful place in this system. As we have seen, primaries are a great way to collect data. However, the problem in France today is that voters on the opposite side of the House are now offered the possibility of disrupting the choice of the candidate of the other side (Cossé, 2018). The problem in France, for a much broader application of the electoral Big Data, remains the control and the ceiling of campaign budgets. These are in fact quite low and very controlled by the CNCCFP (Commission Nationale des Comptes de Campagne et des Financement Politiques) (CNFFP, 2020). But in an increasingly liberal American-style logic, it is quite conceivable that the rules will become more flexible in the more or less near future, which would lead to a much greater use of digital and Big Data. After the primaries, the party would then rally around their presidential candidate, whose program would replace the party's. A good way to move on to the next section, where we will discuss the issue of the presidential candidate's "person".

6.3 Lower and lower barriers to the Presidential Office

As already mentioned in the section on "renewed activism" (cf. Part 2, section 4.2.9), traditional French partisan culture is collapsing and the barriers to becoming president of the republic seem to be falling more and more. The co-founder of Fédéravox states on this subject: *"The change we have seen throughout Europe is that the new tools allow political parties that did not exist a year ago, a year later, to win an election, as we have seen in Paris. That's the story of En Marche, it's amazing! And we saw the same thing with Jeremy Corbyn for the Labor Party in England, we see the same thing in Belgium with the Communist Party which is destroying the socialist party in Belgium. ...]* So I think that with all this innovation, all this technology, the entry barrier is much lower than it has ever been, and as a result, it allows anyone to be president of the Republic, to have a new political party" (Bon, 2017, quoted by Théviot, 2019, p.136). Trump is an illustration with nothing to do with politics. Electoral Big Data make politics more unpredictable because, at any given moment, we observe the possibility of a candidate emerging out of nowhere. It's an important paradox of electoral Big Data.

6.4 Information disruption and polarization of Society

As the Deutsche Bank study (2019) shows, traditional sources of information lose their influence to the profits of social networks and online information (Körner, 2019). As such, an alternative information offer, which is moreover free of charge, is not a problem. It could even be virtuous, in the sense that it would allow information to be diversified, give power back to citizens, and even expose the corruption of certain governments.

However, the problem is the lack of control that often characterizes new information platforms such as Facebook and other social networks or websites. Moreover, the government has reacted by activating an "anti-fake news" platform (LeParisien, 2020). "Désinfox" is controversial because how is it possible to ensure that agencies responsible for disclosing "the truth" are not biased? Moreover, truth is often a matter of perspective, and a well-known and fairly accurate expression from our point of view is *"the Truth, only God knows it"*.

But the problem with Big Data is that these information platforms have an inherent tendency to algorithmically reinforce the various opinions that exist among users. This can then lead to coordinated misinformation and propaganda, and digital filtering bubbles or 'echo chambers' that can polarize society and undermine public dialogue, especially during election periods (Körner, 2019). The problem with Big Data is then not the dissemination of fake news, a phenomenon that has always existed, but the speed and scale at which such false news can spread despite debates on the concrete proposals of the candidates' programs.

6.5 Change in power and influence in the data economy

Finally, a future prospect that is tending to become increasingly reinforced is the concentration of data and economic power in the hands of a small oligarchy (Körner, 2019). The exclusive control of massive data by big technicians gives them an enormous advantage not only in their respective market segments, but also in the development and formation of artificial intelligence and Big Data. This further strengthens their position in the established markets, of which the electoral Big Data market is a part and gives them a head start on the new ones. Facebook, for example, holds 60% of US digital advertising market (Körner, 2019).

This is what Harvard economist Zuboff calls "*the Age of Surveillance Capitalism*" (Zuboff, 2019, p.1). The economies of scale of platforms have led to a concentration of data, knowledge, financial power and control of communication channels in the hands of a small technological elite, often referred to as the GAFAM (Zuboff, 2019). In politics, this translates into a rapidly growing social asymmetry of economic and political influence. The influence of voters, in the broadest sense of the term, will continue to diminish, whereas supposedly free and democratic decisions could ultimately only be a democratic façade (Körner, 2019).

This future perspective could also explain the loss of interest of the French in elections and politics in general. In 2017, France recorded record abstention in legislative elections, more than 57%. It continues to rise, as shown in Appendix 14. The presidential election remains the only election where turnout remains relatively stable. The increasingly visible link between political and economic elites can be considered as a betrayal of the political elites in to their electoral base (Polony, 2014).

6.6 A regularization of the use of data

As Körner explains, Europe is lagging in the war of big technology companies. It is a war in which Americans and Chinese are way ahead. Europe's reaction is to be a world leader in data regulation. In 2016, the European Union adopted a regulation on the use of data called GDPR (General Data Protection Regulation), which is supposed to protect citizens from the massive use of data, especially during election campaigns (European Union, 2016). The aim is also to inform people about the use of data, through learning: "*It is a learning process: shortcoming and loopholes will need to be addressed, and rules and regulations adjusted in order to balance economic, societal and political aspects*" (Körner, 2019, p.16). But surprisingly, companies such as LMP or Fédéravox, market players in the electoral Big Data, do not see this new regulation as a brake on their development (Erhard et al., 2019). Mr. Maes confirmed us that it is absolutely necessary to regulate the use of data, and quoted the GDPR as a good example of such regulation. But Mr. Aladenise, who is a lawyer and data protection specialist, believes that the legal framework for the personal use of data is not at all effective enough today.

In France, in addition to European legalizations, there are also national control institutions such as the CNIL. The CNIL is considered to be "*the data gendarme*" (Girard, 2017, p.9). During the 2016 elections, for example, the CNIL banned the "NationBuilder Match" application, which enriched candidate databases by automatically sucking data from networks using an email address (Théviot, 2019). It remains to be seen, given the importance of Big Data Velocity, whether these regulations will be sufficient to regulate the use of data by political actors.

7. The major trends identified

7.1 The application of electoral Big Data summarized by 4 major trends

Following this analysis of the advantages, opportunities and future prospects of using Big Data in the electoral sector, we would like to draw up a synthesis of the main trends noted. This will allow us to make the transition from this section to the next, which deals with the impact of Big Data in the world of governance. We will also summarize all the techniques used in Big Data in a large summary table.

7.1.1 Personalization

The Big Data is a large, varied set of data collected almost in real time. Globally, the information that is linked is traditional data, data from the Internet and data generated by sensors. The analysis of this cross-referenced data (Big Data Analytics) enables political parties and election market players to better understand the needs and preferences of their voters or "customers" and thus provide them with a better and more personalized service.

7.1.2 Predictability

When an entity has at its disposal a set of data that can be qualified as Big Data, beyond using the data to learn more about its constituents, it can also apply predictive analyses to it. The purpose of these analyses is to identify correlations, interactions and recurring patterns in the data in order to predict future behaviors or events. Once the market actor or political party can predict, it can act accordingly:

- If these future behaviors or events are opportunities to gather votes, then the organization can deploy the necessary means at the right time.
- If these future behaviors or events constitute risks or threats, then the organization can put the means in place to prevent them and try to fight against them.

In the private sector, Mr. Dombret explains to us that predictability is used a lot to analyze the internal market, to see what decisions we are going to make for the future but also to observe what is happening outside the market, in order to be sure that the ecosystem in which the company evolves is always in line with reality. In concrete terms : "*this allows us to target certain customers, certain products, whether being on past or future trends*".

7.1.3 Optimization

Internal data analysis increases the efficiency of internal processes and reduces costs. This optimization is also reflected in the ever more efficient calculations of the voters to be targeted, which allows an optimization of the political techniques deployed during or even after the elections to support a political decision, for example, and to look for support points.

7.1.4 Objectivation

Another major strength of data analysis is that it makes objective decisions and judgements. Indeed, Big Data analytics enable organizations to make decisions and judgments that are no longer based on intuition, assumptions, traditional rules or reasonable hypotheses, but rather on data analyses that are objective, since they are performed by algorithms and no longer by humans. Even though this so-called "objectivity" is the subject of much debate and discussion, as we discussed in Part 1, section 3.3. For this trend, Mr. Dombret gives us a concrete example, which shows that we sometimes must detach ourselves from human intuition. A manager at Ethias was convinced that there were more car window breakage claims in Charleroi than anywhere else in Belgium. But by looking at the data, they could see that this was not true. This is why he stresses the importance of being fact-based.

7.2 Summary Table of Electoral Big Data Techniques in Relation to the Four Major Trends

In the following table, we will summarize the eleven electoral Big Data techniques identified and developed and see which trends are used for each of them.

	Personalization	Predictability	Optimization	Objectivation
1. Mapping			X	X
2. Co-constructing the programs	X	X		X
3. The websites becoming platforms			X	
4. Mailing	X		X	
5. Le phoning	X		X	
6. Door-to-door	X		X	
7. Predicting voter behaviour		X		X
8. Internal Party Management			X	
9. Renewed activism	X			
10. Management and community building	X	X		X
11. Voter profiling	X	X		

Part 3: Data-driven Governance

In the first part, we described the context of the Big Data revolution in our society. In the second part, we analyzed the impact of Big Data in an area where data is becoming increasingly important: the electoral Big Data sector. We concluded this second part by identifying four major trends in the use of Big Data in the sector under review: personalization, predictability, optimization and objectivation. We have seen that through these four major trends, political parties and election market players manage to create value both internally and externally, in the relationship with their customers.

Now, we are interested in the influence of Big Data on governments, on the state in general. Like the electoral Big Data sector, the state is an organization that, in its role of governing citizens, is required to manipulate large amounts of data. As shown in Appendix 15, the government is the second sector with the most stored data in the US, after the manufacturing sector. If the government has so much data in mass storage, then it also has enormous potential for value creation (MGI, 2011).

The purpose of this third part is to explore the possibility of Data-driven governance. We want to envision a state that uses Big Data in the image of political parties during elections, to create value both internally and externally for the citizen. To do this, we ask ourselves two questions:

- First, how can the state use Big Data to create value?
- Secondly, what are the challenges and risks that the state must face if it wants to enter the Big Data trend?

To do so, we will for each trend observed, make an assumption of application in a sector where this trend could be particularly useful. Thus, we will:

- For the personalization trend: talking about the education sector.
- For the predictability trend: talking about the fight against tax and social fraud
- For the optimization trend: talking about improving road traffic
- For objectivation: talking about the connected objects “secured by design”

Then, in a double entry table, with criteria on one hand, and the 4 sectors where we apply a Data-driven governance hypothesis on the other, try to evaluate and audit each sector in order to see in which sector a Data-driven governance is possible or not. The table as well as the criteria’s used will be explained in section 3 of this part. The way of evaluating will also be explained in section 3.

Since we are talking about the State and Big Data, it is obviously interesting to know where Belgium stands in relation to other countries in terms of digital strategy. Every two years, the United Nations Department of Economic and Social Affairs conducts a survey on the state of e-governments around the

world (United Nations, 2018). The survey assesses Governments on the basis of three key aspects of e-government: online services, telecommunications infrastructure and human capital.

The result of the survey includes a table that represents the ranking of world leaders in e-government. We note that in 2018, Belgium ranked 27th in this ranking, which groups the 40 leading countries in e-government. This table shows that our country is outperformed by 14 European countries, 3 Asian countries and other major countries such as the United States in particular. Belgium can undoubtedly do better; we will see how.

In the first section, we explore the main ways in which Big Data can be used in several public domains such as education, road traffic, tax and social fraud and connected objects “secured by design”. We believe, at our level, that the Belgian State could take inspiration from these few avenues for the overall improvement of its services.

Before continuing, a preliminary remark must be made. In this section, we therefore speak of "state" and "government". However, the Belgian State is made up of a federal State and federated entities, which may create confusion as to which entity is the subject of our presentation. We would therefore like to make it clear that by "state" or "government" we mean political power and public services in the broadest sense of the term. The various applications of Big Data that will be developed are therefore, in the context of Belgium, addressed to both the federal State and the federated entities, according to the competences assigned to them.

1. How can the State use Big Data?

To answer this question, we study how the Big Data can revolutionize the way in which the government operates in the light of the four main trends that we discovered in the second part. By tracing these trends to the field of public governance, here are the four main ways in which the Big Data can be used:

- Personalized service to the citizen.
- Predictability: preventing fraud and threats.
- Internal optimization.
- Objectivation of decision making and judgment.

We asked several people about our hypotheses of application of Data-driven governance in certain sectors in Belgium. Our approach has been to interview people from different sectors in order to get different perspectives. For example, we interviewed Mr. Dombret, who works for Ethias, and Mr. Maes, who works for Wizata. Their two interviews allowed us to get the point of view of the private sector. We then interviewed Martin de Waziers, who is a member of the "Order of Malta", in order to get an

ethical point of view. To get a "legal" point of view, we interviewed Mr. Aladenise, a lawyer and data protection specialist. Finally, in order to have a more academic point of view, we interviewed Mr. Faulkner, who teaches Big Data at the University of Namur. We would also have liked to interview someone of the government, in order to complete our analyzes, but unfortunately, we did not succeed.

1.1 Personalization of service to the citizen

In the area of electoral Big Data, we have seen that Big Data analytics allow organizations to improve voter or client satisfaction by offering them an increasingly personalized service. In addition, they also make it possible to increase the number of votes collected by offering, for example, a program adapted to the expectations of the citizen. We are now transposing this dimension of personalized services to the domain of government: how can the state create value by offering a personalized service to its citizen?

1.1.1 From standardization to personalization

Market segmentation to personalize marketing actions is a trend that has long been accepted in the private sector. The public sector, on the other hand, has operated for too long with a standardized system of "same service for everyone". It assumed that economies of scale would make the system less costly than a personalized public service program. However, population segmentation and the personalization of public services have been shown to increase the efficiency of services, reduce related expenditures and increase citizen satisfaction (Diez, 2016).

A personalized public service is *"a public service that is unique in its purposes, but differentiated in its modalities according to the needs of each user"* (Grosdhomme Lulin, 2015, p.19). This differentiation is made possible when the State, which possesses mass data on its citizens, gives meaning to this data by applying advanced techniques that allow for sophisticated and individualized processing.

In the head of government, there is therefore a real opportunity for value creation at the organizational level, but this personalization of public services benefits above all the citizen, with better quality service, more appropriate to his or her needs (Desouza, 2014; Hamel & Marguerit, 2013).

Here, we can note a first difference with respect to the elections sector, studied in the second part. For this sector, the personalization of programs, for example, allows, beyond better customer satisfaction, an increase in the number of votes collected by the candidate. It is therefore a "win-win": everyone wins. On the other hand, in the field of governance, the aim of the State is not to sell services, but to fulfil its duty of public service to the citizen. Consequently, it is above all the citizen who will benefit from this personalization. It is always a win-win situation, but the citizen probably has more to gain than the State.

1.1.2 The case of teaching and education

We illustrate how the personalization of service to the citizen can be achieved by education. Today, Belgium is characterized by a mainly standardized education: a teacher gives the same course, the same exercises, and evaluates according to the same criteria all the pupils in his class, even though each of them has a different level (Diez, 2016). This is the principle of diversity. Nevertheless, it is important to explain that education has evolved compared to previous generations. As Mr. Maes explained, we have already switched from an education in which we had to "store a lot of information" to an education that is more focused on understanding, which is a very positive change according to him. But he also explains that we are lagging far behind in the education sector in Belgium, and he takes Luxembourg, where schools are much better equipped, as a point of comparison. Mr. Aladenise, for his part, explains that there is a need to better equip schools in Belgium in terms of IT, in order to potentially be able to use the data. He also explains that not everyone is in the same class, and that there is a great disparity between certain schools (from primary to secondary). Mr. Dombret explained that for him, the importance of equipping schools with computer equipment would allow students, who have difficulty taking notes and listening at the same time, for example, to listen to the audio lessons again afterwards. This is another argument in favor of better equipping schools with computers, which would make it possible to take into account the differences that students have in their approach to education. Mr. Faulkner, for his part, mentioned a growing disparity between Flanders and Wallonia. According to him, Flanders has overtaken Wallonia in the field of education. He gives the example of start-ups. The most successful start-ups are mostly created by students coming from the Dutch-speaking school system. On the overall level of education in Belgium, he is quite satisfied and this is confirmed when he looks at the results obtained by Belgian students sent abroad on Erasmus. Nevertheless, he confirms, like Mr. Aladenise, that it is imperative to better equip schools and universities with digital technology. Today, computer tools are, according to him, too much used in a playful aspect (games, applications, etc.) and that we should rather focus our efforts on the "professional" aspect of computing because he believes that in the future, it will be very difficult to find a job without having been trained in digital tools. He says: *"This is where education really has a role to play"*.

Drawing on the initial progress of the digital strategy in French education, we have a glimpse of what is to come with Big Data revolution. The first step consists, using the most basic computer technologies (interaction, experimentation, trial-and-error, "e-learning" tools), in proposing to each student individualized learning. Nevertheless, Mr. Maes thinks it is essential to keep a common base. He also explains the importance of the social mix that education provides. According to him, the purely academic aspect is not the only element to be taken into account when talking about education. This is confirmed by Mr. Aladenise, who also adds that transcripts and other records are not the only criteria to be taken into account when evaluating a student. Some academic failures are not only due to the

academic aspect, other external factors can also play a role, such as a student's family situation or mental health. And this is something the Big Data can hardly capture. To sum up, he is not against the use of such tools, but he considers it an obstacle to use only "objective" data such as result transcripts to judge a student. Moreover, he thinks that in Belgium, teachers are not well trained enough to use of Data-driven tools to improve the education system.

This path (using basic computer technologies) would allow the student to work on the subjects in which he or she has more difficulties and to spend less time on the subjects he or she masters. Conversely, we might also think that it could be wise, in certain cases (for example, a student who is gifted in mathematics), to focus efforts on the student's strengths, so as not to slow down his progress and potentially make him or her a future elite in his field. In this respect, Mr. Maes explains that he fears that if data-driven governance is used in the education sector, it is only to train elites, and not to help students in difficulty. He fears the practical application of such a policy. Even if, in his opinion, it can be very beneficial in theory. Another Big Data application in the context of remote examinations, which Mr. Maes was able to observe through his children, is the Testwe platform. As it allows the movements of the head, eyes, etc. to be analyzed for cheating, this is a concrete case of a Big Data application in the education sector, using Volume, Variety and Velocity of data.

The second step would be for the teacher to have access to a "digital work diary", which would allow him/her to have a better vision of the level of his/her students, both at the individual and the class level. The teacher would be able to see the actual work done by his students, identify where they had the most difficulty and thus evaluate them in a more relevant way. Mr. Aladenise believes that such Big Data oriented analyses at the educational level can only be relevant if they are carried out on a very large scale and not at the individual or classroom level, for the reasons he explains above. Mr. Faulkner is in favor of this kind of tool, but he nuances. He believes that this tool should be complementary to the teacher's role, because the teacher cannot multiply himself. It would therefore be necessary to complement the work of the teacher, who works face-to-face, with a tool such as a digital diary, in order to make progress on the individual approach.

For the third step, we are at the national level, in the context of students entering higher education. With the necessary computer tools, the Ministry of National Education could collect and collate all kinds of data relating to education and develop algorithms to assist in decision-making. For example, as in the case of elections, where the political parties use personalized mailings, the state could recommend a student a particular course of study based on his or her background and strengths. But in addition to its strong points, the State could go even further by combining education with the job market. For example, it would make sense to direct students towards sectors where there is a strong demand for employment. In this connection, Mr. Maes explained to us that: *"for example, today the Belgian State needs 3,000*

bricklayers. Even if, according to him, students should not be forced to choose certain trades, the state could still create a platform that would somehow identify sectors where the demand for employment is higher. This would help students who are hesitating in their choice of studies." Today, however, we can see that many students are heading towards studies that interest them, without thinking about job opportunities. The result of this is that we are seeing more and more academics who are unemployed, sometimes even directly after leaving university (VRT, 2012). At the national level, Mr. Faulkner is divided. He is not in favor of an American-style system where universities would recruit students, because that would risk creating a two-tier society where there would be favored students, on the one hand, because of their education and social background, and on the other hand, students who would be sidelined because their parents would not always have insisted on the importance of education. By contrast, he fully agrees that the State has a major role to play in informing students about educational opportunities. He has seen many students leaving University without any prospects and end up on the sidelines of the job market for three or four years. He therefore believes that the State has a real role to play as an informer, using data and statistics on supply and demand in the labor market to inform students about future study opportunities. Mr. Dombret is in favor of the idea of using data, both by universities, in order to guide students in their future studies, and both by the State, which should inform young students much better about what is really going on in the economic sphere.

This could lead quickly to the creation of a digital portal that would provide access to many useful data for students, such as statistics on success in such training based on courses already taken. If we go even further, this digital portal could be linked to social networks, with the aim of creating more interaction between students in the country, but also between students and higher education institutions. Thanks to data mining techniques (the exploration of large quantities of varied data at very high speed to detect correlations) applied to this nascent Big Data, these institutions would even be able to better match courses and student profiles.

Implementing a Big Data strategy enables the State to create real added-value in terms of personalizing the service to the citizen, in this case a young person seeking training to prepare for the job market. Beyond education, the State could also, thanks to the use of Big Data, provide the citizen with a more personalized service with high added value in the areas of:

- better support for jobseekers (a case study on the subject can be found in Appendix 16) ;
- justice, with the individualization of sentences.

Thus, this new Big Data philosophy would enable the State to better respond to the needs of its citizens, by making use of mass data but also by making it available to the general public (Grosdhomme Lulin, 2015). In other words, the State can use Big Data technologies to offer its citizens a "smarter" public service.

1.2 Predictability: preventing fraud and threats

Here we find the second trend arising from the use of Big Data analytics in the electoral sector: predictability. As we have seen, predictive data analytics are becoming increasingly popular in this sector, enabling political parties and market actors in Elections 2.0 to create value by more effectively predicting risks and behaviors.

To prevent fraud and improve public safety, the government's Big Data strategy is to collect and collate all kinds of behavioral data that describe the reality of people's lives first continuously. Then, using ever more efficient algorithms, Big Data analytics, and data mining techniques, enable more effective identification of anomalies, errors and cases of fraud (Desouza, 2014; Grosdhomme Lulin, 2015).

1.2.1 From retrospective to prospective

Today, Big Data analytics and predictive analysis "*are gradually moving closer to artificial intelligence and allow us to move from retrospective to prospective, from observation to anticipation, from analysis to action*" (Grosdhomme Lulin, 2015, p.5). Instead of chasing fraudsters, predictability could be a tool that would allow the State to be proactive, to discover fraud attempts even before the fraud takes place.

Thus, data analysis allows two opportunities for value creation:

- **financial value creation**, as Big Data analytics become a powerful tool in the fight against fiscal and social fraud.
- **improved public safety**, as Big Data analytics enable the government to reduce crime and be more effective in identifying and preventing terrorist threats.

In our presentation, we will focus on the first way to create value: creating financial value by preventing tax evasion. Nevertheless, it is important to note that tax evasion is also a form of crime, although it is a "white-collar" crime and is less often associated with the term "crime". According to Mr. Maes, the problem of prevention (instead of sanctions) is not a technical problem. Algorithms today make it possible, for a whole host of subjects, to prevent fraud, crimes, etc., from being committed. What is missing, according to him, are laws to regulate all that and allow implementation. On this subject, Mr. Aladenise believes that there is a risk of moving from Big Data to Big Brother if we evaluate or predict a citizen's behavior on the basis of personal data. Concerning tax fraud, he also explains that we can defraud tax authorities in good faith, simply because we do not know the latest legislation. He therefore thinks that we can use Big Data techniques to detect and prevent tax fraud beforehand, but that from a legal point of view, it is very complicated to punish someone before he or she has taken action. Mr. Dombret nuances Mr. Aladenise's comments and added that we could use proactivity to better inform people, instead of sanctioning them afterwards. He takes the example of his parents who had big problems with the tax authorities but they were not in bad faith. Using Big Data tools to prevent fraud in a warning and information perspective would have many virtues today.

1.2.2 Cases of prevention against tax and social fraud

Zikopoulos et al (2012) argue that in general, where there is a financial transaction, there is also necessarily a risk of abuse or fraud. In this context, they argue that exploiting the Big Data provides a clear opportunity to identify fraud and, better still, to prevent it. For governments and their tax and social administrations, this is a high value-added opportunity to use Big Data.

In the fight against tax fraud, implementing algorithms that will be able to quickly analyze large amounts of various data (data mining techniques) can enable the government to identify correlations and suspicious activities. For example, the government could (Diez, 2016):

- note that a specific transaction took place at the same time as a change of address.
- discover that a citizen is active in two countries at the same time and regularly carries out transaction patterns that resemble fraud.
- detecting a fraud network and identifying the key players.

The Belgian government is already using these data mining techniques in the fight against tax fraud, as confirmed by Mr. Aladenise. Mr. Dombret gave a concrete example of data cross-checking by the tax authorities. The administration contacted him because there were differences between what Ethias declared to the tax authorities and what he declared as an independent. This was a mistake because what he had declared was VAT excluded whereas Ethias had declared VAT included. But this shows that public administrations cross-check data, even if they are sometimes wrong.

These mass data analysis tools are also used to combat social fraud. For example, with the help of Big Data, a State could detect that a person receives unemployment benefits from one institution and at the same time enters an occupational accident file with another organization. On this subject, Mr. Maes explains that the Belgian State is totally out of touch with social fraud. According to him, many people nowadays benefit at the same time from mutual insurance, family allowances, unemployment, etc. There is far too much *laissez-faire* at this level and Big Data could be a relevant answer to this problem.

Furthermore, in the United States, a system for paying social benefits via prepaid electronic payment cards enables the government to monitor the use of these benefits. In concrete terms, this payment card can be configured to prohibit the spending of these amounts in certain types of businesses, such as alcohol or gambling halls. Even if Mr. Maes did not know about this application, he thinks it could be transposed to Belgium and that it could be beneficial because it is still taxpayers' money. Nevertheless, he points out that we should not go overboard. Mr. Aladenise, for his part, thinks that it is rather complicated to introduce from a legal point of view, because in order to "*limit the rights of certain people, you have to be under provisional administration*". Nevertheless, he is not against the principle, quite the contrary. He also gives the example of eco-cheques in Belgium, which are a means of giving

citizens purchasing power, without them being completely free to use it anywhere. This is partly in line with section 1.4.2, as it is similar to connected objects secured by design. Mr. Dombret does not agree. He believes that with this kind of application, a line is beginning to be crossed. If we allow this kind of thing to happen, what are the next steps? In his view, you have to be able to say “stop” at some point. Mr. Faulkner confirms what Mr. Dombret says, adding that there is a big difference between the cultural approach in Belgium and in the USA. For him, this would be the main obstacle to the use of this tool.

These few examples show that, when used by the State, Big Data Analytics, and data mining techniques, have become formidable weapons in the fight against tax and social fraud.

At the fiscal level, this improved fraud prevention obviously results in increased revenue for the State. When we see that in 2019, in Belgium, 172 billion euros flew to tax havens, we are entitled to say that the Belgian State still has a lot of progress to make in this area (LeSoir, 2020), and that Big Data analytics could be a valuable help. This has been confirmed to us by Mr. Aladenise, who explains that the Belgian State is doing its best to combat tax fraud but does not have enough resources. But he is convinced that we must not give up and not *"stop because we are not doing it right"*. Indeed, Big Data analytics generate a more efficient collection of taxes and consequently an increase, or at least a recovery, of tax revenues. Big Data is a powerful means for tax administrations to reduce the difference between potential and actual tax revenues. The best performing administrations use Big Data to segment taxpayers in order to better organize the collection of taxes. To do this, they bring together geographical, financial, demographic, historical, tax payment fidelity, default risk, and taxpayer collection difficulty data. They then categorize taxpayers in order to better understand the risks and maximize tax revenues.

In concrete terms, in France, for example, through the use of Big Data analytics :

- in one year (2019), Edouard Philippe's French government raised more than 800 million euros thanks to auditors who used the work of a new team of Data Scientists
- half of tax audits are now guided by data mining (LaRevueDigitale, 2020).

Thus, a government that deploys a Big Data strategy creates a real opportunity for value creation in its finances (Grosdhomme Lulin, 2015; Desouza, 2014; Hamel & Marguerit, 2013).

A Big Data tool that Mr. Dombret would like to see, as a freelancer, is a state-level simulator that would allow us to say: *"this is where I'm at in my expenses"*. The data used would then be of the "Volunteered" type as it would be the citizen himself who would enter them into the application. This would save people time and money, and it would avoid having to call in accountants, because, according to him, the rules are not always clear. He thinks there could be a way to create something more dynamic so that we could see in real time whether we're on the right track, instead of waiting a year and sometimes being punished afterwards.

According to Mr. Maes, the problem of the Belgian State is not that it does not put enough means in place to combat tax fraud, but that it has the wrong target. It is focusing on SMEs, but the large sums of money that are flying out are coming from multinationals. But on the other hand, he thinks that the fight against tax evasion at the local level (of a country) is rather naive in today's context. As they do not know how to fight against large-scale tax evasion in a globalization context, due to tax havens and power of multinationals, they target individuals. Which in the end is not a good strategy. Mr. Aladenise does not quite share the same view and believes that fraud at the level of SMEs also exists and must be combated. Nevertheless, he confirms that fraud against multinationals is a more complex matter because Belgium, as a country, is not armed to fight large groups. He gives the example of the cases of Ryanair and Google, which are being sued by the EU for illegal tax advantages granted by Ireland. In each case, the large groups won. Even if Big Data manages to put in place tools that could detect tax evasion, the problem that then arises is the power of multinationals compared to countries and international organizations. Mr. Dombret completes this analyze by explaining that today we are in a two-speed society with, on the one hand, very large-scale fraud, which we are struggling to fight, and, on the other hand, small-scale fraud, which we are much better able to detect today compared with 30 years ago. The problem of tax evasion by large multinationals should be tackled not at the level of individual countries, but at the level of Europe and international organizations. Mr. Faulkner makes a good summary of this, explaining that tax evasion is a battle to be fought, among others, but that we should not be obsessed with it. There is risk, that by using too much data, too many files and monitoring companies, of creating rejection by companies that would then set up elsewhere than in Belgium. Although he confirms that data-driven techniques could make it possible to combat tax evasion much more effectively.

But the fight against fraud is not the only virtue that the predictability of Big Data can bring. It can also lead to a decrease in crime and terrorism, as we try to demonstrate in Appendix 17. Of course, there are other sectors where predictability could play an important role, such as the health sector, especially in the Covid-19 crisis. Mr. Maes explained to us that there was a lot to do with artificial intelligence and Big Data, which could have helped us to better manage this health crisis.

1.3 Internal optimization

Optimization is the third trend we have noted in the impact of Big Data in the elections sector. We have seen that internal data analysis leads to increased efficiency of internal processes and lower costs. By studying the impact of Big Data on the state in the light of this trend, we distinguish two ways of approaching optimization. The first way is the more theoretical way, which allows us to move from a more or less good organization to an optimized organization, i.e. one from which we derive maximum benefit. The second way of approaching optimization is a much more concrete way. This second type of the optimization is the functioning of the country, which we will illustrate by a very concrete example, the optimization of road traffic, and which creates value for all its users (both citizens and foreigners).

1.3.1 From organization to optimization

Many books exist on managing and improving the functioning of an organization. Nevertheless, we can still see that, in general, it is private companies that best optimize their resources. States are lagging behind when it comes to optimizing operations and Belgium is a perfect example of this. In addition to reference works on the organization of a company, to make it more efficient, Big Data could enable governments to achieve a real "optimum" in terms of operation. Even if the aims of a State and private companies are different, nothing prevents us from imagining a State that would maximize and optimize their resources in the same way as private companies do. The state would not become a private company but could function as such.

The first type is the optimization of internal processes within government administrations which, by deploying Big Data technologies, creates value, particularly by reducing costs. In 2011, McKinsey Global Institute estimated that by using Big Data, government administration in Europe's developed economies could save more than 100 billion euros, just by improving operational efficiency. In addition, the technology consulting giant argues that increased data transparency and application of Big Data analytics can lead to increased productivity and internal efficiency in government, including reducing administrative costs by up to 20%.

This overall optimization of productivity through Data-driven techniques is therefore manifested, among other things, in lower government costs, a decrease due to fewer errors in procedures and the acceleration of administrative procedures thanks to more efficient and faster data exchange, such as a reduction in data search times (MGI, 2011; Desouza, 2014).

On a large scale, this results in an intra-connected system of data that is both voluminous (the entire population of the country is concerned), collected and exchanged ever more rapidly between the different actors, and varied (mixtures of human, legal, financial information, etc.). This transformation of public administration is therefore well in line with the Big Data trend.

1.3.2 The case of road traffic

As we get higher, we realize that the country can also be optimized in its overall functioning, just like the public administration, through the use of Big Data analytics. Big Data analytics would enable a shift towards a more harmonious society, where life for the citizen would be more comfortable. With this in mind, we will take the example of traffic on Belgian roads.

In this respect, it is well known that Belgium, being a real crossroads in the middle of Europe and having the capital of the European Union on its territory, is a country where road traffic is a scourge. According to an INRIX study (2019), the Belgian citizen lost an average of 195 hours in traffic jams in 2018, making Belgium one of the most congested countries in the world (INRIX, 2019). According to the same study, Brussels and Antwerp are among the 220 most congested cities in the world. In this connection, both Mr. Aladenise, Mr. Faulkner and Mr. Maes says: "*For a worker, road traffic is a disaster in Belgium*". These figures obviously do not put Belgium in the spotlight, but where there is a weakness, there is necessarily room for improvement.

Big Data Analytics is a real opportunity for the government to create value in road traffic. Indeed, congestion has multiple negative impacts: long hours lost on the road affect the well-being of citizens, contribute to air pollution and cost billions of euros each year. For Belgium, the cost of this waste of time is estimated at 2.3 billion euros (Hoornaert & Van Steenberghe, 2019). If the use of Big Data reduces national congestion, this will result in value creation for the well-being of the citizen, preservation of environment and an increased economic competitiveness. The question now is: how?

Today, geolocation data, collected via citizens' handheld devices or sensors in cars, offers a unique opportunity for the government to better manage its road network and thus significantly reduce congestion. Linking geolocation data to traffic control systems (red lights) would allow public authorities to make traffic flow more smoothly in congested cities (Diez, 2016). According to Mr. Faulkner, the problem in Belgium is that we are led to believe that there are strategies to improve mobility, but that behind these strategies, it is always necessary to have an operationalization of the strategy, and this is what is not happening in Belgium.

Optimod is an example of an urban congestion prevention project developed by the city of Lyon, France. It was a failure, but other applications of the same type could see the light of day in 2020 (LyonCapitale, 2019). Similarly, researchers at the Royal Swedish Institute of Technology have deployed stream analytics software to improve traffic flow in metropolitan areas and reduce the number of accidents due to congestion (Grosdhomme Lulin, 2015; Desouza, 2014). On this subject, Mr. Aladenise has made us a very interesting revelation. He informed us that the Brussels-Capital Region was currently working on an application of this type. It is a "smart city" project for all types of mobility. Thanks to this future application, citizens will be able to say where they want to go, and the application will say which means of transport is the fastest and cheapest. It is also interesting to note that the development of this application is being done in collaboration with companies such as Google, which agree to share their data with the private sector so that it can develop better public services. This is interesting because it contradicts our assumptions that private companies are quite reluctant to collaborate with the State. This application is expected to come out in 2021-2022.

This optimization of traffic in cities is one of the objectives of the INRIX project "*Data driven: Connecting cars to smarter cities*" (INRIX, 2019, p.1). The vision of this organization is to take advantage of the trend towards Big Data and the IoT to optimize the overall functioning of cities by making them smarter (INRIX, 2017). Nevertheless, Mr. Maes thinks that Big Data is not the only solution to the problem of road traffic in Belgium. He believes that there is also a huge infrastructure problem. Other solutions, such as homeworking or the more massive use of public transport, must be complementary to the use of Big Data to regulate traffic. Mr. Dombret also shares this view and believes that the use of Big Data should be complementary to the improvement of road infrastructure. Mr. Faulkner confirms all this and says: "*I do not believe in an exclusive solution. I do believe in a vector, a spectrum of solutions which, when combined, will begin to significantly reduce congestion and long-term mobility problems. The use of data with traffic recommendation figures is one of them*".

1.4 Objectivation of decision making and judgement

The fourth trend we have identified in the area of elections is objectivation. Big Data allows decision making, judgment and evaluation based on objective data analysis rather than human intuition and reasoning. Within the framework of the State, the analyses of mass data allow "*the possibility of taking public decisions automatically, without human intervention, with immediate effect, on the basis of the data thus collected and processed*" (Grosdhomme Lulin, 2015, p.6). With the expansion of Big Data across different government departments, a new model of governance is emerging: one in which decision-making becomes increasingly objective.

Today, it is becoming possible for government to replace or support human decision making with automated algorithms. As in the electoral sector, the goal is to base decision-making on objective data rather than on intuition and human judgment. Thus, Big Data enable the government to make better decisions and provide the country with more relevant solutions to existing social problems (MGI, 2011; Grosdhomme Lulin, 2015).

We approach this opportunity for objectivation through data analysis from two angles. On the one hand, objectivation and the best resulting decisions can be shared by the government by increasing data transparency: this is Open Data. On the other hand, in the context of law enforcement, the objectivation generated by Big Data can allow a kind of "automatic" law enforcement by programming the objects in this direction (Diez, 2016).

1.4.1 From objectivity to objectivation thanks to Open Data

Today, almost all political decisions are "supported" by an analysis of figures, which gives the impression of taking "objective" decisions. But as we have already mentioned, human discretion will always play a role in decision-making. Nevertheless, it is possible to imagine decisions made by algorithms, which would then be truly "objectivated" if not just "objective".

Another benefit of using Big Data analytics in government is the increased transparency of data (Schenk & Berman, 2015). Increasing data transparency means providing access to Open Data. Indeed, as we saw in the first part, Open Data refers to data that the government makes available to the general public.

Open Data is seen as an important potential added value for the public domain. On this subject, Zuiderwijk and Janssen wrote: *“Governments and politicians all over the world have stated to support the disclosure of open government data to accomplish a wide variety of public values”* (Zuiderwijk & Janssen, 2015, p.1). According to them, public values can translate into better public services, an improved relationship of trust between the state and its citizen and the emergence of social capital.

The Agence Numérique Wallone confirms this trend and states that *"the availability of public data free of charge, but under license, in adapted formats, makes it possible to create new applications that can give real added value to this data (composite applications, dashboards and info visualizations, public transport, mapping, public interest applications, etc.). Private companies and other digital project leaders would find here a way to quickly operationalize their ideas"* (DigitalWallonia, 2015, p.4). Indeed, with access to large public sector databases, it is all players in the economy, both individuals and organizations, that can increase the efficiency of their activities. Thus, all these actors become more economically efficient, which is obviously favorable to the country's economic growth (MGI, 2011).

Making the data available to the state more transparent so that citizens and companies can benefit from it is the vision of Vanagt and Rodriguez, founders of data.be and Open Lex BVBA's. In this trend towards data transparency, the two Belgians are at the origin of two Open Data initiatives. The first, the "data.be" site, created in 2012, is a platform that makes the financial and legal information of Belgian companies available to the general public. The second initiative, rolled out in 2016, is the "lex.be" site, a single interface that allows lawyers, citizens and companies to easily and free of charge search for legal texts and court decisions in their full version (Data.be SPRL, 2020; Open Lex SPRL, 2020). These two Open Data projects, and in particular the second one supported by the European Commission, show that data transparency and the sharing of public knowledge have become new opportunities to create value for citizens and companies alike.

Nevertheless, it is important to point out that for Open Data to be effective, there should be an opening of data in both directions. Indeed, today, data sharing goes much more from the public to the private sector than the other way around. However, for an Open Data policy to work, it would also be necessary to rethink the way in which private companies act, which would have to share the data they collect with the State so that the State could also benefit from it to improve certain public sectors. In a utopian world, we could imagine much more collaboration between private and public companies, especially in everything to do with Big Data. Mr. Maes says on this subject that this is not necessarily an utopia, and that some private companies might be willing to share their data with the private sector in exchange for tax benefits, for example. Mr. Dombret also shares this view and believes that the private sector could share their data but that it will not do so for free. Tax benefits are also, in his view, a good example of counterpart. Mr. Faulkner, on the other hand, is even more surprising because he does not even think that financial compensation is necessary so that private companies agrees to share their data. Once it's a win-win situation, what is called Private Public Profits or Partnerships (PPP), can be envisaged completely free of charge because companies could ultimately also benefit from the improvements that would result from Data-driven governance. As long as it's a win-win situation, according to him, private companies would agree to share their data.

1.4.2 The case of connected objects "secured by design"

The "car that drives itself" is now a reality, as shown by the first Google prototypes (Cadot, 2019). For these cars to work, it is obvious that they must be equipped with all kinds of sensors that analyze the vehicle's environment in order to drive without causing an accident. However, sensors also means data. These sensors transmit the environmental data to an on-board computer, which analyses it and then draws conclusions about how the car should be driven in a particular location. As soon as a large number of such cars are on the road, this will mean that mass data will be produced and transmitted. Since this information is varied (mixture of geolocation, environment, speed, weather, etc.) and can be collected in real time, the dimensions of volume, variety and velocity give rise to Big Data.

How can the government turn this into an opportunity for better law enforcement? For example, by equipping these "intelligent" cars with sensors that would be able, thanks to image recognition systems, sonar and long-range radars, to slow down automatically at the sight of a speed limit sign. These technologies are not fictions, but characterize the vehicle built by the Tesla company, the "Tesla D". In this way, compliance with the law is no longer left to the subjectivity of drivers, but is guaranteed by the objectivity of the data. Mr. Maes has discussed this subject with a car manufacturer and he explains that here too, the problem is not technical, because we are quite capable of building cars that would comply with the law "by design". The problem on the one hand is ethical, with the danger of drifts, and on the other hand legal. A legal framework for this type of object is sorely lacking. A danger he sees is

that, for example, if an autonomous car such as the Tesla D is used, it would no longer be possible to incriminate the driver of the car in the event of an accident. Those who can afford these cars will therefore be privileged over other drivers, because they will be "protected" by the object in a certain way, and thus relieved of responsibility. But Mr. Aladenise, as a lawyer, explains this in a more nuanced way. He explains that there is no problem of legality, because as soon as an object of this type is authorized on the market and you decide to buy it, it becomes the consumer's choice. The problem he sees with this type of object is the lack of enthusiasm of manufacturers, whether in the automotive field or in other models. It would not, in his view, be economically profitable to manufacture objects that would work according to the laws of each country. That's why he doesn't really see a future for this type of object. He gives the example in Belgium of devices to be installed in cars, which would detect the blood alcohol level. He states: *"for 10 years now, there has been talk of installing it for people in the case of a repeat drunk driving offence, and yet nothing is being done"*.

In the same way, a payment card could prevent its user, directly and without any possible dispute, from making a purchase in an unauthorized area. A baffle that no longer exceeds a certain volume after a certain time (10 p.m. in Belgium) to avoid nocturnal noise, is also another concrete example of application. These few examples symbolize a transformation of the State towards an objective model of governance which, in certain areas, allows for better compliance with the law by citizens. If the law is respected, it is therefore not necessarily because the citizen wants it, but because the object he or she is using is configured to do so. In short, Big Data allows an automatic decision-making process that leads directly to action (Grosdhomme Lulin, 2015). In this connection, Mr. Dombret explains that he has a very mixed opinion. He explains that humans must remain humans and that with connected objects secured by design, we touch too much on free will. He fears a society where many such objects would appear. He explains that the "Black Mirror" series enlightened him on this point. He does not want a *"sanitized world"*. On the other hand, he is aware that this is the direction we are heading in, and that we have much less freedom than in the 1970s, for example.

But Mr. Faulkner gave us another perspective on "secure by design" connected objects. There also exist connected objects that would not necessarily be designed to "comply with the law" but simply to "protect people". He takes the example of the health sector, with cardiac regulation systems that track a patient and automatically trigger electrodes, shock waves at the heart to act on the person's voltage and so on. For this kind of object, he sees a great future. But in this case, we are no longer in the objects secured by design to "respect the law". It's interesting to have another approach on this subject.

Increasing the transparency of public data and Big Data analytics for law enforcement show the potential for value creation in the hands of the state that effectively uses the data it has at its disposal. The Big

Data revolution is creating a new model of governance, a more objective and efficient model in which decision-making and judgement processes are "driven" by data analysis.

We have just transposed the trends of Big Data in the electoral sector to the field of public governance. While this list of Big Data applications is not exhaustive, it already gives us a good insight into the possibility of Data-driven governance.

Indeed, the State that succeeds in using Big Data in a pragmatic way can deploy many opportunities for value creation. It has the opportunity to improve productivity within its administrations, to make the country more harmonious and pleasant to live in for citizens and businesses, to provide them with more personalized services and finally to switch to an objective model of governance, more efficient to tackle current social and societal problems.

2. What are the challenges and risks of Data-driven Governance?

Now that we know how government can create value by deploying a Big Data strategy, we are asking ourselves the question of challenges and risks. Firstly, we ask ourselves what are the conditions that the state must meet in order to use big data effectively. Secondly, we look at the extent to which the state's use of big data can entail risks for the citizen.

2.1 Challenges for the State

As we have developed through the trends of personalization, predictability, optimization and objectivation, Big Data is definitely an opportunity to create value for the State. But what are the challenges that the State must face in order to make this value creation possible? What are the conditions for it to work?

According to MGI, *"Public sector leaders who want to realize the value potential offered by Big Data need to address a number of internal and external issues. These include implementing appropriate technology, recruiting and training talented personnel, and managing change within their organizations. Just as important, government will need to use policy to support the capture of value from Big Data and the sharing of data across agencies"* (MGI, 2011, p.63). This quotation highlights three challenges for government: a technical challenge, a skills challenge and a mindset challenge.

2.1.1 A technical and infrastructure challenge

Investment in appropriate technologies to be able to collect, store and ensure the security and confidentiality of data is the first requirement. Given the cost of these investments and the limited public finances, the technical challenge is the first barrier to Data-driven Governance. If the state wants to reap the benefits of Big Data, it must give itself the means to do so by making the necessary investments (Grosdhomme Lulin, 2015). Moreover, these days, we do not see the State investing heavily in new technologies. Indeed, with the Covid-19 crisis, which will cost Belgium around 45 billion euros in 2020, it is unlikely that investing in Data-driven governance is the actual government priority (Lecho, 2020).

2.1.2 A skills challenge

Data-driven Governance will only be possible if the state succeeds in recruiting and training competent people who are able to unlock the value of Big Data. Recruitment and retention of talent is one of the main weaknesses of the public sector in Belgium. Indeed, today, the private sector holds far more financial resources than the state and therefore has the means to attract the best people. In Sweden, for example, the best university students work for the state after graduation (Körner, 2019). It is the whole logic of our liberal system that needs to be reviewed. The state must therefore absolutely remedy this if it wants to make the most of all the data it holds (Grosdhomme Lulin, 2015; Schenk & Berman, 2015).

2.1.3 A challenge of mentality

As we have seen in the area of elections, mentalities can sometimes be a greater obstacle than the first two challenges. Managing change from traditional culture to a new model of governance "driven" by data is the third challenge facing the state (Grosdhomme Lulin, 2015; MGI, 2011). Indeed, the administration is too rigid and needs to shift towards a mentality that promotes dynamism, flexibility and innovation.

In Appendix 18 is a computer graphic from Deloitte University Press (Eggers & Bellman, 2015, Figure 2) which illustrates these three barriers that the state must overcome in order to carry out its digital transformation towards Data-driven Governance. In this illustration, the "people", "processes" and "preparedness" poles represent respectively the challenges of skills, technical issues and mentality. Let's now take a look at the risks of Big Data for the citizen.

2.2 Risks for the citizen

As we have seen, Big Data analytics have many advantages, both for the state and the citizen. It seems that Big Data can lead us to a better world and a more harmonious society. However, as already mentioned in the first two parts, there is also the flip side. Indeed, the risks of Big Data that were mentioned earlier in the field of elections are quite transposable to the field of public governance. Thus, two main risks threaten the system: a risk to data security and a risk to individual liberties. These two risks have one thing in common: first, only the state is in a position to manage them, and second, it is the citizen who is the potential victim (Diez, 2016).

2.2.1 A data security risk

The security and confidentiality of data is a major risk, and the State must do everything in its power to guarantee it. We live in an increasingly digital world and no individual or organization is immune to hacking. Moreover, we have talked about terrorism, but cyber-terrorism also exists. If the state deploys a much greater Big Data strategy, it is clear that it will have at its disposal even greater volumes of sensitive data about its citizens. It cannot therefore afford to allow a malicious person or organization to access this data (MGI, 2011; Grosdhomme Lulin, 2015; Schenk & Berman, 2015). In this situation, the vulnerable party is the citizen, so the state has the responsibility to be the custodian of the data and the enemy is a third person or organization.

To take the example of the car plates that we have to introduce for parking in Brussels, Mr. Dombret explains that, in addition to the danger of data leakage', the danger also comes from the fact that we are obliged to give them away. He believes that we should, at the very least, inform the citizen of what is being done with this data. Furthermore, he thinks that we should leave the choice to the citizen to provide the data or not.

2.2.2 A risk to individual liberties

As far as the risk to individual freedoms is concerned, the vulnerable party remains the citizen, but now the State is both the one that can commit abuses and the one that must protect the citizen from such abuses (Diez, 2016). As mentioned in the first two parts, some of the most fundamental human rights may be threatened by the Big Data revolution, in particular the right to privacy and the right to be different. Therefore, in order to regulate the use of Big Data within government departments and to avoid abuses, the State must put in place new laws and regulations. It must also demonstrate transparency (data use policy, benefits and risks) and ensure that the citizen "*remains the owner of his data and can use it to his benefit*" (Grosdhomme Lulin, 2015, p.8). One of the risks, mentioned in the

interview by Mr. Maes, is to marginalize a certain part of the population by opting for algorithmic governmentality. Mr. Aladenise, for his part, takes the example of the use of Big Data during the Covid-19 crisis. He explains that we now have to fill out forms about our travels, which are then cross-referenced with other data such as health data. All of this is to ensure that everyone is kept under lock.

But Mr. Faulkner has enlightened us a great deal on this subject, about the danger of individual freedoms. He explains: *"So there are people who say, I'm afraid to use applications because I don't want to share my private data. But in this case, I would tell them that you shouldn't use GPS either, or Uber for example, if you're not ready for that. You have to play the game because we're in a society where we're sharing data with applications that are going to provide us with services, so we need private data, that's logic."* What he means is that we must give citizens the choice not to use these applications. They have every right to exclude themselves from the game, but then you have to be prepared to exclude yourself from certain services as well. We can't have it both ways.

To summarize this third part, we can say that the State benefits, with Big Data, from the same opportunities for value creation as the elections sector. Indeed, the trends of personalization, predictability, optimization and objectivation are entirely transposable to the field of public services. However, in order for the deployment of a Data-driven strategy to bring real added value to the citizen and to the public administration internally, several barriers must be overcome: the State must face a technical challenge, a skills challenge and a mentality challenge. Finally, if the benefits of Big Data can be transposed from the electoral sector to the public sector, so too are the risks. This is why the state and its citizens are threatened by a data security risk and a risk to individual liberties.

Appendix 19 contains a computer graphic produced by the TechAmerica Foundation and the IT giant SAP, which summarizes the main points of the presentation in this third part. It presents, visually and with the help of a few figures, the areas where Big Data has real potential for added value and in which the state should invest. In addition, the computer graphics also remind us of the challenges that lie on the road to Data-driven Governance.

3. Audit of the four sectors studied

In order to complete our work, we now wish to audit the four areas discussed, where we have raised the possibility of Data-driven Governance. This audit will be represented by a table in section 3.3. The objective is to answer our research question and whether or not it is reasonable to opt for Data-driven Governance in the following four sectors/areas:

- Education (analyzed through the trend of personalization).
- Tax and social security fraud (analyzed via the predictability trend).
- Road traffic (analyzed via the optimization trend).
- Connected objects "secured by design" (analyzed via the objectivation trend).

3.1 The evaluation grid

This will be based on various criteria, which will be set out in section 3.2. The evaluation of these criteria will be made through the various interviews conducted as well as through our own research work. The context is of course the Belgian territory, the Belgian population and the Belgian State as we define it in the introduction of part 3: Data-driven Governance.

The evaluation criteria will be assessed as follows:

- ☹ = "Bad" level: we will consider that Data-driven Governance is not applicable (for this criterion only) in the proposed sector
- 😐 = "Medium" level: we believe that Data-driven Governance could be applied (for this criterion only) but without conviction, there is still room for improvement in the proposed area
- ☺ = "Good" level: we consider that Data-driven Governance is possible and desirable (for this criterion only) in the proposed sector

The method of evaluating the criteria is based on personal experience. Indeed, during a finance internship at IBA in 2020, Deloitte audited the various sectors of the company as follow. We felt that it was relevant to audit our four sectors, according to the different criteria's, in the same way.

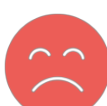
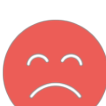
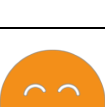
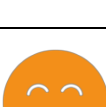
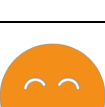
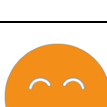
3.2 The criteria used

We decided to select 8 evaluation criteria's in order to audit the four sectors studied in our thesis. The first 6 criteria were inspired by the study conducted by Elisabeth Grosdhomme Lulin (2015) "*Gouverner à l'ère du Big Data*" (Grosdhomme Lulin, 2015, p.1). In her study, she identifies 6 challenges for government to achieve data-driven governmentality. In our thesis, we will use them as evaluation criteria. The 7th and 8th evaluation criteria come from our own reflection. We are now going to explain the 8 criteria selected:

1. **Skills:** this criterion refers to the work of the project owner and the IT project manager. Here we question the capacity of the Belgian State to carry out the processes of developing Data-driven Governance oriented models in the different sectors mentioned (Grosdhomme Lulin, 2015).
2. **The economic model:** the switch to Big Data implies huge investments in data collection. Given the imperative of controlling public spending, we wonder here whether the "financial" means available to the State are sufficient to carry out Data-driven Governance in the proposed sectors (Grosdhomme Lulin, 2015).
3. **The social model:** in order to fully exploit the opportunities offered by data, administrations will have to modernize their managerial practices to be able to manage two types of adjustment.

- a. **Qualitative adjustments**, relating to the skills and qualifications of its agents (recruiting data management "experts", such as data scientist). We therefore wonder about the quality of the personnel available in the government to carry out Data-driven Governance missions in the different sectors (Grosdhomme Lulin, 2015).
 - b. **Quantitative adjustments**, relating to the allocation of potential productivity gains. Here, we question the State's capacity to rationalize and optimize (particularly because of high costs) Data-driven processes as a high-performing private company could do (Grosdhomme Lulin, 2015).
4. **Reliability**: by integrating more and more computer systems into the management of public infrastructures, the State is exposed to technical vulnerabilities and new risks. These may be risks of hacks, but also "choices" to be made regarding what must be done internally with regard to data and what can be outsourced externally. We are therefore wondering here about the ability of the State to protect the citizen regarding the security of their data, by opting for Data-driven Governance in the sectors mentioned (Grosdhomme Lulin, 2015)
5. **Decision-making culture**: when decisions are made by individuals, it is always possible to find a culprit. Once we opt for a policy driven by data and algorithms, it is necessary to be much more transparent and rigorous, because the decisions taken no longer depend solely on people. This also includes equipping our democracy with a capacity to audit (evaluate) the methods used (algorithms), so that all citizens can understand how decisions will be taken. Here we question the capacity of the Belgian State to keep control over decisions and especially its capacity to justify/explain them afterwards, in the perspective of a Data-driven Governance (Grosdhomme Lulin, 2015).
6. **Individual liberties**: the issue is no longer really to prohibit data collection as it has invaded our daily lives, but at least to ensure its transparency (everyone must know who holds what data about them). We therefore wonder here about the ability of the State to conduct a Data-driven policy while ensuring confidentiality and protecting the free will of citizens, which is a fundamental right in our democracy (Grosdhomme Lulin, 2015).
7. **Necessity**: the 7th criterion that we have retained, and which is not mentioned in the challenges by Grosdhomme Lulin, is that of the necessity or not to apply a Data-driven policy in certain sectors of the economy. We have every right to believe that some sectors are functioning properly as they are today and that there is no need for more Data-driven Governance, especially when you consider the huge investments that this requires. We are therefore wondering whether or not it is necessary for the State to opt for Data-driven Governance in the sectors mentioned.
8. **Legality**: given the strong trend of data protection in Europe (GDPR and other laws), we wonder here about the legality of the proposed practices in relation to the different sectors mentioned.

3.3 The final audit table

<i>Sectors</i> Criteria		<i>Education</i>	<i>Tax and social fraud</i>	<i>Road traffic</i>	<i>Connected objects secured by design</i>
1. Skills					
2. The economic model					
3. The social model	a. Qualitative adjustments				
	b. Quantitative adjustments				
4. Reliability					
5. Decision-making culture					
6. Individual liberties					
7. Necessity					
8. Legality					

Conclusion

Before concluding, let us recall our initial problem. As part of the Big Data revolution, we are considering the possibility of a Data-driven governance in Belgium, similarly to what is done in the electoral Big Data sector. We therefore ask ourselves the following question: is it possible to conceive a Belgian State that uses Big Data to improve its governance, as political parties and electoral market players do in order to acquire votes and generate profit?

Our presentation consisted of three parts. In the first part, we presented the context of the Big Data revolution. In the second part, we studied the impact of the use of Big Data during the last French election, where four trends were identified: personalization, predictability, optimization and objectivation. Finally, in the third part, we transposed these trends to the field of public governance and explored how the State could, thanks to Big Data, create value, both for itself and for its citizens. Now we are able to answer our research question: *"On the basis of the major trends identified in the use of Big Data in the electoral sector, what would be the possible applications of a Data-driven governance in some public sectors in Belgium?"*

We decided to answer this question in three steps. First of all, we will try to answer the question globally. Then, we will try to answer the question sector by sector and provide our recommendations. Finally, we will answer the question by looking at the challenges that Data-driven governance represents, both for the citizens and for the state.

The overall answer to the research question is: yes, it is possible for the State to conduct Data-driven governance. Why is it possible? Because the State benefits from the same opportunities to create value as the business do. Nevertheless, it must be stressed that, if the modalities of use are the same (cf. the four trends), the purpose remains different: unlike the business world, the State does not have a profit-making goal. The government exists to serve the citizens and not to generate profit; it has competences and it is important that it does not exceed them. That is why one of our recommendations is to create independent institutes, which are set up for the sole purpose of monitoring political power and preventing abuses. Their role would also be to inform citizens about what is done with their data, personal data, and we will take the example of the car plates that we are forced to give in order to park in Brussels. It is essential that citizen are being informed of what the State intends to do with this data. However, in Belgium, such institutes already exist, such as the Commission pour la Protection de la Vie Privée (CPVP, 2016). The problem that then arises is the independence, sometimes dubious, of such institutes. To close this first part of our conclusion, we ask ourselves the following two questions: how can we ensure that the State continues to act for the public good? Who will monitor and prevent the state from acting in "its own interest"? This subject is particularly relevant in the context of the Covid-19

health crisis. Personal data, such as data on foreign travel, are collected by the State. The use of this data and particularly the collaboration with private companies during this crisis raise doubts among citizens. The figures concerning the Covid-19 crisis, as illustrated in Appendix 20, are rather reassuring, yet the State has decided to opt for stricter measures. The problem therefore no longer comes from objective data in order to carry out data-driven governance, but from the decisions taken on the basis of these data. As Mr. Faulkner confirmed to us, the final decision remains a human decision. In this case, in our opinion, we observe a concrete case of poor decision-making process, despite the availability of objective data. We are therefore entitled to ask ourselves whether sometimes we should not go further than Data-driven governance and opt for algorithmic governmentality, which is the next, even more advanced, step in the use of data.

Now, we will answer the research question, sector by sector, and provide our recommendations, based on our audit table, our interviews, and our diverse and multiple readings.

For the first sector studied, i.e. education, our answer is quite clear: no, today we should not opt for Data-driven Governance in this sector. The key factor that makes us think this is that the state lacks, on the one hand, of skills, and on the other hand, of critical IT infrastructures that are a prerequisite for the use of Big Data in education. Indeed, as virtually everyone we interviewed confirmed to us, schools and universities are sorely lacking in computer tools that would allow them to collect relevant data on students. Nevertheless, most people, and we agree, confirm that investing in schools and universities to equip schools with more digital tools is a priority. Firstly, because today, mastery of computer tools is a major asset when it comes to entering the working world, and secondly, because Big Data techniques, such as a "digital diaries", could be an additional tool for teachers to improve teaching, assuming they are well trained in the use of these tools, of course. At the national level, the majority of the interviewees think that it would not be a good idea to opt for an elitist American-style system where universities would come to recruit students. On the one hand, because it is in contradiction with the "Belgian cultural model", and on the other hand, because it would run the risk of excluding too many students from the school system, who are not all equal in terms of education, nor in terms of their social and family situation. On the other hand, we share the opinion that the role of the state must absolutely be accentuated to inform students about the supply and demand of the labor market as well as the future opportunities of their studies, and this will be our major recommendation for this sector. By cross-referencing labor market data with statistics such as "employment rate according to X or Y academic background", the State could play a major role in improving the labor market situation in Belgium. Moreover, studies are largely state-funded in Belgium. It is all the more not understandable that the State and the federated entities provide so little information to students about job opportunities. We are against the current, rather individualistic tendency to say to students: study what interests you, and you will see what

happens afterwards. That's the speech I heard during my schooling. I think the data now allows us to be much more fact-based and it is imperative to start using it wisely.

For the second sector studied, tax and social fraud, our answer is also quite clear: yes, we must opt for Data-driven governance in this sector. Why is this necessary? Quite simply because the State now has all the data it needs to apply the right tax and to fight against social fraud, whether at the level of raw data, or at the level of algorithmic models and data cross-referencing possibilities. After having interviewed and studied the subject, we are very surprised, even worried, about the lack of responsiveness of politics with regard to social fraud for example. The problem is therefore, according to us, at the level of the decision-making culture. Unfortunately, as for the electoral Big Data sector, any data-driven policy must be accompanied with an efficient strategy. Nevertheless, we have identified two important elements, which qualify our initial comments. On the one hand, we wanted to retain the argument of Mr. Maes. We also believe that in a context of economic globalization, the State alone cannot fight against large-scale tax fraud and confront large multinationals such as GAFAM. This is why our first recommendation for this sector is to fight against tax fraud not at the national level, but rather at the level of multinational or even international organizations. Today, data-driven tools undoubtedly make it possible to combat tax fraud or tax evasion. The real issue underlying this debate is the balance of power between multinationals and States or even supra-state organizations. Since companies are very powerful and also have powerful algorithms capable of optimizing, at the limit of legality, or even evading the tax authorities, we fear that the balance of power is rather to the disadvantage of the States. What should we do about this? Unfortunately, we cannot provide an answer and we see that States cannot either. It is therefore a political problem, which goes beyond the issue of data-driven governance. The second argument that we have decided to adopt is that of Mr. Faulkner, who explains that by wanting to fight tax fraud at all costs, we risk scaring companies away. We did not consider this argument, but it got us thinking. Indeed, tax competition between states, even within the European Union, is a problem. The recommendation we would therefore make is to use the Big Data and Data-driven governance tools, but keeping in mind that Belgium must remain competitive and attractive compared to other countries.

For the third and penultimate sector studied, i.e. road traffic, our answer is much more nuanced: yes, data-driven governance can improve the fluidity of road traffic, but it is not the only solution to the issue. Indeed, as the interviewees confirmed to us, in Belgium there is also a big infrastructure problem and Big Data alone will not fully optimize road traffic. Data-driven governance, which would allow for example the optimization of red traffic lights in real time, depending on the traffic situation, must be a complementary tool to other political actions. The first recommendation we make is therefore quite basic: investing in the road infrastructure and the road network to improve traffic in Belgium. Nevertheless, the fact that we have to invest in infrastructure does not exclude the use of a data-driven

policy. Our second recommendation is therefore: working closely with the private sector, applications such as Waze or tools such as Google Maps could enable the state, via algorithms and cross-referencing of public and private databases, to see where it is a priority to invest and to detect where (from which roads, which routes, which timetables) the main problem of road traffic comes from. In our opinion, such data-driven governance would then make sense. Some interviewees think that this collaboration will not be free of charge, like Mr. Maes or Mr. Dombret, while others, like Mr. Faulkner, think that it is totally possible that this data sharing will be free of charge as long as private companies also have something to gain from it. In both cases, we believe that close collaboration with the private sector is necessary. The third recommendation we make for the mobility sector, and already partially mentioned, is the following: to mix the data available from cars (sensors, weather data etc.) with other initiatives such as homeworking and the increased use of public transport. But here too, we see no contradiction with a data-driven governance policy. Private companies, as well as citizens, could communicate their timetables, places of residence, etc. so that the State can cross-reference the data and create applications to optimize mobility in Belgium. Here too, this requires close collaboration with the private sector. But as Mr. Aladenise told us, such work is already underway, particularly with the "Smart-City" application that the Brussels Capital Region is developing.

As for the fourth and last sector studied, i.e. connected objects secured by design, our answer is also quite clear: no, we should not opt for data-driven governance in this sector. Why not? Because a data-driven governance policy for objects secured by design as we have defined them, that is, designed to comply with the law, is not desirable for two major reasons. The first one is the economic reason. Indeed, as Mr. Aladenise confirmed to us, the problem is not so much technological, but rather economic. To take the example of cars, which would comply with speed limits by detecting speed limit signs, from a technical point of view, it is totally feasible. But the problem really lies in the manufacturing sector, for which Mr. Maes works. Manufacturers have studied the subject and concluded that designing such objects would not be economically profitable at all. Indeed, the design would be very sophisticated, in addition to the difficulties of application. Traffic laws differ in each country. Today, when creating a "manufacturing" products, it is essential not to limit oneself to the Belgian market, otherwise companies would be subject to enormous and potentially fatal economic constraints. The other argument we have retained is the aspect of individual freedoms. Most of the interviewees believe that there is a limit not to be crossed when opting for data-driven governance and that, in this case, it would be crossed. We have to leave a certain amount of free will to people, in order to remain human and not become robots. Nevertheless, we still have a recommendation for this sector: broaden the definition and the spectrum of connected objects secured by design. This is in fact a recommendation that we make to ourselves. We should have broadened the definition of objects secured by design in our interviews. Indeed, objects can also be secured by design for safety or health reasons, not to comply with the law. In this sense, objects such as cardiac regulation systems, which would allow people who use them to be warned when

they make too much efforts, in order to avoid the risk of a heart attack for example, would make sense in our society. They would allow the use and cross-referencing of health data, movement data, heart rate data, etc. via predictive algorithms. This would therefore be used for personal and private purposes and not for public and legal purposes. We are convinced that if we had broadened the definition of this type of objects, the interviewees would have been much less reluctant to the concept.

Finally, we would now like to answer our research question with particular regard to the risks for individuals and the challenges for the state regarding data-driven governance. Our answer is the following: we cannot affirm today, that Data-driven governance is essentially about improving the life of the citizen and improving our democracies. We believe that to answer the question from this angle, we should be looking back about 50 years. Today, we are at the beginning of this kind of practice. That is why we have the humility not to take position. The question we asked ourselves at the end of our work is the following: now that we have observed that data-driven governance is possible in Belgium, is it really desirable?

We will take up the three challenges facing the State, as well as the risks for the citizens, to attempt to answer this delicate question. As far as the technical challenge is concerned, we doubt that the Belgian state is capable of making the right investments. The management of the Covid-19 crisis confirms our intuition. However, we would like to say that this criticism does not only apply to the Belgian state, but to European and even worldwide states. Predictive algorithms could have "predicted" the spread of the virus from China to other countries in the world. However, we have all acted in hindsight and the consequences could be catastrophic for the economy. In relation to the skills challenge, here too, we have a downside. Our intuition, confirmed by the interviewees, is that education is lagging behind in everything concerning the training of Data Scientists, for example. This has repercussions at the state level. On the one hand, the state is not capable and not attractive enough to retain these people and on the other hand, there is a serious lack of people to deploy an operational Big Data strategy at the national level. The traditional mentalities of the political world are obstacles to technological innovation and a transition to a "data-driven culture" would be necessary, and that is the main recommendation we make to the State. Certainly, there are initiatives that are being deployed in the digital area, such as the Banque Carrefour for Social Security or the Tax-on-Web tax service. But will that be enough?

Consequently, we ask ourselves another question: if the Belgian State does not manage to enter effectively into the Big Data trend, could someone else do it instead? Our society is experiencing a very important phenomenon of disintermediation, that companies as Uber, Airbnb, Deliveroo, etc. represent perfectly. Individuals are increasingly being put in contact with each other, some looking for a service and others offering a service. We noticed that all this is being done entirely outside and without the collaboration of the public sector. We therefore wonder whether the State could one day be "uberized",

i.e., in the end, citizens would completely do without public services by providing each other with services.

In the same perspective, we could also be concerned about the economy generated by Big Data, i.e. companies whose business model is based on the resale and reuse of data, such as GAFAMs. These companies are increasingly powerful and often outperform governments. To monitor terrorist threats, the Belgian government has, for example, asked Google help. The question we are now asking ourselves is the following: Data-driven governance, of course, but will this governance be in the hands of the State or could it be privatized? As Körner (2019) explains, the concentration of political and economic power tends to be "*in the hands of small technology elite*" (Körner, 2019, p.6).

As far as the risks for the citizen are concerned, the picture is also somewhat alarming. One major risk we see is that non-democratic countries are increasingly using Data-driven tools to monitor their population. We will take the case of China, with the "citizenship" score attributed to each citizen according to his behavior and which will subsequently influence his fundamental freedoms. But in our Western democracies, where we are given the illusion of increasingly advanced democracies, we observe that, in reality, our freedoms have been steadily diminishing over the last 50 years. Many people miss the time when we could smoke everywhere, when there were no radars, no blood-alcohol tests, when we could moonlighting from time to time. Today all that has become impossible, especially with the increasingly visible disappearance of cash for example. In Europe, we are kings of giving lessons of democracy to other countries (Russia, China, Syria, Libya, etc.). However, everything seems to indicate that behind the democratic facade, we are moving towards an increasingly police state, where everything is under surveillance, largely through the control of personal data. The Covid-19 reflect this perfectly. Looking back, we are today practically "under house arrest" by the Belgian state. Penalties of up to 250 euros can be imposed for journeys that the government considers "unnecessary". Our entourage must be limited to 5 people. To conclude this work, we would like to ask our readers the following question: has the figure of "Big Brother" from Orwell's novel (1984) not become a reality today?

Bibliography

Agoria. (2014). *Big Data : comment le fisc a récupéré plus d'1 milliard d'euros dans la lutte contre la fraude*. Online <https://www.agoria.be/fr/Big-Data-comment-le-fisc-a-recupere-plus-d-1-milliard-d-euros-dans-la-lutte-contre-la-fraude>, accessed on the 7th of May 2020.

Apache Hadoop. (2020). *Hadoop*. Online <https://hadoop.apache.org/>, accessed on the 17th of April 2020.

Bajos, S. (2020, 2 mai). Le site anti- fake news du gouvernement fait grincer des dents. *Le Parisien*. Online <https://www.leparisien.fr/culture-loisirs/tv/le-site-anti-fake-news-du-gouvernement-fait-grincer-des-dents-02-05-2020-8309730.php>, accessed on the 18th of Mai 2020.

Bastien, L. (2018). *Hadoop, tout savoir sur la principale plateforme Big Data*. Online <https://www.lebigdata.fr/hadoop>, accessed on the 6th of April 2020.

Bastien, L. (2018). *Big data : le volume des données mondial multiplié par 5 d'ici 2025*. Online <https://www.lebigdata.fr/big-data-2025-idc#:~:text=A%20l'heure%20actuelle%2C%20en,soit%20175%20milliards%20de%20t%C3%A9raoctets>, accessed on the 17th of April 2020.

Beal, V. (2020). *Unstructured data*. Online https://www.webopedia.com/TERM/U/unstructured_data.html#:~:text=Experts%20estimate%20that%2080%20to,than%20structured%20databases%20are%20growing, accessed on the 19th of April 2020.

Béglé, J. (2017, 18 avril). Comment peut-on prédire une élection ? *Le Point*. Online https://www.lepoint.fr/politique/comment-peut-on-predire-une-election-18-04-2017-2120434_20.php, accessed on the 8th of June 2020.

Belga. (2020, 27 février). 172 milliards d'euros se sont envolés de la Belgique vers des paradis fiscaux en 2019. *Le Soir*. Online <https://www.lesoir.be/283075/article/2020-02-27/172-milliards-deuros-se-sont-envoles-de-la-belgique-vers-des-paradis-fiscaux-en#:~:text=Au%20cours%20de%20l,communiqu%C3%A9s%20par%20l'administration%20fiscale>, accessed on the 5th of May 2020.

Belga. (2019, 23 Janvier). L'évasion fiscale coûte chaque année 30,4 milliards à la Belgique. *7 sur 7*. Online <https://www.7sur7.be/economie/l-evasion-fiscale-coute-chaque-annee-30-4-milliards-a-la-belgique~aa972b20?referrer=https://www.google.com/>, accessed on the 9th of July 2020.

Blosh, M. (2020, 27 février). Sondage, la popularité moyenne de Macron au plus bas depuis mai 2019. *Le Journal du Dimanche*. Online <https://www.lejdd.fr/Politique/sondages-la-popularite-moyenne-de-macron-au-plus-bas-depuis-mai-2019-3952096>, accessed on the 17th of June 2020.

Bizet, (2018). Le Big Data est-il l'Eldorado des gouvernements ouverts ? *Revue internationale des droits des données et du numérique*, 4.

Boireau, M. (2017). D'où vient l'expression les éléphants du PS. Online <https://www.caminteresse.fr/economie-societe/dou-vient-l-expression-les-elephants-du-ps-114373/>, accessed on the 8th of May 2020.

Bourany, T. (2018). Les 5 V du Big Data. *Regards croisés sur l'économie*, 2018/2 (n°23), 27-31. Online <https://www.cairn.info/revue-regards-croises-sur-l-economie-2018-2-page-27.htm>, accessed on the 27th of January 2020.

BrightPlanet. (2020). *About us*. Online <https://brightplanet.com/about-us/>, accessed on the 18th of March 2020.

BrightPlanet. (2020). *Structured vs. Unstructured Data*. Online <https://brightplanet.com/2012/06/28/structured-vs-unstructured-data/>, accessed on the 19th of July 2020.

Cadot, J. (2019). *Google lance ses voitures complètement autonomes sans supervision en Arizona*. Online <https://www.numerama.com/vroom/560079-google-lance-ses-voitures-completement-autonomes-sans-supervision-en-arizona.html>, accessed on the 18th of June 2020.

Cambridge Dictionary. Algorithm. Dans *cambridge Dictionnaty*. Accessed on the 2nd of January 2020 from <https://dictionary.cambridge.org/dictionary/english/algorithm>

Cardon, D. (2015). *A quoi rêvent les algorithmes (1ère éd.)*. Paris, France : Seuil.

Cardon, D. (2018). Le Pouvoir des algorithmes. *Pouvoirs*, 1(194), 63-73. Online <https://www.cairn.info/revue-pouvoirs-2018-1-page-63.htm>, accessed on the 5th of March 2020.

Cardon, D. (2019). Open data et état plateforme. *Culture numérique*, 330-339. Online <https://www.cairn.info/culture-numerique--9782724623659-page-330.htm>, accessed on the 19th of May 2020.

CNCCFP. (2020). *Contrôle des comptes de campagnes*. Online <http://www.cnccfp.fr/>, accessed on the 6th of June 2020.

Coleman, S. (2008). *Doing IT for Themselves: Management versus Autonomy in Youth E-Citizenship (1ère éd.)*. Cambridge, UK : The MIT Press.

Cossé, P-Y. (2018). *Vers un bipartisme à l'américaine ?* Online <https://www.telos-eu.com/fr/politique-francaise-et-internationale/vers-un-bipartisme-a-lamerica.html>, accessed on the 9th of June 2020.

Data. (2020). *About*. Online <https://data.be/fr/content/about>, accessed on the 19 of July 2019.

Davenport, T-H. (2014). *Big Data at Work* (1ère éd.). Boston, USA: Harvard Business Press.

Dayley, A. (2018). *Market Share: Data Quality Tools, Worldwide, 2017*. Online <https://www.gartner.com/en/documents/3882363/market-share-data-quality-tools-worldwide-2017>, accessed on the 13th of January 2020.

Deligia, F. (2019, 19 novembre). Lyon : où en est l'application pour supprimer les bouchons ? *Lyon Capitale*. Online <https://www.lyoncapitale.fr/actualite/lyon-ou-en-est-l-application-pour-supprimer-les-bouchons/>, accessed on the 8th of May 2020.

de Montjoye, Y-A., Radaelli, L., Kumar Singh, V. & Pentland, A. (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347(6221), 536-539. Online <https://science.sciencemag.org/content/347/6221/536#:~:text=We%20study%203%20months%20of,by%2022%25%2C%20on%20average>, accessed on the 5th of May 2020.

De Munk, J. (2017). *Emmanuel Macron, une startup ambiguë*. Online https://www.researchgate.net/profile/Jean_De_Munck/publication/317188907_Emanuel_Macron_une_startup_ambigue/links/592c0ed7a6fdcc44435dfc82/Emmanuel-Macron-une-startup-ambiguee.pdf, accessed on the 8th of April 2020.

Desouza, K-C. (2014). Realizing the Promise of Big Data, Implementing Big Data Projects. Online <http://www.businessofgovernment.org/sites/default/files/Realizing%20the%20Promise%20of%20Big%20Data.pdf>, accessed on the 15th of April 2020.

Diez, S. (2016). *Data Driven gouvernance : hypothèse d'application à l'état des modèles big data déployés par le secteur financier (Mémoire de master)*. HEC, école de gestion de l'ULg, Liège.

Donnat, O. (2007). Pratiques culturelles et usages d'internet. *Culture Etude*, 3(3), 1-12.

Droit, R-P. (2015). *Figures libres. De Big Brother à Big Data*. Le Monde. Online https://www.lemonde.fr/livres/article/2015/04/02/figures-libres-de-big-brother-a-big-data_4607929_3260.html, accessed on the 3th of April 2020.

Durand, C. (2016). *Qu'est-ce que NationBuilder, le CMS qui veut renouveler la démocratie ?* Online <https://www.numerama.com/politique/144986-quest-ce-que-nationbuilder-le-cms-qui-veut-sauver-la-democratie.html>, accessed on the 7th of April 2020.

Eggers, W & Bellman, J. (2015). The journey to government's digital transformation. *A Deloitte Digital global survey*. Online https://www2.deloitte.com/content/dam/insights/us/articles/digital-transformation-in-government/DUP_1081_Journey-to-govt-digital-future_MASTER.pdf, accessed on the 18th of April 2020.

Ehrhard, T., Bambade, A. & Colin, S. (2019). A la conquête des élections 2.0, étude des acteurs et du marché de la technologie politique en France. *Politiques de communication*, 1(12), 97-133. Online <https://www.cairn.info/revue-politiques-de-communication-2019-1-page-97.html>, accessed on the 17th of June 2020.

Europe 1. (2012). *L'UMP "spamme" les boîtes mails*. Online <https://www.europe1.fr/politique/L-UMP-spamme-les-boites-mails-922226>, accessed on the 18th of April 2020.

Essa, Y. (2015). *The 3 V's of Big Data*. Online https://www.researchgate.net/figure/The-3-VS-model-of-Big-Data-9_fig1_302580400, accessed on the 8th of July 2020.

Falque-Pierrotin, I. & Schrameck, P. (2016). *Campagnes électorales : tout savoir sur les règles CNIL et CSA*. Online : https://www.cnil.fr/sites/default/files/atoms/files/guide_cnil_et_csa.pdf, accessed on the 8th of April 2020.

Ferrand, O. (2009). *Moderniser la vie politique : innovations américaines, leçons pour la France*. Online <http://tnova.fr/rapports/moderniser-la-vie-politique-innovations-americaines-lecons-pour-la-france>, accessed on the 7th of April 2020.

Fillieule, O. (2015). *Le désengagement militant*. Paris, France : Belin.

Fontaine, R. (2018). *Apache Hadoop is thriving*. Online, <https://blog.cloudera.com/apache-hadoop-thriving/>, accessed on the 6th of April 2020.

Forbes. (2016, 20 decembre). Big Data, the 6th V everyone should now about. Forbes. Online, <https://www.forbes.com/sites/bernardmarr/2016/12/20/big-data-the-6th-v-everyone-should-know-about/#4acb40e52170>, accessed on the 29th of April 2020.

Forbes. (2016, 20 decembre). Big Data, the 6th V everyone should now about. Forbes. Online, <https://www.forbes.com/sites/bernardmarr/2016/12/20/big-data-the-6th-v-everyone-should-know-about/#4acb40e52170>, accessed on the 29th of April 2020

Erinault, T & Le Bart, C. (2009). *Revue française de science politique*, 59, 629-631.

Gandomi, A & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35, 137-144.

Gantz, J. & Reinsel, E. (2011). *Extracting value from chaos*. Online <https://www.pdfFiller.com/jsfiller-desk10/?projectId=468698265#1df1be028ace0a8aaa359dbb24772b31>, accessed on the 7th of April 2020.

Gartner. Big Data. Dans Gartner Glossary. Accessed on the 5th of February 2020 from <https://www.gartner.com/en/information-technology/glossary/big-data>

Gartner. (2015). Gartner Says It's Not Just About Big Data; It's What You Do With It: Welcome to the Algorithmic Economy. Online <https://www.gartner.com/en/newsroom/press-releases/2015-10-05-gartner-says-its-not-just-about-big-data-its-what-you-do-with-it-welcome-to-the-algorithmic-economy>, accessed on the 18th of March 2020.

Gartner. Internet of Things. Dans Gartner Glossary. Accessed on the 5th of February 2020 from <https://www.gartner.com/en/information-technology/glossary/internet-of-things>

Gartner. Open Data. Dans Gartner Glossary. Accessed on the 5th of February 2020 from <https://www.gartner.com/en/information-technology/glossary/open-data>

Gauthier, N. (2014). *Innovations démocratiques et logiques partisanes : le cas de la campagne Ségolène Royal en 2017*. Université de Droit et de la Santé, Lille.

Géoris, V. (2017, 3 juin). Le meilleur conseil en politique, c'est le big data. *L'Echo*. Online <https://www.lecho.be/dossier/presidentiellesfrancaises2017/le-meilleur-conseil-en-politique-c-est-le-big-data/9900886.html>, accessed on the 8th of June 2020.

Girard, J. (2017). *Quel usage du big data dans les élections présidentielles ?* Online https://www.manageo.fr/blog/data-_bca1/quel-usage-du-big-data-elections-presidentielles-_bar13854.html, accessed on the 3th of February 2020.

Grosdhomme Lulin, E. (2015). *Gouverner à l'ère du Big Data*. Online http://www.paradigmes.com/wp-content/uploads/Gouverner_Ere_Big_Data_IDEP.pdf, accessed on the 7th of June 2020.

Grunberg, G & Haegel, F. (2007). Introduction. *La France vers le bipartisme, 1*, 7-9.

Hamel, M-P & Marguerit, D. (2013). Quelles possibilités offertes par l'analyse des big data pour améliorer les téléservices publics ? *Revue française d'administration publique*, 2013/2 (N° 146), 437- 448.

Hammond, K-J. (2013). The Value of Big Data Isn't the Data. *Harvard Business Review*. Online <https://hbr.org/2013/05/the-value-of-big-data-isnt-the>, accessed on the 18th of March 2020.

Hoornaert, B & van Steenberghe, A. (2019). *Le coût de la congestion du trafic en Belgique Une estimation utilisant le modèle PLANET*. Online https://www.plan.be/uploaded/documents/201909060847190_WP_1909_11966_F.pdf, accessed on the 23th of June 2020.

IBM. (2017). *Making sense of Big Data: a day in a life of an Enterprise Architect*. Online <https://www.ibm.com/fr-fr/analytics/hadoop/big-data-analytics>, accessed on the 16th of May 2020.

IBM Cloud. (2020). *Cloud computing, a complete guide*. Online <https://www.ibm.com/cloud/learn/cloud-computing>, accessed on the 13th of February 2020.

IBM Cloud. (2020). *Data mining, a complete guide*. Online <https://developer.ibm.com/technologies/analytics/articles/ba-data-mining-techniques/>, accessed on the 13th of February 2020.

IBM. (2020). The four V's of Big Data. Online https://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg, accessed on the 3th of February 2020.

INRIX. (2019). *Global Traffic Scorecard*. Online https://static.poder360.com.br/2019/02/INRIX_2018_Global_Traffic_Scorecard_Report__final_.pdf, accessed on the 17th of July 2020.

INRIX. (2019). *Connecting cars to smarter cities*. Online <https://inrix.com/resources/data-driven-insights-connects-cars-to-smarter-cities/>, accessed on the 9th of May 2020.

Insee. (2020). *Accueil*. Online <https://insee.fr/fr/accueil>, accessed on the 17th of June 2020.

Ion, J. (1997). *La fin des militants* (1^{ère} éd.). Paris, France : Edition de l'Atelier.

Jehane Noujaim, J. & Amer, K. (2019). *The great hack* [Documentaire]. Netflix.

Jirbandey, A. (2015). *Comment l'email marketing a façonné la politique moderne*. Online <https://fr.mailjet.com/blog/news/email-marketing-politique/>, accessed on the 9th of April 2020.

Kaufmann, M. (2016). *Successes and failures of knowledge management* (1^{ère} éd.). Harrisburg, Pennsylvania : Jay Liebowitz.

Kitchin, R. (2013). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1-14. Online, https://www.researchgate.net/publication/262008155_The_real-time_city_Big_data_and_smart_urbanism, accessed on the 1st of July 2020.

Korner, K. (2019). *Digital Politics : AI, big data and the future of democracy*. Online : https://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000497768/Digital_politics%3A_AI%2C_big_data_and_the_future_of_d.pdf, accessed on the 9th of April 2020.

Kreiss, D. (2012). *Taking Our Country Back: The Crafting of Networked Politics from Howard Dean to Barack Obama* (1st ed.). North Carolina, USA : Oxford Scholarship Online.

Lambert, R. (2012). *Le financement de la campagne des candidats à l'élection présidentielle au travers des comptes de campagne*. Online <https://www.conseil-constitutionnel.fr/nouveaux-cahiers-du-conseil-constitutionnel/le-financement-de-la-campagne-des-candidats-a-l-election-presidentielle-au-travers-des-comptes-de>, accessed on the 4th of May 2020.

Laney, D. (2001). *3D Data Management : Controlling Data Volume, Velocity, and Variety*. Online <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>, accessed on the 2nd of July 2020.

La revue du Digital. (2020). *Lutte contre la fraude fiscale : 800 millions d'euros encaissés grâce au Data Mining*. Online <https://www.larevuedudigital.com/lutte-contre-la-fraude-fiscale-800-millions-deuros-encaisses-grace-au-data-mining/>, accessed on the 19th of June 2020.

Leroy, S. (2020, 08 avril). La crise du Covid-19 coûtera 45 milliards d'euro à la Belgique en 2020. *L'Echo*. Online <https://www.lecho.be/dossiers/coronavirus/la-crise-du-covid-19-couterait-45-milliards-d-euros-a-la-belgique-en-2020/10219584.html>, accessed on the 9th of June 2020.

Leuth, K. (2014). *Why the Internet of Things is called Internet of Things: Definition, history, disambiguation*. Online <https://iot-analytics.com/internet-of-things-definition/#:~:text=But%20the%20actual%20term%20%E2%80%9CInternet,new%20exciting%20technology%20called%20RFID,> accessed on the 19th of May 2020.

Lex. (2020). FAQ. Online <https://lex.be/fr/content/faq>, accessed on the , accessed on the 9th of April 2020.

Loukides, M. (2015). *We need open models, not just open data*. Online <http://radar.oreilly.com/2014/11/we-need-open-models-not-just-open-data.html>, accessed on the 16th February 2020.

Mayer-Schönberger, V & Cukier, K. (2014). *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (1ère éd.). Boston, USA : Houghton Mifflin Harcourt.

McGinty, J. (2016, 14 octobre). Algorithms Aren't Biased, But the People Who Write Them May Be. *The Wall Street Journal*. Online <https://www.wsj.com/articles/algorithms-arent-biased-but-the-people-who-write-them-may-be-1476466555>, accessed on the 17th of March 2020.

McKinsey Global Institute. (2011). *Big data: the next frontier for innovation, competition and productivity*. Online [https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx#:~:text=%E2%80%9CBig%20data%E2%80%9D%20refers%20to%20datasets,store%2C%20manage%2C%20and%20analyze.&text=With%20those%20caveats%2C%20big%20data,petabytes%20\(thousands%20of%20terabytes\),](https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_exec_summary.ashx#:~:text=%E2%80%9CBig%20data%E2%80%9D%20refers%20to%20datasets,store%2C%20manage%2C%20and%20analyze.&text=With%20those%20caveats%2C%20big%20data,petabytes%20(thousands%20of%20terabytes),) accessed on the 9th of May 2020.

McKinsey Global Institute., McKinsey Center for Government & McKinsey Business Technology Office. (2013). Open Data: unlocking innovation and performance with liquid information. Online https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_Open_data_FullReport_Oct2013.ashx, accessed on the 8th of March 2020.

Medina, E. (2015). Rethinking algorithmic regulation. *Kybernetes*, 44(6), 1005-1019. Online https://www.researchgate.net/publication/283104626_Rethinking_algorithmic_regulation, accessed on the 14th of January 2020.

MongoDB. (2020). *What is noSQL ?* Online <https://www.mongodb.com/nosql-explained>, accessed on the 4th of May 2020.

Moysan, I. (2020). *Big data et porte à porte : la nouvelle manière de faire la politique*. Online <https://www.soft-concept.com/surveymag/big-data-et-porte-a-porte-la-nouvelle-maniere-de-faire-de-la-politique.html>, accessed on the 19th of June 2020.

O'Neil, C. (2013). *Cool open-source models ?* Online <https://mathbabe.org/2013/11/27/cool-open-source-models/>, accessed on the 19th of January 2020.

O'Neil, C. (2016). *Weapons of math destruction, how Big Data Increases Inequality and Threatens Democracy* (1ère éd.). Londres, UK : Allen Lane.

Open Knowledge Foundation. (2020). *What is Open ?* Online

<https://okfn.org/opendata/#:~:text=Open%20knowledge%20is%20any%20content,legal%2C%20technological%20or%20social%20restriction,> accessed on the 17th of March 2020.

Owais, S-S & Hussein N-S. (2016). Extract Five Categories CPIVW from the 9V's Characteristics of the Big Data. *International Journal of Advanced Computer Science and Applications*, 7(3), 254-258. Online

https://thesai.org/Downloads/Volume7No3/Paper_37-Extract_Five_Categories_CPIVW.pdf, accessed on the 12th of March 2020.

Oxford Learners Dictionaries. Big Data. Dans *Oxford Learners Dictionaries*. Accessed on the 2nd of January 2020 sur <https://www.oxfordlearnersdictionaries.com/definition/english/big-data#:~:text=%E2%80%8Bsets%20of%20information%20that,benefit%20from%20more%20personalized%20experiences>.

Parti Socialiste. (Réalisateur). (2009). Martine Aubry présente la Coopol. [Video]. France : Dailymotion.

Peugeot, V. (2012). L'ouverture des données publiques : convergence ou malentendu politique ? *Confiance, croyance, crédit dans les mondes industriels*. Online

https://www.academia.edu/2760114/L_ouverture_des_donn%C3%A9es_publiques_convergence_ou_malentendu_politique, accessed on the 15th of March 2020.

Pillon, A. (2016). *Connaissez-vous DigitaleBox, le concurrent français de NationBuilder ?* Online

<https://www.numerama.com/politique/146021-quest-ce-que-digitalebox-le-concurrent-francais-de-nationbuilder.html>, accessed on the 4th of April 2020.

Polony, N. (2005, 24 mai). Référendum 2005, la fracture démocratique. Le Figaro. Online

<https://www.lefigaro.fr/vox/politique/2015/05/22/31001-20150522ARTEFIG00332-natacha-polony-referendum-de-2005-la-fracture-democratique.php>, accessed on the 19th of July 2020.

Rallet, A & Rochelandet, F. (2015). Protéger la vie à l'ère numérique. *Réseaux*, 189, 240.

Richaud, N. (2017, 19 avril). Comment le Big Data s'est invité dans la campagne présidentielle. *Les Echos*. Online

<https://www.lesechos.fr/2017/04/comment-le-big-data-sest-invite-dans-la-campagne-presidentielle-165412>, accessed on the 7th of June 2020.

Rousse, M. (2020). *Internet des objets (IOT)*. Online <https://www.lemagit.fr/definition/Internet-des-objets-IoT>, accessed on the 7th of March 2020.

SAS Institute. (2015). *Visualisation des données : quoi, pourquoi et comment ?* Online

https://www.sas.com/content/dam/SAS/en_be/doc/other1/ebook/sas-data-visualization-ebooklet-fr.pdf, accessed on the 9th of March 2020.

- Schenk, T & Berman, B. (2015). *Big Data and analytics, a research from the center for Digital Government*. Online
http://assets.teradata.com/resourceCenter/downloads/WhitePapers/Public_CIO_Special_Report_Big_Data_and_Analytics.pdf, accessed on the 9th of April 2020.
- Sénécat, A & Motet, L. (2017, 12 juin). Cinq chiffres qui montrent l'échec du FN. *Le Monde*. Online
https://www.lemonde.fr/les-decodeurs/article/2017/06/12/legislatives-cinq-chiffres-qui-montrent-l-echec-relatif-du-front-national_5143199_4355770.html, accessed on the 9th of April 2020.
- Skourletopoulos, G., Mastorakis, G., Mavromoustakis, C., Dobre, C., & Pallis, E. (2018). *Mobile Big Data : a roadmap from models to technologies (1ère éd.)*. Genève, Suisse : Springer.
- Statista. (2020). *Volume of data/information created worldwide from 2010 to 2024*. Online
<https://www.statista.com/statistics/871513/worldwide-data-created/>, accessed on the 5th of March 2020.
- TechAmerica Foundation. (2012). *Dymestifying Big Data: a practical guide to transforming the business of government*. Online file:///Users/alexisverbiest/Downloads/Demistifying%20Big%20Data.pdf, accessed on the 24th of April 2020.
- Ternisien, X. (2011, 12 octobre). 2012 : la bataille du web. *Le Monde*. Online
https://www.lemonde.fr/politique/article/2011/10/12/2012-la-bataille-du-web_1586268_823448.html, accessed on the 8th of April 2020.
- Ter Haar, B. (2018). *Government of the people, by the people, for the people*. Online
<https://www.clingendael.org/publication/government-people-people-people>, accessed on the 8th of May 2020.
- The economist. (2017). *The world's most valuable resource is no longer oil, but data*. Online
<https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>, accessed on the 4th of May 2020.
- Théviot, A. (2019). *Dis moi qui tu es, je te dirai pour qui voter (1ère éd.)*. Lormont, France : Bord de l'eau.
- Théviot, A. (2016). Towards a standardization of digital strategies mimicking the 'Obama model', dossier "Comparing the French and American Presidential Campaigns". *French politics*, 14(2), 158-177.
- Théviot, A. (2018). *Faire campagne sur internet*. Lille, France : Septentrion.
- Théviot, A. (2019). Entretien avec Jacques Priol. *Politiques de communication*, 12, 135-145.

Thomas H. Davenport, T.-H., Barth, P. & Bean, R. (2012). How 'Big Data' is Different. *MITSloan Management Review*, 54, 22-24.

United Nations. (2018). *Etude des nations Unies sur les e-gouvernements*. Online https://www.unescap.org/sites/default/files/E-Government%20Survey%202018_FINAL.pdf, accessed on the 8th of June 2020.

Verstichel, P. (2015). *Open (Big) Data (as a service)*. Online <https://www.digitalwallonia.be/fr/publications/open-big-data-as-a-service>, accessed on the 19th of March 2020.

VRT News. (2012). *Un nombre record d'universitaires au chômage*. Online https://www.vrt.be/vrtnws/fr/2012/02/09/un_nombre_recordduniversitairesauchomage-1-1216039/, accessed on the 16th of April 2020.

Webopia. (2020). *Structured Data*. Online https://www.webopedia.com/TERM/S/structured_data.html, accessed on the 18th of July 2020.

Werly, R. (2019, 10 juin). La France Insoumise, bien trop soumise à Mélenchon. *Le Temps*. En ligne <https://www.letemps.ch/monde/france-insoumise-bien-soumise-melenchon>, accessed on the 9th of June 2020.

World Wide Web Foundation. (2017). *Open Data Barometer 4th Edition*. Online <https://opendatabarometer.org/4thedition/report/#:~:text=a%20Although%2079%20out%20of,the%20official%20open%20data%20portal>, accessed on the 8th of May 2020.

Zikopoulos, P. & Eaton, C. (2012). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data* (1st ed.). New York, USA : Clearance Book Sales

Zubbof, F. (2019). *L'ère du capitalisme de surveillance* (1ère éd.). Paris, France : Zulma.

Zuiderwijk-van Eijk, A & Janssen, M. (2015). Towards decision support for disclosing data: Closed or open data? *Information Policy*, 20, 103-117.

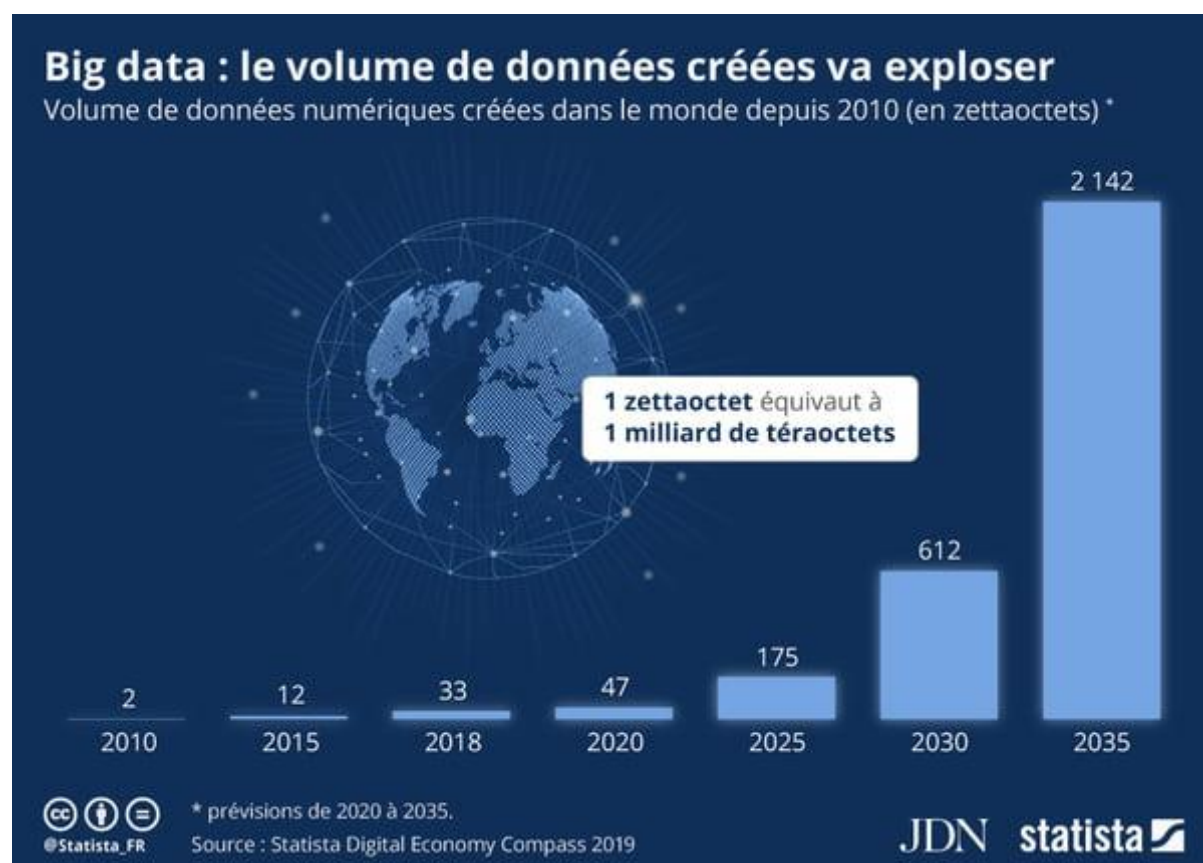
Appendix

Appendix 1: what do we really know about Big Data?



5 Big Data Facts You Need to Know (King, 2015, p. 1).

Appendix 2 : Big Data: le volume de données créées va exploser



Big data: le volume de données créées va exploser (Tasset, 2019, p. 1)

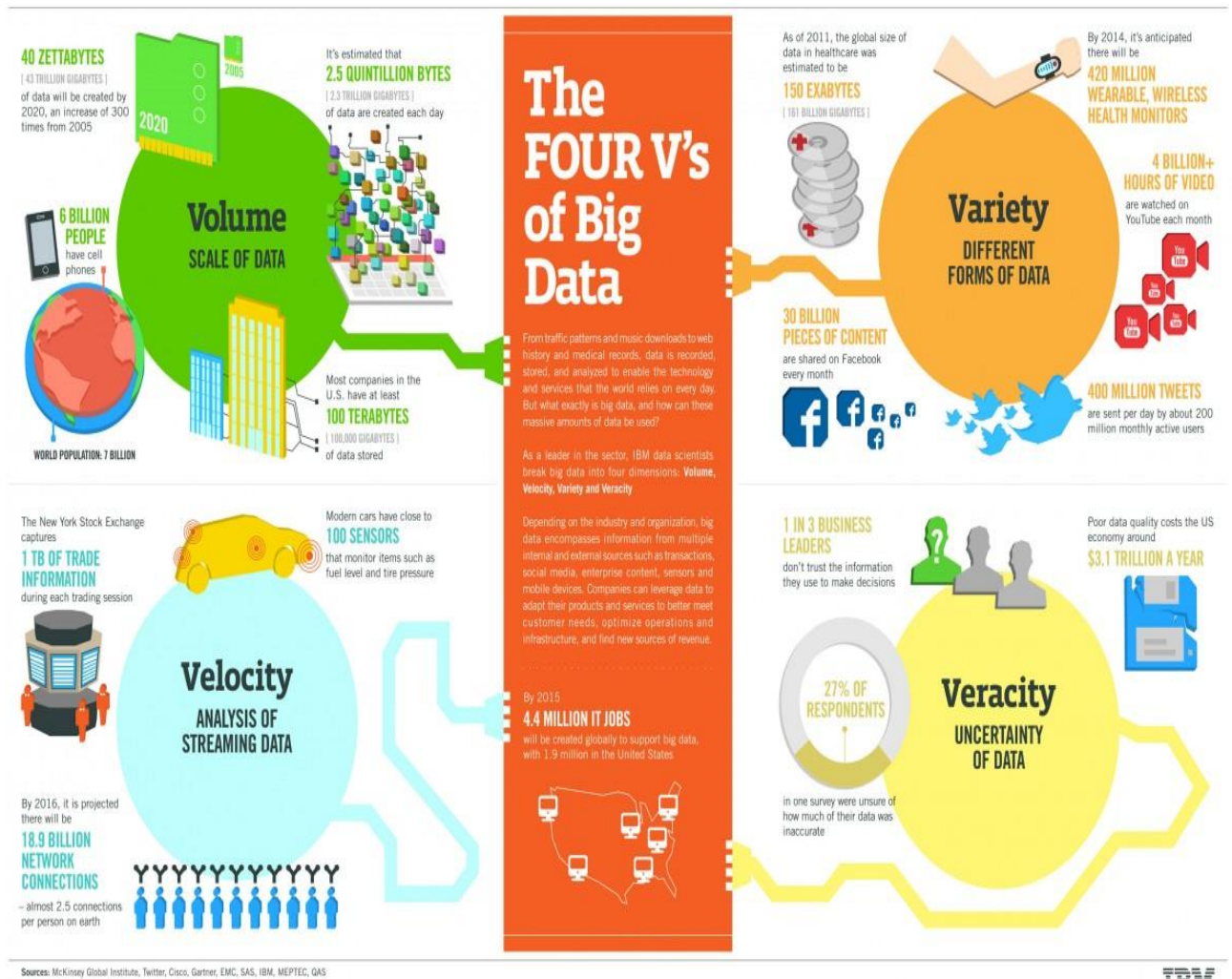
Appendix 3 : 2019, This is what happens in an Internet minute

2019 *This Is What Happens In An Internet Minute*



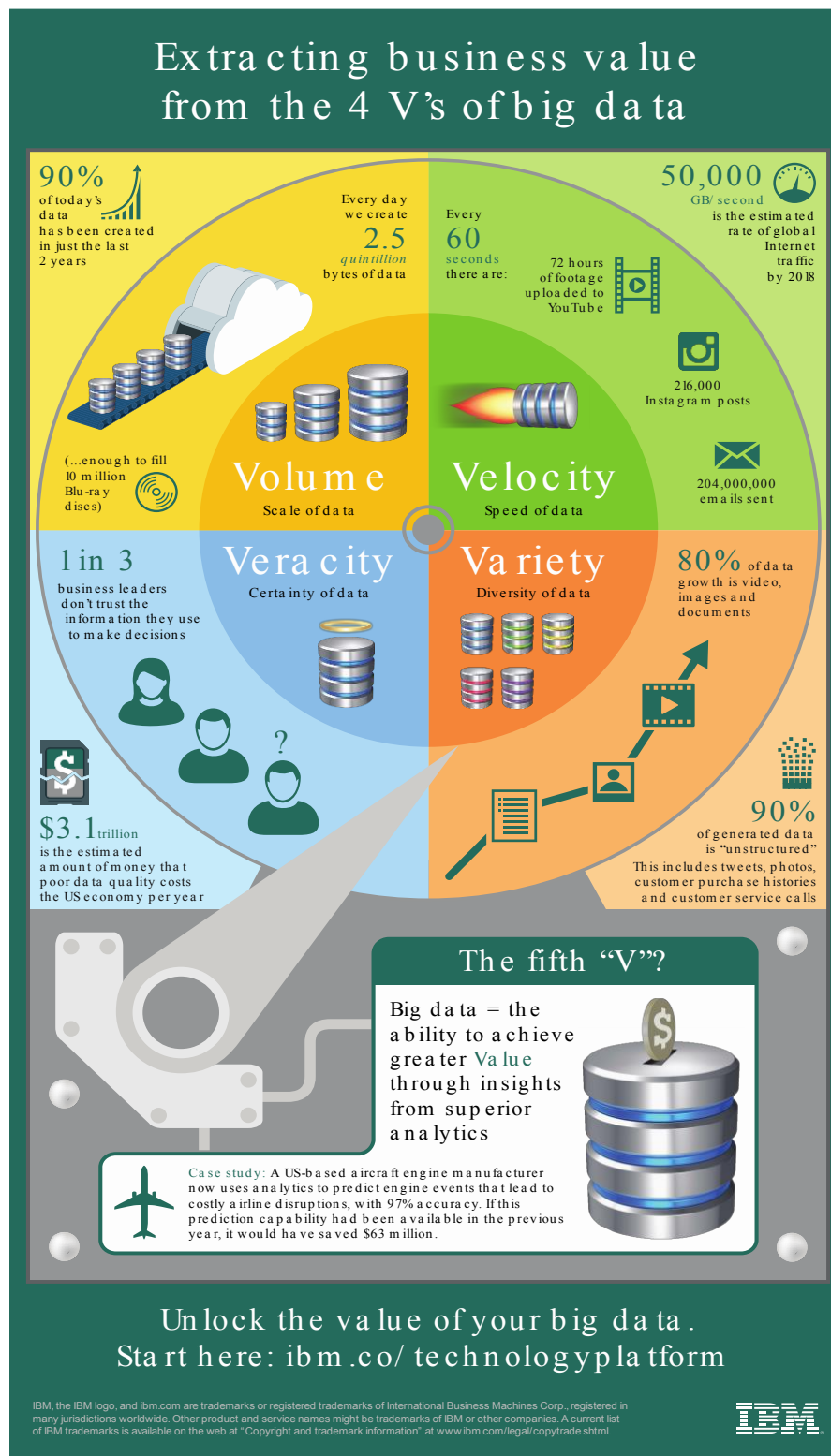
This is what happens in an internet minute (Desjardin, 2019, p. 1)

Appendix 4: The four V's of Big Data

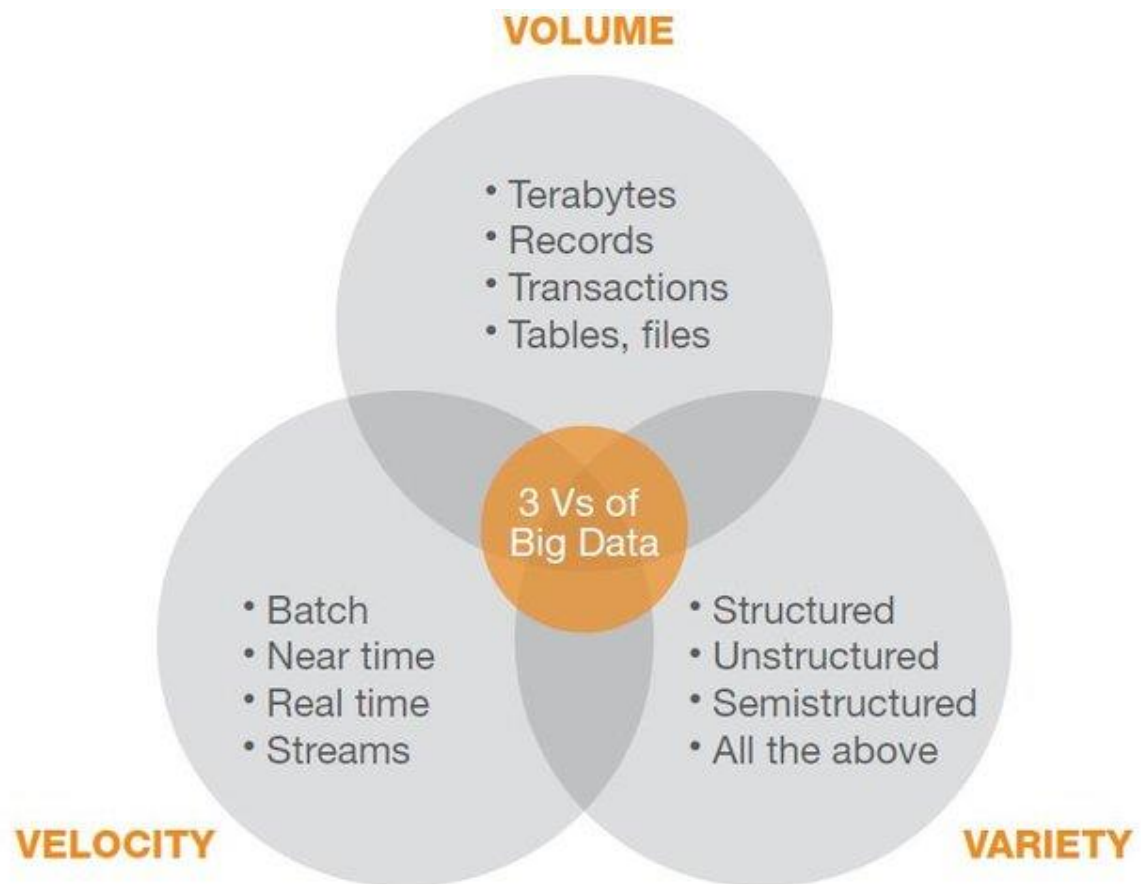


The four V's of Big Data (IBM, 2020, p. 1)

Appendix 5 : Extracting business value from the 4 V's of Big Data



Extracting Business Value from the 4 V's of Big Data (IBM, 2020, p. 1)

Appendix 6 : The 3 V's model of Big Data

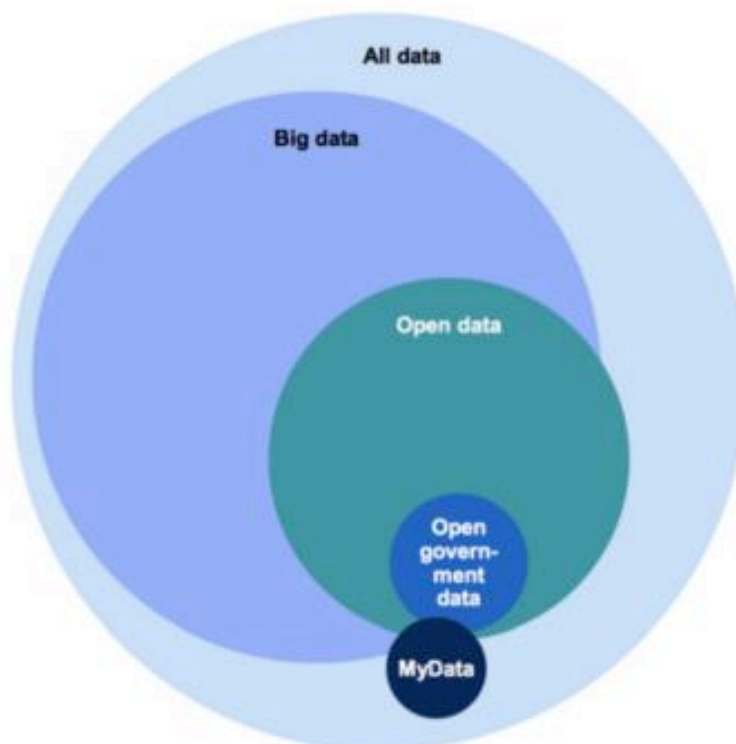
The 3 V's of Big data (Essa, 2015, p 1)

Appendix 7: Open Data & Big Data

Open Data & Big Data

How open data relates to other types of data

ILLUSTRATIVE

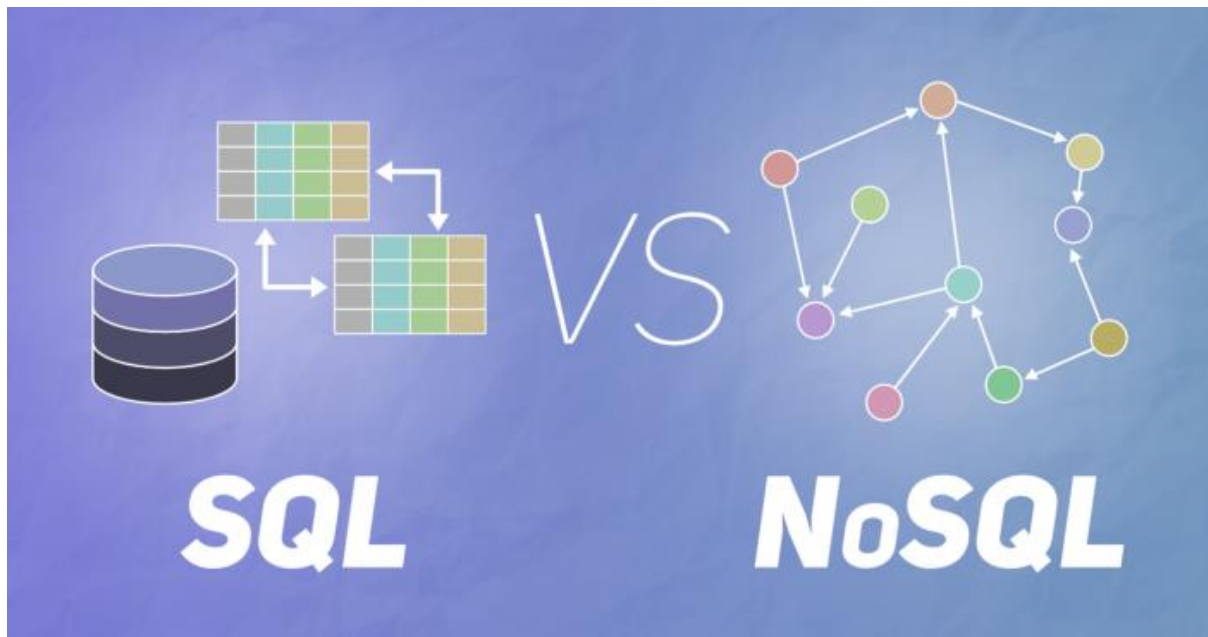


SOURCE: McKinsey Global Institute analysis

McKinsey Global Institute ; Open Data: Unlocking innovation and performance with liquid information (2013, p.4)

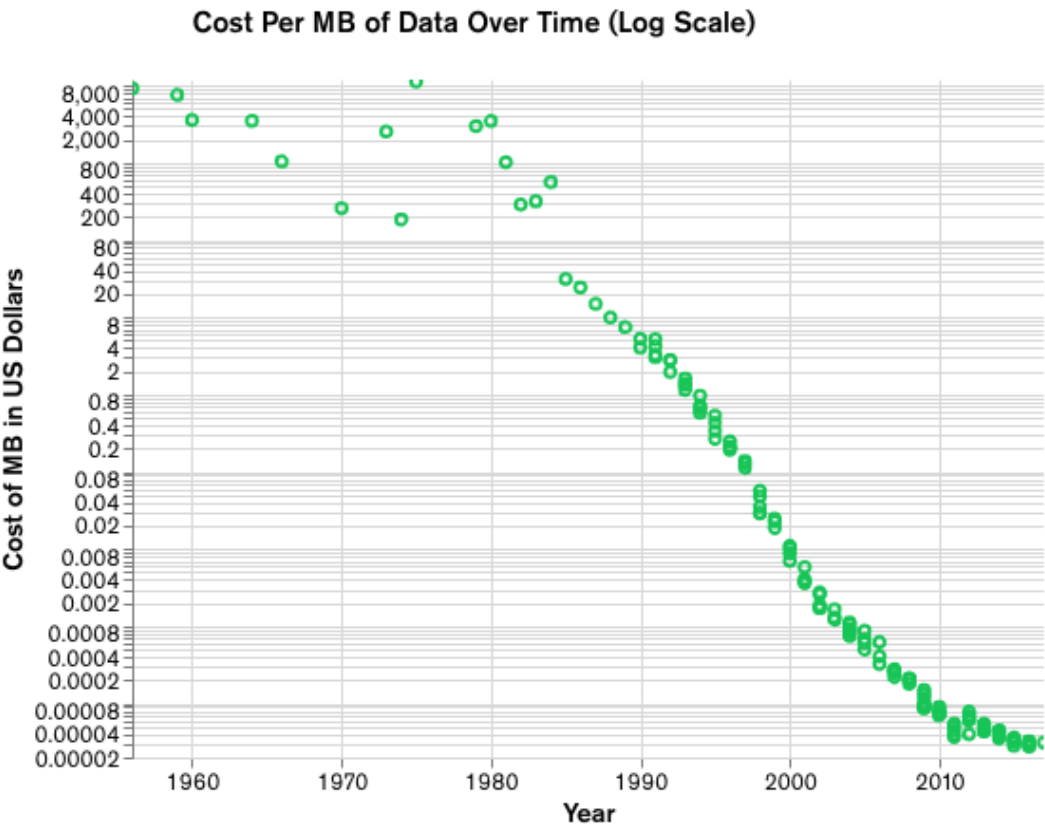
Appendix 8 : Global Hadoop Market Revenue, 2016-2022 (USD billion)

Global Hadoop Market Revenue (Zion Market Research, 2017, p. 1)

Appendix 9 : SQL vs NoSQL

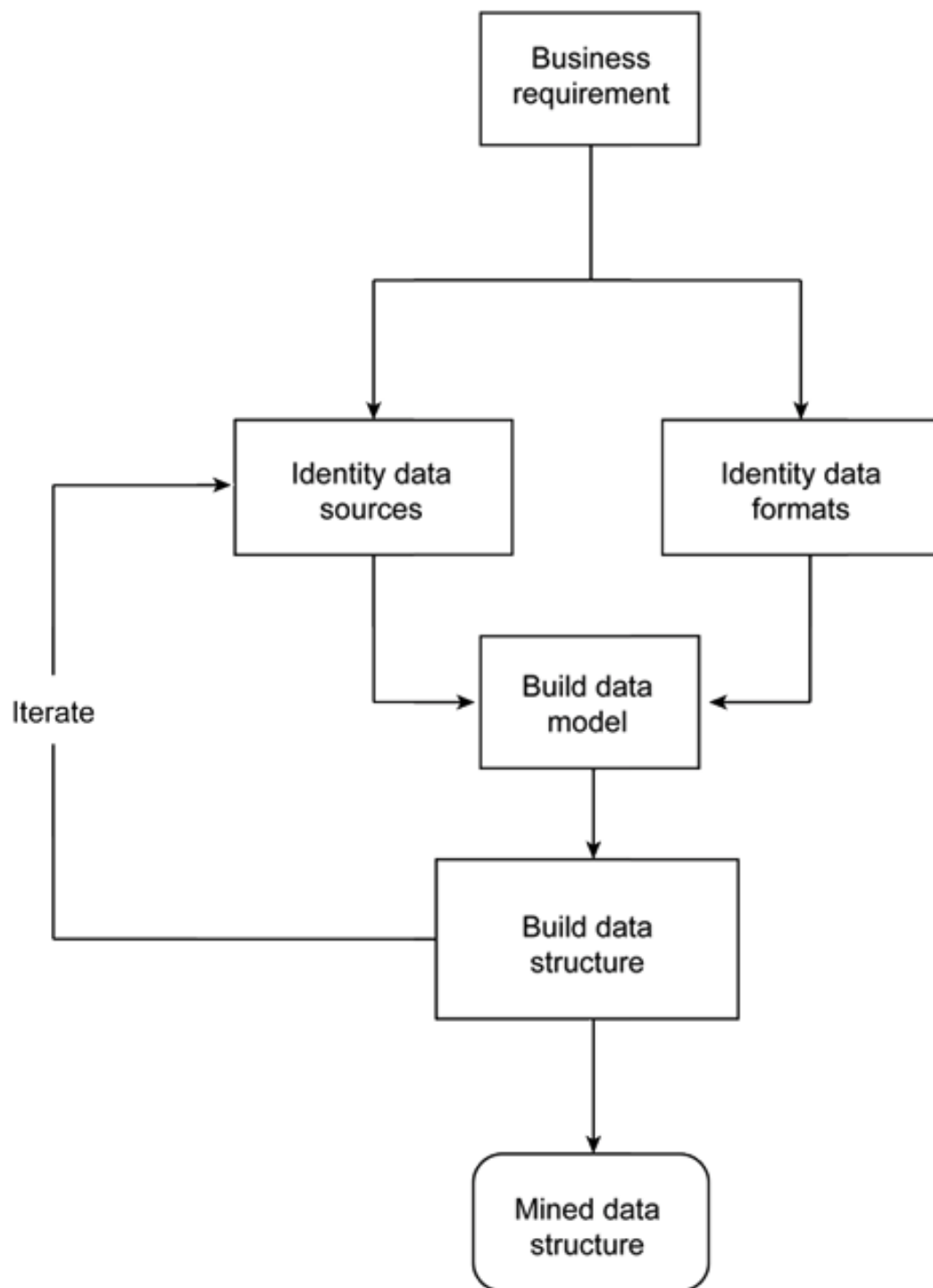
NoSQL vs. SQL: Which is better? (Septiyan, 2019, p. 1)

Appendix 10 : Cost per MB of Data Over Time (Log Scale)



mongoDB.

Cost Per Mb of Data over Time (Log Scale) (MongoDB, 2020, p. 1)

Appendix 11: Outline of the process of Data Mining

Outline of the process of Data Mining (Brown, 2012, p. 1)

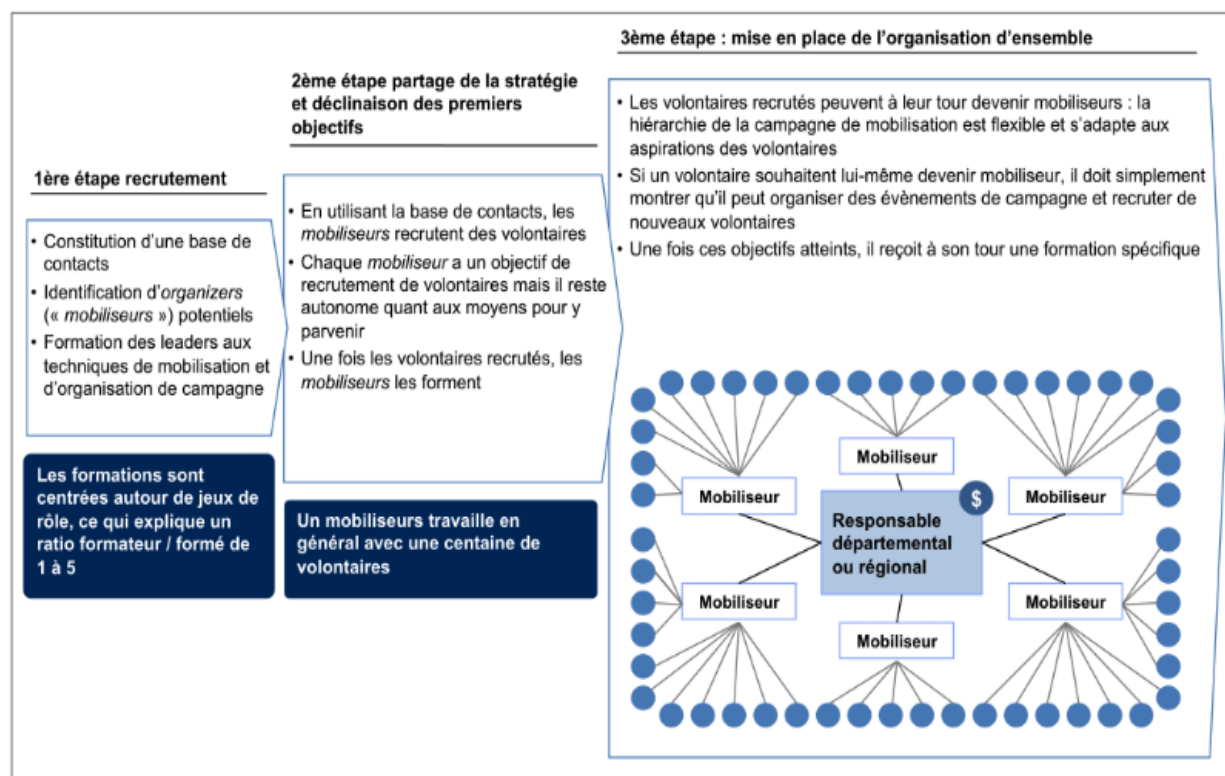
Appendix 12 : Typologie des acteurs des élections 2.0 en France

Tableau n° 1. Typologie des acteurs des élections 2.0 en France.

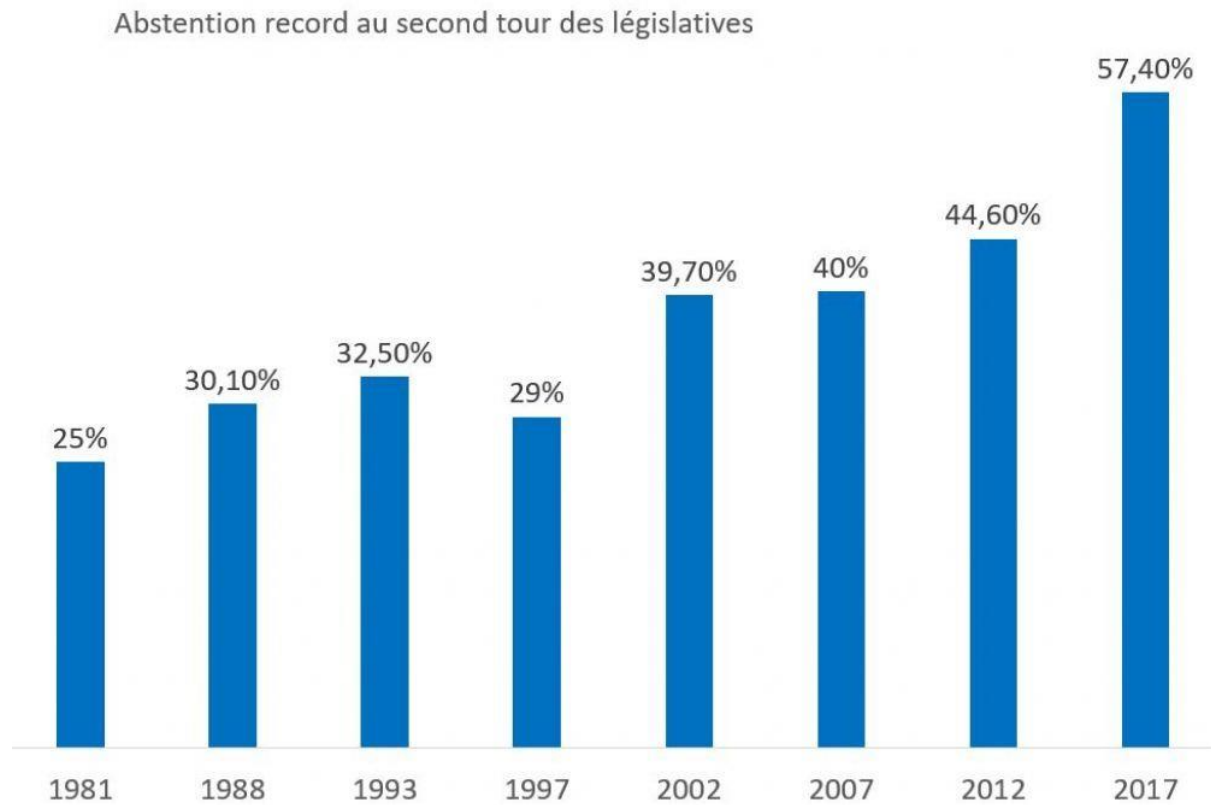
Classe de paramètres	Paramètres	Zuckerman, E.	Veillon, P.-A.	Ecanwasser	DigitalaBox	NationBuilder	Liegey Muller Pons	Spallian	Fédéravox
Services proposés	CRM			X	X	X	X		X
	Carte		X	X			X	X	X
	Targeting		X	X	X		X	X	X
	Modèle de prédiction		X				X		X
	Consulting politique	X					X		X
Accompagnement	Formation logiciel			X	X	X	X	X	X
	Formation terrain						X		
	Éditeur de logiciel			X	X	X	X	X	
Particularité des acteurs	Logique de plateforme				X	X			
	Uniquement politiques		X	X					X
Données utilisées (en plus des résultats électoraux)	Données terrain			X	X		X	X	X**
	Données INSEE		X	X*			X	X	X**
	Listes électorales			X*	X		X	X**	X**
Degré de spécialisation		Acteurs spécialisés			Acteurs relais			Acteurs intégrés	

Typologie des acteurs des élections 2.0 en France (Ehrhard, Bambade & Colin, 2019, p. 102)

Appendix 13 : Les trois étapes de l'organisation du porte-à-porte



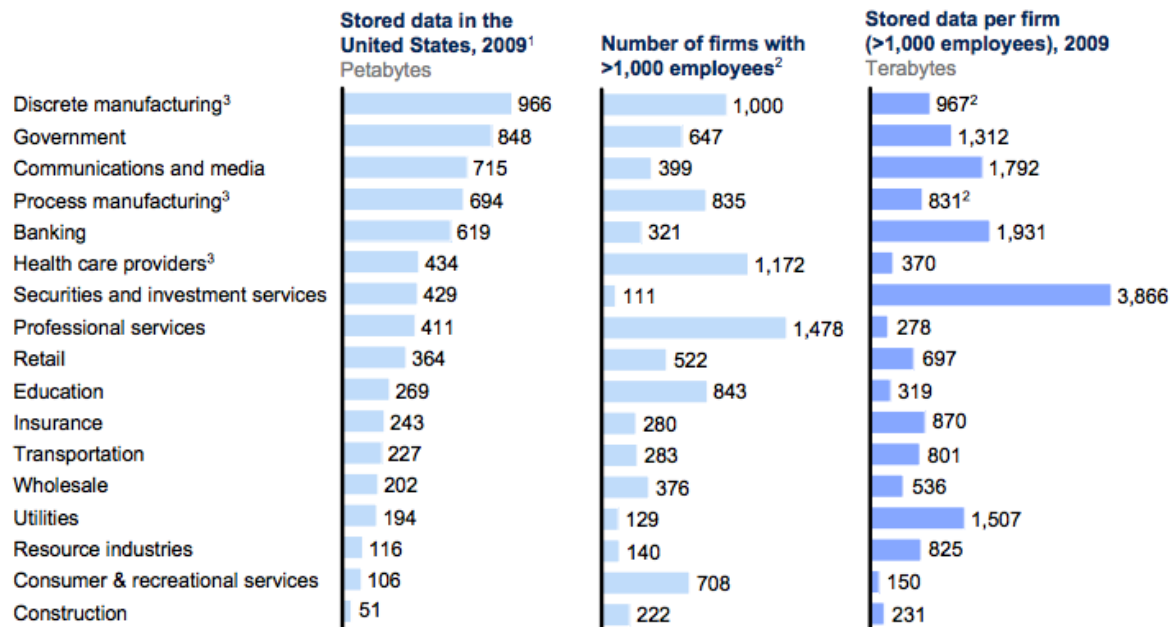
Les trois étapes de l'organisation du porte-à-porte (Liegey, Muller & Pons, 2010, p. 13)

Appendix 14 : Taux d'abstention au second tour des législatives depuis 1981

Taux d'abstention au second tour des législatives depuis 1981 (Challenges, 2017, p. 1)

Appendix 15 : Companies in all sectors have at least 100 terabytes of stored data in the US

Companies in all sectors have at least 100 terabytes of stored data in the United States; many have more than 1 petabyte



1 Storage data by sector derived from IDC.

2 Firm data split into sectors, when needed, using employment

3 The particularly large number of firms in manufacturing and health care provider sectors make the available storage per company much smaller.

SOURCE: IDC; US Bureau of Labor Statistics; McKinsey Global Institute analysis

McKinsey Global Institute “Big Data: The next frontier for innovation, competition and productivity” (2011, p.19)

Appendix 16 : Case study – Big Data and the Public Employment Service

Today, unemployment is undoubtedly one of the biggest challenges facing Western governments. While helping jobseekers find work has become a policy priority, some see the solution in the Big Data. We present here an example of Big Data deployment in the French public employment service, taken from Grosdhomme Lulin's book (2015).

Jobijoba is a "*start-up company in Bordeaux that offers an intelligent job search engine*" (p.50). The founder of Jobijoba started from the observation that in France, an individual looking for a job has to search on many different websites such as Monster, cadreemploi.fr and other more specialized sites. Jobijoba fills a void and offers a unique platform that not only aggregates offers from traditional sites, but also presents these offers in an intelligent way thanks to advanced algorithms. Jobijoba offers three main innovations:

- Jobs are no longer referenced according to the job names filled in by recruiters, but are deduced from the job descriptions. When a job description is published by a recruiter, Jobijoba's algorithms identify the keywords in the description and associate them with already registered keywords. A certain combination of keywords will correspond to a job, the algorithm will make the link and then associate a job to the job description on the basis of its content rather than on the basis of the theoretical name of the job. This way, the job names really correspond to the content of the jobs, which gives more relevant search results.

- Jobijoba's algorithms exploit all the data to make a system of recommendations possible for both employers and jobseekers. On the one hand, employers will receive suggestions based on what other employers are doing (for example, completing a job description with skills that most recruiters require for a similar position). On the other hand, recommendations will be made to the job seeker to increase his or her chances of finding a job (for example, pointing out a skill that recruiters often require for the job sought and that the job seeker does not yet possess).

- Job offers and requests for work are no longer matched using the traditional method that considers occupations and sectors. Algorithms match offers and requests rather according to skills and talents. This allows a job seeker to come across offers he or she would not have come across with the traditional system. It is then a question of searching for jobs on the basis of skills rather than on the basis of job names that are believed to correspond to certain skills.

"Today, Jobijoba works with nine hundred partner sites, references 6 million job offers in eleven countries and receives 4 million unique visitors a month. In France, there are 600,000 job offers from four hundred partner sites" (p.51). With his "Big Data mentality", Jobijoba's founder has created a tool that brings real added value to the public employment service.

And the platform is still developing. New services are being developed, such as a virtual advisor concept to "coach" jobseekers and a labor market analysis service for both supply and demand. Although it is not yet proven that the unemployment rate in France has decreased as a result of this initiative, the case study shows the potential of Big Data to create value by offering the citizen a personalized service that responds more effectively to his or her needs.

Appendix 17 : Case Study – Big Data and Terrorist Threat Detection

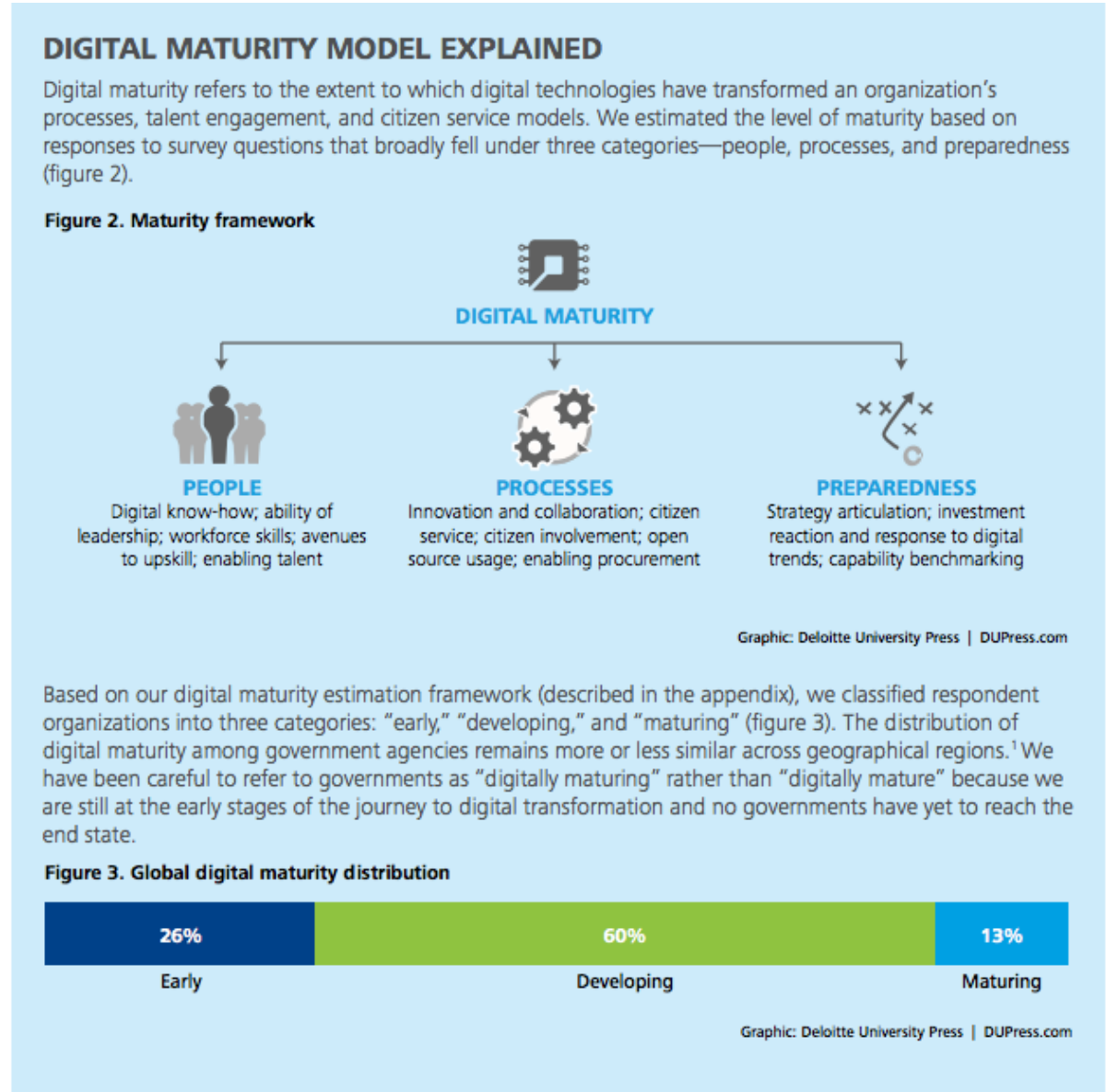
At the beginning of the 21st century, our contemporary society is marked by a war that is a first in the history of Humanity: the war against terrorism. The first outbreaks of this war took place on 11 September 2001 and the more recent attacks in Paris on 13 November 2015 and in Brussels on 22 March 2016 have shown that, unfortunately, this war is far from over. If terrorism is a difficult enemy to fight and defeat, it is undoubtedly because it is hiding among the innocent. It is therefore clear that security has become a major concern, not only for states but for all organizations in general. Indeed, everyone is concerned. Moreover, as we have already discussed, cyber-terrorism also exists.

In this hostile context, as in the case of crime, Big Data Analytics is a wonderful opportunity for the State to improve public security. This is why intelligence and public security agencies are mobilizing to gather large amounts of data from multiple sources in order to identify potential terrorist and cyber security threats. It is within this framework that Big Data analytics can play a key role in identifying and preventing risks and threats to improve public safety. In addition, sophisticated systems of automated algorithms are now capable of analyzing large datasets in order to more effectively detect and counter terrorist threats more quickly.

In concrete terms, we can take the example of the Passenger Name Records (PNR), which became the talk of the town after the attacks in France and Belgium. PNR are passenger data held by airlines. If law enforcement agencies have access to this data, then they can cross-reference it with their own intelligence data. Thus, Big Data analytics applied to this mass of cross-referenced information would make it possible to detect suspicious behavior in air transport (MGI, 2011).

In discovering that Big Data could drastically improve public security, a delicate question can easily come to mind: could the tragic events of 22 March 2016 in Brussels have been avoided if the Belgian State had first deployed a real Big Data strategy?

Appendix 18: Digital Maturity Model Explained

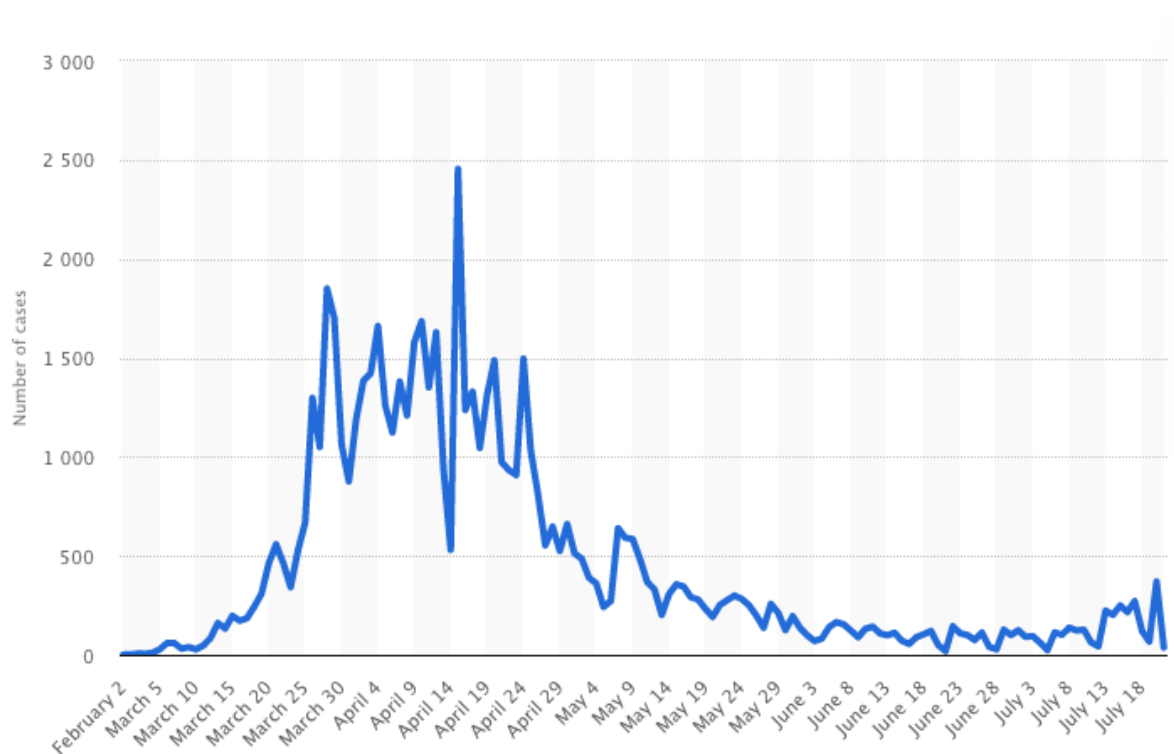


Digital Maturity Model Explained (Deloitte, 2015, p. 2)

Appendix 19: Big Data, Big Possibilities, Big Challenges



Big Data, Big Possibilities, Big Challenges (TechAmericaFoundation, 2020, p. 1)

Appendix 20: Number of new daily coronavirus (COVID-19) cases in Belgium

Number of new daily coronavirus (COVID-19) cases in Belgium (Statista, 2020, p.1)

Interviews

1. Interview 1

Répondant : Monsieur Philippe Maes, dirigeant de la société Wizata active dans l'AI pour le manufacturing.

Alexis : Bonjour Monsieur Maes, je tiens d'abord à vous remercier de prendre le temps pour répondre à quelques questions pour la rédaction de mon mémoire dont le sujet est « Au travers des quatre grandes tendances Big Data identifiées dans le secteur du Big Data électoral (personnalisation, prédictibilité, optimisation, objectivation), nous avons développé des hypothèses d'application d'une possible Data Driven Governance pour les secteurs suivants : l'éducation ; la fraude fiscale et sociale ; le trafic routier ; les objets connectés sécurisés par leur conception », comme je vous l'ai présenté dans mon mail. Le but de ce questionnaire est d'avoir une opinion extérieure, de la part de personnes qualifiées qui présentent un intérêt particulier dans le domaine, sur la possibilité ou non de mener une politique de gouvernance dirigée par les données en Belgique dans les quatre secteurs évoqués ci-dessus.

Philippe : Avec plaisir Alexis, j'ai lu les documents que vous m'avez envoyés, en tout cas ce qui concerne votre sujet, donc je vais essayer de répondre au mieux mais il y a des domaines où je ne suis absolument pas compétent et je vous dirai où c'est.

Alexis : Aucun problème Monsieur Maes, il est très difficile de trouver des personnes vraiment expertes sur le sujet, mon but est donc d'avoir l'avis de personnes qui travaillent dans le domaine de la donnée et qui pourraient un avis plus pertinent et des informations complémentaires.

Philippe : J'espère pour voir vous aider, j'ai appris en lisant votre texte, que le secteur du manufacturing, était celui où il y avait le plus de données, même plus gouvernement. La raison pour laquelle nous travaillons à développer des modèles de Data Driven Decision Making pour l'industrie, c'est qu'il y a aucune notion de vie privée dans les données avec lesquelles nous travaillons, elles ne sont pas non plus anonymisées, ce qui rejoint complètement votre thématique. Il n'y a aucune reconnaissance juridique pour une machine pour l'instant donc cela nous permet d'avoir accès sans aucun problème pour travailler ces données.

Nous avons également été approchées par des compagnies d'assurances et des banques. Ce sont des projets où il y a énormément de choses à faire, comme le gouvernement mais les barrières à l'entrée en terme de confidentialité étaient telles que nous avons immédiatement arrêté de travailler.

Alexis : Ce que venez de m'expliquer est vraiment intéressant, je vais maintenant vous poser ma première question, comment définiriez-vous la notion de « Data Driven Governance », soit la gouvernance dirigées par les données ?

Philippe : Dans l'univers dans lequel je travaille, la Data Driven Gouvernance est un outil supplémentaire pour optimiser des processus de production. C'est une opportunité d'aller au-delà de l'expertise humaine ou des modèles statistiques habituels et de pouvoir prendre des décisions basées sur les données.

Alexis : Quelles est votre opinion concernant la pratique du Big Data dans le cadre d la gouvernance en général ?

Philippe : Dans le cas d'une gouvernance publique si j'ai bien compris, c'est tout à fait pertinent et je pense qu'on peut très certainement améliorer les choses mais il y a énormément de risque liés à l'Etat qui ne basent leur décision que là-dessus.

Alexis : Considérez-vous la notion de « Data Driven Governance » comme une avancée/un progrès pour nos démocraties occidentales ?

Philippe : Je ne sais pas si on peu considérer qu'il y a un lien entre Data Driven Governance et démocratie, c'est à voir.... Mais je pense quand-même que l'on peut considérer que c'est un progrès qui peut nous apporter des choses mais l'utilisation de ces données doit absolument être cadrée par exemple comme la réglementation RGPD.

Alexis : En effet, mais le problème c'est que le Big Data évolue tellement vite, que les lois votées ont toujours un retard considérable par rapport à ce qu'il faudrait, donc elles ne sont souvent obsolètes. Je passe maintenant à la question suivante, quels sont les principaux obstacles et risques qui vous viennent à l'esprit lorsqu'on évoque la gouvernamentalité algorithmique ?

Philippe : Je pense qu'il y aurait un gros risque social, un risque de classification. L'Etat a quand-même un rôle de développer une nation et son peuple, et il y aurait un risque de marginaliser certaine partie de la population.

Alexis : Et Au contraire, quelles sont les qualités que vous pourriez attribuer à une gouvernamentalité algorithmique (s'il y en a), ou du moins, à une augmentation de l'utilisation de données par les autorités publiques ?

Philippe : Là nous sommes dans le vif du sujet au niveau du secteur de la santé mais également au niveau de la sécurité. Avec la crise actuelle du Coronavirus, il y a certainement beaucoup de choses à faire avec des données algorithmiques qui pourraient être très utiles dans la gestion actuelle de cette crise mais également pour le futur.

Alexis : Merci beaucoup pour votre réponse, vous avez par la même occasion répondu à ma question suivante qui était « S’il y a des secteurs/domaines publics où vous accentueriez l’utilisation des données afin d’améliorer la politique menée par le gouvernement et pourquoi ? » Je passe donc à la prochaine question, Quel futur imaginez-vous pour la « Data Driven Governance » ? Pensez-vous que cette tendance va plutôt s’accroître à l’avenir, ou au contraire, que c’est un effet de mode qui va disparaître, et pourquoi ?

Philippe : Je pense que nous n’en sommes qu’au début, je pense que les sources de données vont devenir de plus en plus abondantes et de plus en plus précises. Cela va s’accroître dans tous les domaines et j’espère que l’on pourra compter sur une bonne gouvernance des données.

Alexis : Je vais à présent vous poser quelques questions à propos de l’éducation. Comme vous avez lu la partie de mon travail que je vous ai envoyée, je pense que vous voyez de quoi il s’agit. Je voudrais connaître votre opinion sur le niveau global de l’enseignement en Belgique. Je ne sais pas si vous avez des enfants ?

Philippe : oui j’ai une fille qui est un peu plus jeune que vous, je ne peux pas vraiment vous donner une réponse par rapport à l’enseignement global mais par contre je constate que par rapport à l’enseignement que moi j’ai suivi où on devait emmagasiner les informations. Quand je regarde l’éducation que l’on donne à ma fille, on lui apprend à trouver l’information et à l’utiliser. Ce qui est pour moi quelque chose de très positif. Je suis donc plutôt satisfait.

Alexis : Selon-vous, est-il nécessaire pour l’Etat d’investir massivement dans l’enseignement afin d’équiper les écoles et les universités de plus d’outils numériques, qui sont un préalable à l’utilisation d’une Data Driven Governance dans le domaine de l’éducation ?

Philippe : On parle aujourd’hui beaucoup des STEM (science, technology, engineering and mathematics). Pour moi il faut apprendre aux élèves à utiliser les outils, à coder de manière très simple sans réelle compétence. Il faut en tout cas continuer à investir dans ce domaine, ça c’est certain.

Alexis : Par rapport à mon expérience personnelle, j'ai constaté qu'il y avait une énorme différence entre les humanités et l'université. En humanité, nous n'avons aucun équipement, je trouve ça un peu dommage. Quel est votre avis à ce sujet ? Cela complèterait la réponse à la question.

Philippe : Je suis tout à fait d'accord avec ça et je pense qu'en Belgique il est grand temps qu'on se bouge à ce niveau-là. Quand je compare simplement avec les écoles au Luxembourg, nous sommes très en retard.

Alexis : la question suivante est , des applications Big Data tels que la prédiction des performances d'un élève, l'ajustement en temps réel des programmes scolaires, le repérage de données communes entre les élèves en échec scolaire, etc... pourraient-elles, selon vous, améliorer la qualité de l'enseignement et augmenter le niveau des élèves en Belgique ?

Philippe : En théorie je dirais oui mais en pratique j'ai une crainte par rapport à l'utilisation des données. Est-ce qu'elles vont vraiment être utilisées pour favoriser, aider ou uniformiser l'éducation, ou est-ce que ça va aussi être utilisé aussi pour essayer de déterminer une certaine « élite », une classification et que certaines écoles commencent à avoir des meilleurs scores en gardant les meilleurs élèves et avoir plus de subventions, etc....En résumé, oui en théorie mais plutôt dangereux en pratique.

Alexis : Quel est votre avis concernant une éducation plus « personnalisée » en matière d'enseignement, par opposition à l'enseignement « standardisé » pratiqué aujourd'hui ?

Philippe : Il faut une base commune mais il est clair que théoriquement on pourrait déceler des potentiels plus rapidement . Plus tôt on les décèle, plus tôt sait investir et développer cela. Maintenant, ce dont il faut tenir compte ce n'est peut-être pas simplement la performance académique pure mais aussi voir comment les élèves évoluent au niveau social, collaboratif. Je pense que c'est une partie extrêmement importante de l'apprentissage de la vie.

Alexis : Vous avez donc déjà répondu en partie à la question suivante qui est , en admettant l'hypothèse d'un enseignement « personnalisé » pour les élèves, à quels niveaux de l'enseignement (maternel, primaire, secondaire, universitaire) pensez-vous qu'il serait plus utile de l'appliquer et pourquoi ?

Philippe : oui en effet, comme je l'ai dit il faut commencer le plus tôt possible, donc dès la maternelle et ensuite à chaque stade de l'enseignement. Mais je ne pense pas qu'il faille le faire matière par matière, donc ne plus donner de Math à quelqu'un qui est nul en Math par exemple mais plutôt créer des groupes. Je pense que ce n'est pas utile de faire cela avant la 3^{ème} ou 4^{ème} secondaire.

Alexis : Quels sont, selon vous, les principales qualités et les principaux défaut de l'enseignement à distance ?

Philippe : Alors, je peux vous rapporter l'expérience d'amis qui ont des enfants plus jeunes et certains enfants se sont retrouvés dans une situation d'isolement. Je ne suis pas contre l'enseignement à distance mais plutôt pour un mix si cela se révèle important mais à priori, et certainement pas au début, je ne suis pas du tout favorable à l'isolement qu'entraîne parfois l'enseignement à distance. Il y a toute la vie sociale en communauté dans les écoles, les relations humaines, également les règles à respecter qui sont pour moi très importantes.

Alexis : Merci pour votre réponse, je passe à la question suivante. Dans le cadre de la crise du COVID 19, avez-vous observé/utilisé des applications orientées « Big Data » dans l'enseignement, que ce soit en tant que parent, professeur ou personne neutre ? Si oui lesquelles ?

Philippe : Ma fille a eu des cours par vidéo conférence, des tutoriaux, des vidéos etc...mais on ne peut pas considérer ça vraiment comme du Big Data. Par contre j'ai des amis qui ont des enfants plus grands qui ont passé des examens via une plateforme qui analyse les mouvements de la tête, des yeux, pour essayer de repérer si l'élève triche. Ça je pense que c'est vraiment du Big Data.

Alexis : Oui en effet, il s'agit de la plateforme [Testwe](#), et on peut effectivement dire selon la définition reprise dans mon mémoire que c'est vraiment du Big Data. Quel est votre avis concernant la potentielle utilisation des données des élèves (relevé de notes électroniques, journal de travail numérique, statistiques de réussite) par :

- a. Les universités, afin d'orienter/recruter les élèves dès l'enseignement secondaire et de leur proposer un cursus qui, de par une étude dirigée par des algorithmes, pourrait leur correspondre ?
- b. L'Etat, afin d'orienter les élèves vers des études qui mènent à des débouchés, en fonction de l'offre et de la demande sur le marché de l'emploi

Philippe : En ce qui concerne la partie A, cela existe déjà aux États-Unis, pourquoi pas, cela peut être un indicateur, si les universités viennent recruter des élèves sur base de cela, elles ne vont évidemment pas venir chercher un élève qui est nul en Math pour des études d'ingénieur civile. Je ne suis pas du tout contre mais il faut quand-même laisser la liberté à l'élève de choisir en fonction de ses envies et de ses intérêts. Il ne faut mettre personne dans des « cases ».

En ce qui concerne la partie B, c'est un peu plus compliqué car par exemple actuellement on est à la recherche de 3000 maçons, on ne peut quand-même pas conseiller à 3000 élèves de devenir maçon, ce n'est donc pas si simple. Mais effectivement à contrario, pas mal de jeunes font des études pour

lesquelles il n'y a pas réels débouchée aujourd'hui. Je suis donc d'avis que l'Etat devrait d'une manière ou d'une autre créer une plateforme où l'on pourrait trouver facilement quelle est l'offre et la demande sur le marché de l'emploi. Cela aiderait sans aucun doute certains élèves qui hésitent sur le choix de leurs études. On peut également constater que pas mal de jeunes qui ont terminé leurs études quittent la Belgique pour travailler ailleurs. Ce qui est dommage. Avec une meilleure information du marché de l'emploi cela pourrait être différent.

Alexis : Oui tout à fait, j'ai plusieurs amie qui font des études de psycho et qui ne trouvent même pas de société où faire leur stage, c'est vraiment dommage, il n'y a visiblement pas beaucoup de débouchés là dedans. Je pense que si elles avaient été mieux informées, elles auraient peut-être choisi d'autres études. Et la Belgique perd parfois de très bons éléments qui partent travailler à l'étranger alors qu'il y aurait peut-être eu des possibilités en Belgique.

Merci pour vos réponses, je vais maintenant passer aux questions concernant la fraude fiscale et sociale. Quelle est votre opinion générale sur la fraude fiscale et sociale en Belgique ? Que pensez-vous des efforts faits par le gouvernement belge à ce sujet ?

Philippe : Je pense que l'Etat met beaucoup d'énergie là dedans mais je ne suis pas certains qu'ils prennent les bonnes cibles. Ils devraient s'orienter plus sur les grosses fraudes.

Alexis : Selon vous, est-ce que la lutte contre la fraude fiscale est un combat prioritaire à mener ou plutôt une fatalité dans le contexte de mondialisation et de globalisation de l'économie ?

Philippe : Je pense qu'au niveau local c'est assez naïf. Quand je dis local je parle d'un pays. C'est pour cette raison qu'ils se retournent énormément vers les personnes physiques, les entrepreneurs, les gens qu'ils peuvent cibler et contrôler plus facilement. Je pense donc qu'ils se trompent de cibles mais les cibles qu'ils devraient viser (multi nationales) sont extrêmement difficiles à contrôler car ces grandes entreprises détiennent beaucoup trop de pouvoirs. Il y a des multi nationales qui sont plus grandes que certains pays.

Alexis : Estimez-vous que l'utilisation du Big Data et des données personnelles pourraient être une arme efficace pour lutter contre la fraude fiscale et sociale ? Si oui, pourquoi et comment ?

Philippe : Oui bien entendu, en ce qui concerne la fraude fiscale avec toutes les données qui existent il y a des moyen de creuser et analyser les comportement qui engendrent de là cette fraude mais à nouveau je pense que l'Etat se trompe de cible. Il faudrait faire cela pour les plus grandes entreprises. En ce qui concerne la fraude sociale, c'est la même chose, par exemple pour les gens qui bénéficient du chômage, de la mutuelle, allocations familiales, il y a beaucoup trop de laisser aller de l'Etat à ce niveau-là. Avec

toutes les données qu'ils y a sur les personnes actuellement, l'Etat pourrait certainement les utiliser pour détecter la fraude sociale.

Alexis : A votre connaissance, quelles sont les techniques de Data Driven Governance utilisées par le gouvernement belge aujourd'hui afin de lutter contre la fraude fiscale et sociale ?

Philippe : J'imagine qu'ils font des croisement de bases de données, par exemple quand on fait une vente d'un bien immobilier, tout ça est croisée avec d'autres bases de données. Parfois on a donc des petits retour et des correctifs. J'ai déjà vécu cela. Je n'ai pas de cas concret. Par contre il semblerait que l'Etat va également voir certaines choses sur Facebook mais je ne sais pas si c'est la vérité ou une simple rumeur.

Alexis : D'accord, merci, je passe donc à la question suivante. D'un point de vue pratique, légale et éthique, que pensez-vous d'opter pour une politique proactive en matière de fraude, c'est-à-dire de l'empêcher et de la prévenir plutôt que de la sanctionner, comme c'est le cas aujourd'hui ? Des algorithmes prédictifs permettent aujourd'hui d'identifier les fraudeurs potentiels à priori et non plus a posteriori, pensez-vous que l'utilisation par l'Etat de ce type d'instrument est réaliste et réalisable dans notre société ?

Philippe : La complexité ce n'est pas tant d'utiliser les algorithmes mais plutôt d'établir des lois qui cadrent tout cela. Maintenant commencer à détecter à l'avance, je suppose que cela doit être possible. C'est ce que moi je fais dans le cadre de l'industrie pour les lignes de productions et les machines. Il y a certains comportements qui pourraient être des signaux d'alarme. Cela doit être possible mais je ne pourrais pas du tout vous dire si le gouvernement utilise déjà cela et si c'est vraiment réaliste. Au niveau global cela reste toujours très difficile.

Alexis : Quel est votre avis concernant le système aux États-Unis qui interdit aux personnes bénéficiant d'allocations sociales, certaines dépenses (alcool, salles de jeux) ? Est-ce que ce système serait transportable en Belgique ?

Philippe : Je ne connaissais pas du tout ce système. Je ne savais pas que cela existait. Je pense que ce ne serait pas une mauvaise idée, tant qu'on ne tombe pas dans l'excès. Mais effectivement étant donné qu'ils reçoivent quand-même l'argent des contribuables, je suis effectivement d'avis qu'un certain contrôle des dépenses pourrait être mis en place.

Alexis : Très bien, merci beaucoup. Je vais donc maintenant passer à un autre sujet très sensible en Belgique. Le trafic routier. Quel est votre ressenti par rapport au trafic routier en général en Belgique ?

Philippe : Effectivement, très sensible, j'habite avenue Montjoie et ils viennent de fermer le bois de la Cambre ! Ce n'est qu'un exemple parmi tant d'autres. Pour un travailleur, le trafic routier est une catastrophe en Belgique.

Alexis : L'utilisation intelligente des données disponibles (géographiques, météorologiques, capteurs de vitesse, etc) pourrait-elle être un moyen d'améliorer significativement la fluidité du trafic routier en Belgique selon vous ? Ou pensez-vous que la congestion est plutôt un problème structurel (infrastructures des routes) et que d'autres solutions seraient préférables (home working le matin, transports publics, etc) ?

Philippe : Là encore, il n'y a pas une solution. C'est un ensemble de solutions. L'utilisations des données peut-être utile mais ne pourra pas tout résoudre car il y a aussi de gros problèmes d'infrastructure. A partir du moment où des personnes du gouvernement pensent que mettre une piste cyclable pour faire rouler des vélos sur une route initialement destinées aux voitures, pour moi c'est tout simplement une catastrophe et c'est malheureusement de plus en plus souvent le cas.

Alexis : Afin d'optimiser la Data Driven Governance, nous évoquons le partage de données types « Open Data » du secteur public vers le secteur privé et inversement, ce qui est peu le cas aujourd'hui. Pensez-vous que l'ouverture des données du secteur privé est un impératif pour mener à bien une Data Driven Governance dans le domaine du trafic routier, ou considérez-vous cela comme une utopie ?

Philippe : Effectivement c'est une idée. Si par exemple Waze partageait ses données avec le gouvernement, cela pourrait certainement être utile pour la gestion du trafic routier. Mais je ne pense pas que ces société privées seraient d'accord de partager leurs données gratuitement. Peut-être en échange d'un avantage fiscal ?

Alexis : Dernière question sur le trafic routier. Que pensez-vous de l'option d'une gestion des routes par le secteur privé (comme c'est le cas en France par exemple) ? Quelles seraient les avantages, notamment concernant l'utilisation des données que posséderait le secteur privé par rapport au secteur public ?

Philippe : J'ai vraiment du mal à répondre à cette question car la taille des 2 pays est tellement différente. Je ne sais pas si ce serait économiquement viable On disposerait de plus de données de transit.

Alexis : Je passe maintenant aux questions sur le dernier sujet, les objets connectés et sécurisés par leur conception. Ces objets sont des objets qui, de par leur nature, empêchent l'utilisateur d'enfreindre la loi. Prenons le cas des voitures : certaines voitures sont aujourd'hui capable de limiter la vitesse lorsqu'elles

détectent, grâce à des capteurs, des panneaux de limitation de vitesse. Que pensez-vous de ce genre d'objets (d'un point de vue pratique, éthique et légal) ?

Philippe : C'est vraiment très délicat car effectivement c'est du « contrôle ». A partir du moment où ce sont des choses qui respectent les lois pour des raisons sécuritaires, je peux comprendre que cela soit utile. Le danger ce sont les dérives. J'ai parlé à un constructeur de voitures il y a 4 ou 5 ans, le problème ce n'est pas la partie technologique, la technologie automobile est très développée depuis longtemps. Le problème c'est la partie éthique et légale. Avec des dérives comme par exemple, être prêt à devoir payer plus pour une voiture où le constructeur garantit que le soft va probablement privilégier le conducteur dans certaines situations (accident, limite vitesse ou autre) par rapport à d'autres marques de voitures. C'est vraiment assez délicat.

Alexis : Quel est, selon-vous, l'avenir des objets connectés sécurisés par leur conception en Belgique ? Pensez-vous que de plus en plus d'objets de ce genre vont voir le jour dans un futur plus ou moins proche ?

Philippe : Je n'y connais pas grand-chose le domaine des objets connectés sécurisés par leur conception et je suis d'avis que cela pourrait être utile mais uniquement s'il s'agit de la sécurité de la population. Cela pose donc un vrai problème éthique et légal. Je ne vois pas très bien non plus qui aurait le droit d'implémenter cela. Cela devrait être une collaboration étroite entre le secteur privé et le secteur public. J'ai du mal à voir où ça pourrait mener exactement et n'ai pas le sentiment que de plus en plus d'objets vont voir le jour dans un futur proche. En tout cas pas en Belgique.

Alexis : C'est vrai que c'est un sujet un peu méconnu, mais c'est un secteur qui se développe de plus en plus dans le monde. Passons à la question suivante. Connaissez-vous ou avez-vous entendu parler d'autres projets pour ce type d'objets dans votre entourage ? Si oui, serait-il possible de les décrire et de les expliciter ?

Philippe : Non, ce n'est pas le cas

Alexis : Que pensez-vous d'une société où énormément d'objets de ce genre feraient leur apparition ? Quel serait encore la part de liberté individuelle laissée au citoyen ? Pensez-vous que, par exemple, le citoyen doit parfois pouvoir enfreindre les lois pour se faire entendre (grève de l'impôt, manifestation non autorisées, etc) ?

Philippe : Comme je l'ai dit précédemment, je suis d'avis que ce genre d'objet peut-être utile en ce qui concerne la sécurité des personnes mais pas si cela devait devenir une intrusion dans la vie privée ni un frein aux libertés individuelles de chacun.

Alexis : L'interview touche à sa fin. J'ai cependant encore 2 dernières questions plus formelles à vous poser. Y-a-t-il des informations transmises en réponses à ce questionnaire qui sont confidentielles et ne peuvent être qu'utilisées anonymement pour mon mémoire ? Et Puis-je vous citer dans le cadre de mon mémoire ?

Philippe : Non aucune réponse n'est confidentielle et vous pouvez sans problème me citer dans le cadre de votre mémoire.

Alexis : Merci beaucoup pour vos réponses claires et pour le temps que vous y avez consacré.

Philippe : Avec grand plaisir, j'espère que votre mémoire aura du succès et vous souhaite du courage pour la suite. Au revoir Alexis.

Alexis : Au revoir Monsieur Maes

		L'éducation	La fraude fiscale et sociale	Le trafic routier	Les objets connectés sécurisés par leur conception
1. les compétences		1	2	3	2
2. Le modèle économique		1	3	2	1
3. Le modèle social	Ajustements qualitatifs	1	3	2	1
	Ajustements quantitatifs	1	3	3	2
4. La fiabilité		2	1	1	3
5. La culture décisionnelle		1	2	2	1
6. Les libertés individuelles		1	2	2	1
7. La nécessité		1	2	2	2
8. La légalité		1	2	3	3

2. Interview 2

Répondant : Monsieur Martin de Waziers (founder at Performance in Action & Managing director at Alpha Omega)

Alexis: Bonjour Monsieur de Waziers. Je tenais tout d'abord à vous remercier pour votre aide précieuse. Comme je vous l'ai mentionné dans notre échange de mail précédemment, mon mémoire aborde la possibilité d'une data driven governance, étudié de par les modèles utilisés par le Big Data dans le secteur des élections. Le but de ce questionnaire est d'obtenir des informations complémentaires à mon sujet de recherche. Si vous êtes prêt, on peut commencer avec les questions.

Martin : Avec plaisir ! allez-y.

Alexis : Très bien, cela me va parfaitement. Premièrement, comment définiriez-vous la notion de « Data Driven Governance », soit la gouvernance dirigée par les données ?

Martin : Selon moi, la gouvernance guidée par les données est l'utilisation appropriée des données pour soutenir les activités, grâce à des systèmes qui garantissent des normes éthiques strictes, le respect de réglementations rigoureuses et la protection de la vie privée.

Alexis : Ensuite, quelle est votre opinion concernant l'utilisation de la pratique du Big Data dans le cadre de la gouvernance en général ?

Martin : On ne peut pas aller à contre-courant et imaginer le monde sans ce qui est devenu un élément important dans tous les aspects de la société ; ainsi, je suis tout à fait pour, à la limite de la décence !

Alexis : Considérez-vous la notion de Data Driven Governance comme une avancée ou un progrès pour nos démocraties ?

Martin : C'est, sans aucun doute, une bonne chose tant que les gens ne sont pas paranoïaques et que nous arrêtons d'ajouter des couches stupides et inutiles comme celle de devoir cocher la question RGPD à chaque fois que nous consultons un site web...

Alexis : Quels sont les principaux risques et obstacles qui vous viennent à l'esprit lorsqu'on évoque la gouvernamentalité algorithmique ?

Martin : Aucun à mon avis, si ce n'est la puissance de la machine d'un côté, le pouvoir politique de l'autre (Ironie)!

Alexis : Au contraire, quelles sont les qualités que vous pourriez attribuer à une gouvernamentalité algorithmique (s'il y en a), ou du moins, à une augmentation de l'utilisation des données par les autorités publiques ?

Martin : Elle est fondamentale pour les États et les gouvernements, exactement comme elle l'est pour les entreprises.

Alexis : Quel futur imaginez-vous pour la « Data Driven Governance » ? Pensez-vous que cela va s'accroître ou que c'est plutôt un effet de mode amené à disparaître, et pourquoi ?

Martin : Elle est là pour rester, tant qu'elle ne crée pas une telle pression sur les organisations, qu'elle ne bloque les bonnes pratiques et ne rend pas la vie impossible. Comme il y a une limite à la puissance des machines, elle doit se corriger au fur et à mesure du temps.

Alexis : Merci pour les réponses aux questions générales. Passons maintenant au sujet suivant analysé dans mon mémoire, à savoir l'éducation. Pensez-vous qu'il est nécessaire pour l'état d'investir massivement dans l'équipement des écoles, qui est un préalable « technique » dans l'optique d'une gouvernance dirigée par les données dans le domaine de l'éducation ?

Martin : Oui, nous avons besoin de plus, mais est-ce l'État qui doit tout faire ?

Alexis : Des applications Big Data tels que la prédiction des performances d'un élève, l'ajustement en temps réel des programmes scolaires, le repérage de données communes entre les élèves en échec scolaire, etc. pourraient-elles, selon-vous, améliorer la qualité de l'enseignement et augmenter le niveau des élèves en Belgique ?

Martin : Restons simples, il me paraît stupide de vouloir utiliser ce genre de pratiques beaucoup trop compliquées à mon avis.

Alexis : Quel est votre avis concernant une éducation plus « personnalisée » en matière d'enseignement, par opposition à l'enseignement « standardisé » pratiqué aujourd'hui ?

Martin : L'enseignement national doit être standard ; il appartient à chaque élève d'affiner son programme ; aider chacun ne permet pas de penser par soi-même !

Alexis : Quels sont, selon-vous, les principales qualités et les principaux défauts de l'enseignement à distance ?

Martin : Moins de stress lié au transport et plus d'autonomie, mais pas assez de brassage social et de socialisation.

Alexis : Quel est votre avis concernant l'utilisation de données par l'état afin d'orienter les élèves vers des études avec des débouchés au niveau de l'emploi, en fonction de l'offre et de la demande ?

Martin : Si l'État était capable de planifier correctement la scène de l'emploi, nous n'aurions pas besoin du système d'aide aux licenciements, qui coûte un bras et une jambe, et qui empêche les initiatives et la volonté individuelles de se remettre au travail. Ne laissons donc pas l'État trop s'impliquer...

Alexis : Merci pour vos réponses. Passons maintenant au sujet suivant : la lutte contre la fraude fiscale et sociale en Belgique. Selon-vous, est-ce que la lutte contre la fraude fiscale est un combat prioritaire à mener ou plutôt une fatalité dans un contexte de mondialisation et de globalisation de l'économie ?

Martin : Comme je l'ai dit plus haut, je pense que nous sommes allés trop loin en forçant les administrations, notamment bancaires, à remplir des tonnes de normes, ce qui entrave le bon fonctionnement des affaires !

Alexis : Estimez-vous que l'utilisation du Big Data et des données personnelles pourrait être une arme efficace de lutter contre la fraude fiscale et sociale ? Si oui, pourquoi et comment ?

Martin : Cela se fait déjà sur la scène internationale avec les échanges d'informations établis et cela a permis de réprimer l'évasion fiscale ; si elle a une règle de 80/20 et n'essaie pas d'atteindre 100 %, ce qu'elle ne fera jamais, oui, il est bon de l'utiliser largement.

Alexis : D'un point de vue pratique, légal et éthique, que pensez-vous d'opter pour une politique proactive en matière de fraude, c'est-à-dire de l'empêcher et de la prévenir plutôt que de la sanctionner, comme c'est le cas aujourd'hui ? Des algorithmes prédictifs permettent aujourd'hui d'identifier les fraudeurs potentiels à priori et non plus à posteriori, pensez-vous que l'utilisation par l'Etat de ce type d'instruments est réaliste et réalisable dans notre société ?

Martin : Cela commence par des valeurs communes et non par des systèmes informatiques ; pourtant, même si je suis un optimiste, nous ne parviendrons jamais à tout étouffer... Donc, restons raisonnables dans les applications.

Alexis : Quel est votre avis concernant le système de cartes prépayées « contrôlées » utilisées aux États-Unis, qui interdit aux personnes bénéficiant d'allocations sociales certaines dépenses (alcool, salles de jeux) ? Est-ce que ce système serait transposable en Belgique ?

Martin : La liberté reste primordiale et les États-Unis ne sont pas un standard: il y a une limite à l'étouffement des gens.

Alexis : Merci pour vos réponses concernant ce sujet. Nous allons maintenant passer au sujet du trafic routier. Commençons par une question générale : quelle est votre ressenti sur le trafic routier en Belgique ?

Martin : L'entretien régulier est le véritable problème ; en plus de l'utilisation intensive de la voiture. Ce que l'État doit faire, c'est encourager la civilité individuelle, la protection de l'environnement et la mobilité douce, car tout cela ramènerait une certaine harmonie dans le système.

Alexis : L'utilisation intelligente des données disponibles (géographiques, météorologiques, capteurs de vitesse, etc) pourrait-elle être un moyen d'améliorer significativement la fluidité du trafic routier en Belgique selon-vous ? Ou pensez-vous que la congestion est plutôt d'un problème structurel (infrastructures des routes) et que d'autres solutions seraient préférables (home working le matin, transports publics, etc) ?

Martin : Le homeworking, à cause du Covid-19, était la meilleure solution et elle est là pour rester ; comme nous sommes passés de 5% globalement à 50% en 15 jours, nous allons probablement nous stabiliser à 25%. Si c'est un bon exemple de changement majeur, cela n'empêche pas l'utilisation des données pour fluidifier le trafic... Tout peut aider et doit faire partie de la solution.

Alexis : Que pensez-vous de l'option d'une gestion des routes par le secteur privé (comme c'est le cas en France par exemple) ? Quelles seraient les avantages, notamment concernant l'utilisation des données, que possèderait le secteur privé par rapport au secteur public ?

Martin : Tout ce qui peut être fait par des entreprises privées doit être optimisé ; le problème avec l'intervention de l'État est qu'il n'a pas les normes économiques appropriées.

Alexis : Merci de votre réponse. Passons à présent au dernier sujet : les objets sécurisés par leur conception. Que pensez-vous de ce genre d'objets (d'un point de vue pratique, éthique et légal) ?

Martin : C'est formidable et je le trouve très utile, car je l'utilise moi-même ; si les gens ne le font pas, aidez-les !

Alexis : Quel est, selon-vous, l'avenir des objets connectés sécurisés par leur conception en Belgique ? Pensez-vous que de plus en plus d'objets de ce genre vont voir le jour dans un futur plus ou moins proche ?

Martin : C'est évident mais je ne suis pas sûr que ce soit toujours utile, juste un autre gadget.

Alexis : Que pensez-vous d'une société où énormément d'objets de ce genre feraient leur apparition ? Quel serait encore la part de liberté individuelle laissée au citoyen ? Pensez-vous que par exemple, le citoyen doit parfois pouvoir enfreindre les lois pour se faire entendre (grève de l'impôt, manifestations non autorisées, etc.) ?

Martin : Liberté de parole, liberté d'action... tant qu'elle ne bloque pas le système, donc avec des limites.

Alexis : Merci de vos réponses. Pour terminer, pourriez-vous m'indiquer quelles informations transmises en réponse à ce questionnaire sont confidentielles (et ne peuvent être qu'utilisées anonymement pour mon mémoire) et dans l'affirmative, lesquelles ?

Martin : Tout peut être utilisé.

Alexis : Puis-je vous citer dans mon mémoire ?

Martin : Oui, vous pouvez.

Alexis : Voilà, l'interview touche à sa fin. J'ai obtenu des réponses très intéressantes de votre part. Je vous remercie très vivement pour vos réponses, elles m'aideront sans aucun doute dans mon travail.

Martin : le plaisir est pour moi.

Alexis : Merci beaucoup, bonne soirée.

Martin : Au revoir, et bonne chance surtout !

		L'éducation	La fraude fiscale et sociale	Le trafic routier	Les objets connectés sécurisés par leur conception
<i>1. les compétences</i>		2	3	2	1
<i>2. Le modèle économique</i>		2	3	1	1
<i>3. Le modèle social</i>	<i>Ajustements qualitatifs</i>	3	3	2	1
	<i>Ajustements quantitatifs</i>	1	2	1	1
<i>4. La fiabilité</i>		2	3	2	1
<i>5. La culture décisionnelle</i>		2	2	2	1
<i>6. Les libertés individuelles</i>		2	1	3	1
<i>7. La nécessité</i>		2	3	2	1
<i>8. La légalité</i>		3	3	2	1

3. Interview 3

Répondant : Monsieur Matthieu Aladenise (avocat associé dans le cabinet Buyle Legal et spécialisé en droit commercial et économique et Protection des données à caractère personnel).

Alexis : Bonjour Monsieur Aladenise, je tiens d'abord à vous remercier de prendre le temps pour répondre à quelques questions pour la rédaction de mon mémoire dont le sujet est « Au travers des quatre grandes tendances Big Data identifiées dans le secteur du Big Data électoral (personnalisation, prédictibilité, optimisation, objectivation), nous avons développé des hypothèses d'application d'une possible Data Driven Governance pour les secteurs suivants : l'éducation ; la fraude fiscale et sociale ; le trafic routier ; les objets connectés sécurisés par leur conception », comme je vous l'ai présenté dans mon mail.

Le but de ce questionnaire est d'avoir une opinion extérieure, de la part de gens qualifiés qui présentent un intérêt particulier dans le domaine, sur la possibilité ou non de mener une politique de gouvernance dirigée par les données en Belgique dans les quatre secteurs évoqués.

Matthieu : Avec plaisir. J'ai eu le temps de parcourir dans les grandes lignes ce que vous m'avez envoyé par mail. Je vais donc essayer de répondre au mieux à vos questions.

Alexis : Je vais d'abord commencer par les questions générales et ensuite par des questions plus précises dans les secteurs étudiés. La première question est la suivante : Premièrement, comment définiriez-vous la notion de « Data Driven Governance », soit la gouvernance dirigée par les données ?

Matthieu : C'est assez difficile à définir. Pour ma part c'est actuellement encore un prototype. Je ne pourrais pas vraiment vous donner une définition mais je vois mal une gouvernance « dirigée » uniquement par des données Big Data.

Alexis : Et, quelle est votre opinion concernant l'utilisation de la pratique du Big Data dans le cadre de la gouvernance en général ?

Matthieu : Les données Big Data sont nécessaire dans énormément de cas, indispensable pour beaucoup d'applications, elles permettent une optimisation de solutions, un gain de temps et d'efficacité dans beaucoup de situation. On le voit pour l'instant dans la situation sanitaire et la crise que l'on vit. Mais à côté de cela il y a énormément de dangers lié à un encadrement qui à mon sens n'est pas suffisant actuellement. Si on prend par exemple les formulaires que l'on doit remplir quand on est parti à l'étranger et qu'on revient en Belgique qui vont créer des Big Data sur les origines des déplacements de la population pendant une période et qui vont ensuite être croisés avec d'autres Data de Sciesano ou

d'autres établissements hospitaliers pour ensuite voir si les personnes restent bien confinées. Le Big Data est très intéressant au niveau météorologique, téléphonie, technologique. Au niveau marketing il a certainement une grande puissance également. Après, il peut devenir très invasif. Je pense qu'il est indispensable et qu'il va faire partie de notre vie de plus en plus, qu'on ne pourra plus faire sans mais l'encadrement aujourd'hui n'est pas du tout assez performant.

Alexis : Merci pour votre réponse, vous avez donc répondu en-même temps aux autres questions que je comptais poser à ce sujet, à savoir si pour vous c'est un progrès, si vous estimez qu'il y a des risques et des obstacles, les qualités et le futur de la Data Driven Governance.

Je vais donc passer au sujet suivant qui est l'éducation.

D'abord une question générale. Quel est votre opinion sur le niveau global de l'enseignement en Belgique ?

Matthieu : Je dirais, pas terrible du côté francophone en tout cas. Je trouve que la Fédération Wallonie Bruxelles nivelle l'enseignement par le bas. Les relevés on l'air de dire que les résultats dans les Hautes écoles et les Universités en Flandre sont plus performants.

Alexis : Selon-vous, est-il nécessaire pour l'Etat d'investir massivement dans l'enseignement afin d'équiper les écoles et les universités de plus d'outils numériques, qui sont un préalable à l'utilisation d'une Data Driven Governance dans le domaine de l'éducation ?

Matthieu : Oui je pense que c'est important. Mes filles sont des jumelles de 4 ans qui sont dans une école où il y a de la programmation informatique dès la maternelle. Donc le concept de technologie et d'outils numériques va faire partie prenante de tout un chacun dans l'évolution. Donc autant le maîtriser et apprendre à connaître les limites dès le départ. Il faut que chacun l'appréhende avec ses propres outils. Tout le monde n'est pas logé à la même enseigne pour ce genre de choses.

Alexis : Très bien, merci. La question suivante est : des applications Big Data tels que la prédiction des performances d'un élève, l'ajustement en temps réel des programmes scolaires, le repérage de données communes entre les élèves en échec scolaire, etc. pourraient-elles, selon-vous, améliorer la qualité de l'enseignement et augmenter le niveau des élèves en Belgique ?

Matthieu : Je dirais oui sauf en ce qui concerne la prédiction de performance des élèves.

Le Big Data utilisé à des fins massives de croisement de données pour arriver dans une direction et une orientation principale je dirais oui. Par contre le Big Data utilisé au niveau personnalisation sur le profilage particulier, en ce compris le niveau de performance d'un élève dont ces performances peuvent

dépendre de facteurs qui sont milles par rapport à son intelligence, son attention, des éléments qui ne sont pas encrés dans le Big Data et qui à mon sens ne devraient pas l'être, comme le milieu social, la santé, la dépression et autre font pour moi que la prédiction pour certains élèves n'est pas possible. Après pour les reste, comme par exemple les programmes scolaires, les taux de fréquentations, les taux de réussite par zone, le niveau d'accès aux outils informatiques par rapport au décrochage dans telle ou telle matière, je dirais oui, c'est très intéressant. Mais là on parle de données qui sont de masse

Alexis : Votre réponse rejoint ma question suivante qui est, quel est votre avis concernant une éducation plus « personnalisée » en matière d'enseignement, par opposition à l'enseignement « standardisé » pratiqué aujourd'hui ?

Matthieu : S'il y a un outil et qu'on forme les professeur et les établissements à détecter via croisement de données qu'il y a un décrochage scolaire ou un problème, pourquoi pas. Si il y a un modèle mathématique d'algorithme qui permet de faire cela de façon plus personnalisée, pourquoi également. Mais quelles sont les données qui vont être intégrées pour être le plus juste possible dans cette appréciation ? Pour moi les données qui devraient être intégrées dépassent largement le respect de la vie privée ou seraient incontrôlable, typiquement des données personnelles qui ne sont jamais partagées et qui pourraient entraîner un décrochage scolaire. Là j'y vois déjà un obstacle.

L'orientation de masse est un indice, une incitation qui vient experts ou des décideurs qui permettent d'orienter une stratégie politique ou autre.

Un professeur qui va avoir une solution qui va être basée sur le Big Data, va devoir avoir un professionnalisme et un esprit critique pour remettre en considération les données qui lui sont fournies par rapport à d'autres données qu'il n'a pas pu avoir et je ne pense honnêtement pas que les professeurs soient formés à ce genre de choses. On a pu le constater pendant la crise du Covid, Certains professeurs étaient très à l'aise dans leur fonction et avaient leurs prestations à fournir mais n'allaient pas vraiment plus loin. Beaucoup de professeurs n'ont pas donné de devoirs par exemple, pas beaucoup de suivi alors que d'autres ont fait cela à merveille. On a pu constater pendant cette crise que certains professeurs n'étaient vraiment pas armés pour faire de la personnalisation. Alors s'ils doivent se retrancher derrière un algorithme alors qu'il y a également le facteur humain et d'autres éléments, pour moi non, ça ne peut pas fonctionner comme ça.

Alexis : D'accord, je vais vous poser la question suivante même si vous y avez déjà partiellement répondu. Quel est votre avis concernant la potentielle utilisation des données des élèves (relevés de notes électroniques, journal de travail numérique, statistiques de réussite) par : 1.les universités, afin d'orienter/recruter les élèves dès l'enseignement secondaire et de leur proposer un cursus qui, de par une étude dirigée par des algorithmes, pourrait leur correspondre ou 2.par l'Etat, afin d'orienter les

élèves vers des études qui mènent à des débouchés, en fonction de l'offre et la demande sur le marché de l'emploi ?

Matthieu : Alors moi j'y suis totalement opposé. Oui pour faire des test sur base volontaire. Non pour un accès automatique à ces données. Il y a un milliard de choses qui peuvent faire que pendant votre scolarité en humanité vous êtes moins bon pour une raison ou une autre.. Cela va vous fermer l'accès à certaines écoles ou universités comme dans le système à l'américaine. Car comme je l'ai déjà stipulé, il y a d'autres éléments que simplement les résultats dont ils faut tenir compte et qu'il ne sera probablement jamais possible de mettre dans des données Big Data. Qui va faire en sorte que tout soit conforme et où vont aller ces données ?

C'est trop dangereux, on irait trop loin

Alexis : Merci beaucoup. Je vais maintenant passer au prochain sujet qui est la fraude fiscale. Quel est votre opinion générale sur la fraude fiscale et sociale en Belgique ? Que pensez-vous des efforts faits par le gouvernement belge à ce sujet ?

Matthieu : La fraude fiscale existe et on essaye de la combattre. A mon sens le gouvernement fait des effort qui sont à son niveau, il n'a pas les moyens de faire plus d'efforts.

Alexis : Et selon-vous, est-ce que la lutte contre la fraude fiscale est un combat prioritaire à mener ou plutôt une fatalité dans un contexte de mondialisation et de globalisation de l'économie ?

Matthieu : C'est un combat prioritaire, c'est certain. Les moyen qui y sont mis devraient être prioritaires également mais aujourd'hui ils ne sont pas du tout suffisants. Il faut absolument continuer à combattre la fraude fiscale et développer les moyens, ce serait un mauvais signal que de dire « on n'y arrive pas donc on laisse tomber ». Ce n'est pas parce que l'on n'arrive pas à contrôler tout le monde que l'on doit laisser tomber sinon personne ne respecterait plus rien.

Alexis : Vous ne craignez pas que l'Etat s'en prenne encore aux PME et aux particuliers où ils ont « encore les moyen de contrôler » leur données et plus du tout aux multinationale qui ont « énormément de pouvoir » ?

Matthieu : C'est un sujet d'actualité économique ou autre, mais le groupe Ryan Air qui s'est vu octroyer une aide par l'Irlande et qui a obtenu gain de cause, Google a également obtenu gain de cause en Irlande. Les optimisation fiscales, économiques et mondiales, il y en aura toujours, il y a des sociétés offshore, c'est l'équilibre mondial. Après la fraude fiscale et sociale, il n'y a pas que des grands groupes qui ont des mécanismes, mais il y a aussi les pme, des gens qui ne déclarent pas les gens, qui paient au « black »,

entre autre chose. Il n'y a pas un fétichisme à avoir dans ce genre de combat. Tout le monde est logé à la même enseigne, du meurtrier en série à la personne qui commet un accident involontaire. La fraude fiscale a ses différents degrés et donc reste un combat prioritaire mais est-ce que la Belgique est armée pour combattre seule les grands groupes industriels Français, Allemands ou autres, ou les multinationales ? Non. Est-ce que cela devrait s'organiser au niveau international ? Oui.

C'est un combat prioritaire mais on sait très bien que c'est David contre Goliath quoi.

Alexis : En fait vous avez déjà répondu aux 2 questions suivantes concernant votre avis sur une arme efficace pour lutter contre cette fraude et si c'est faisable pour le gouvernement belge.

Je vais donc passer à la prochaine question qui est : D'un point de vue pratique, légal et éthique, que pensez-vous d'opter pour une politique proactive en matière de fraude, c'est-à-dire de l'empêcher et de la prévenir plutôt que de la sanctionner, comme c'est le cas aujourd'hui ? Des algorithmes prédictifs permettent aujourd'hui d'identifier les fraudeurs potentiels à priori et non plus à posteriori, pensez-vous que l'utilisation par l'Etat de ce type d'instruments est réaliste et réalisable dans notre société ?

Matthieu : Donc c'est de l'anticipation et alors ils ne seraient pas sanctionnés ?

Alexis : Du moins avertis et disons par exemple quadruplement sanctionnés si fraude il devait quand-même y avoir malgré l'avertissement.

Matthieu : C'est toute l'histoire du film de la nouvelle Minority Report où la science-fiction a identifié vers une zone qui est grise. A mon sens c'est tout à fait contraire aux valeurs que nous défendons. La personne qui est grise jusqu'à ce qu'elle ait commis quelques chose...

Alors que le Big Data soit au service pour dire « attention la direction que vous prenez est potentiellement dangereuse » pourquoi pas mais déjà ce sont des dérives que l'on mélange avec le Big Data, c'est pour cela que le Big Data a mauvaise presse, c'est le Big Data et le Big Brother en gros. L'Etat policier, et ici la question est orientée en disant Big Data au service d'un Etat policier, cela veut dire un contrôle en tout temps, qu'ils décident qu'un comportement est répréhensible sans que vous en ayez forcément conscience et décident que sur base des données vous allez dans cette direction-là. Ça vous demande, un, de toujours être conscient de tous comportements répréhensibles ou autres et donc évidemment de la fraude fiscale mais la fraude fiscale c'est large hein. Vous pouvez avoir fait de la fraude fiscale alors que vous étiez de bonne foi mais que vous n'avez juste pas compris une règle. Tout le monde n'a pas un comptable. Donc utiliser ça en tant que préalable à toute infraction sur base des grands principes oui pourquoi pas. Après, aller plus loin, on dépasse la limite Big Brother/ Big Data quoi.

Alexis : D'accord, je vais maintenant passer à la dernière question concernant la fraude fiscale et sociale. Quel est votre avis concernant le système de cartes prépayées « contrôlées » utilisées aux États-Unis,

qui interdit aux personnes bénéficiant d'allocations sociales certaines dépenses (alcool, salles de jeux)? Est-ce que ce système serait transposable en Belgique ?

Matthieu : Alors, le système est bon, enfin bon... c'est un système qui a une aide conditionnée. Vous savez très bien que quand vous avez une assurance avez quelque chose en échange, quand vous avez un logement social, vous avez des conditions à respecter, donc dire que vous recevez une allocation mais que cette allocation doit bénéficier au ménage, aux enfants, elle ne doit pas être dépensée à autre chose, la conditionner moi je ne trouve pas cela choquant.

Est-ce que ce système serait transposable en Belgique, et bien difficilement en l'état car pour être interdit de jeu, c'est soit une interdiction qu'on se crée soi-même ou alors il faut être sous administration provisoire et ne avoir toutes ses facultés donc en l'état notre système législatif ne permet pas d'introduire cette question mais sur le principe d'octroyer une allocation et de conditionner son utilisation je suis d'accord. C'est déjà le cas avec les éco chèques par exemple. Le concept est donc déjà chez nous, pourquoi pas le pousser un peu plus loin et que ce soit un objectif de santé et d'équilibre familial car ces allocations son données pour le bien être des personnes et les allocations familiales pour le bien-être des enfants, devoir les affecter à des dépenses qui sont orientées, moi je ne trouve pas cela choquant.

Alexis : Parfait, merci beaucoup pour vos réponses, je vais maintenant passer à l'avant dernier sujet qui sont des questions à propos du trafic routier. Quel est votre ressenti par rapport au trafic routier en général en Belgique ?

Matthieu : Catastrophique !

Alexis : C'est marrant, tout le monde me répond la même chose. Maintenant la question suivante est l'utilisation intelligente des données disponibles (géographiques, météorologiques, capteurs de vitesse, etc) pourrait-elle être un moyen d'améliorer significativement la fluidité du trafic routier en Belgique selon-vous ? Ou pensez-vous que la congestion est plutôt d'un problème structurel (infrastructures des routes) et que d'autres solutions seraient préférables (home working le matin, transports publics, etc) ?

Matthieu : Alors le problème structurel il est réel. Le RER, ça fait 25 ans qu'on en parle et il n'existe toujours pas. Je parle pour Bruxelles mais aussi pour d'autres villes comme Anvers où c'est un Chao sur le ring également. Il y a un manque de stratégie et d'investissement qui remonte aux 40 dernières années et qui se font ressentir aujourd'hui. Après le Big Data, évidemment, et c'est un des éléments, va à mon sens trouver sa place et aurait une plus-value dans le cadre du trafic routier, tout à fait. Pouvoir orienter en temps réel, c'est ce que font mais à une échelle moindre, les Waze, les Google Map ou autres, en calibrant les itinéraires des personnes, en regardant les zone qui sont plus fréquentées par rapport au trajet annoncé comme TomTom. Je ne fais que constater. Toutes ces données existent déjà, donc c'est

certainement un des enjeux les plus profitable pour le consommateur, en tout cas le particulier. C'est tout l'enjeu du déplacement et du trafic routier. Mais cela ne peut pas être la seule réponse, il faut absolument une structure à côté.

Alexis : En fait cela rejoint un peu la question suivante. Aujourd'hui les données sont partagées essentiellement du secteur public vers le secteur privé mais comme vous le disiez, Waze par exemple pourrait également partager ses données vers le secteur public afin que le secteur public puisse plus facilement étudier le trafic routier en Belgique et faire des travaux en conséquences. Pensez-vous que l'ouverture des données du secteur privé est un impératif pour mener à bien une Data Driven Governance dans le domaine du trafic routier, ou considérez-vous cela comme une utopie ?

Matthieu : Mais c'est une réalité, elle existe. La Région Bruxelles Capitale travaille sur une application de ce type-là. Le projet de la « Smart City » comme on l'appelle, c'est le terme à la mode. La mobilité dans une ville, tout confondu, piétonne, vélo, transports publics, tram, métro, bus, et voitures, dans certaines villes le bateau peut même être rajouté. Par cette application, l'utilisateur peut dire « je veux me rendre à tel endroit » et l'application va croiser toutes les données de tous les transports possibles et imaginables c'est-à-dire que tous les opérateurs privés convergent vers cette application en disant « voilà, là vous avez une station vilo là, vous prenez le tram jusqu'à telle station, là vous avez éventuellement un taxi ou un Uber qui peut vous amener à destination finale etc, donc le croisement et le partage de données privées est en cours.

Alexis : Ah d'accord, je n'avais pas connaissance de ce projet. Est-ce que vous connaissez le nom de cette application ?

Matthieu : Non, je n'ai encore vu aucun nom en dehors de Smart City, je pourrais regarder, mais il y a un cahier des charges qui a été lancé pour l'application. Ils espèrent la sortir dans un horizon prochain, j'imagine que la conjoncture ne va pas aider, mais peut-être 2021-2022. En tout cas le Big Data va avec la Smart City ça c'est certain.

Alexis : super, ce que vous venez de m'expliquer est très intéressant. Vous me confirmez que c'est bien une initiative du secteur public, donc de l'Etat ?

Matthieu : Oui oui, c'est la Région Bruxelles Capitale.

Alexis : Parfait. C'est typiquement le genre d'information que je pourrai intégrer dans mon mémoire. Merci beaucoup. Si vous êtes d'accord, je vais maintenant passer au dernier sujet qui concerne les objets connectés sécurisés par leur conception. Je voulais d'abord savoir si vous savez ce que c'est.

Matthieu : oui oui le design by conception, c'est un des principe du GDPR.

Alexis : En effet, ce sont des objets qui, de par leur nature, empêchent l'utilisateur d'enfreindre la loi. Prenons le cas des voitures : certaines voitures sont aujourd'hui capables de limiter la vitesse lorsqu'elles détectent, grâce à des capteurs, des panneaux de limitation de vitesse. Que pensez-vous de ce genre d'objets (d'un point de vue pratique, éthique et légal) ?

Matthieu : Et bien légal, il n'y a aucun problème de légalité. A un moment j'achète l'objet, j'ai choisi, Je fais le choix de prendre l'objet où il y a une limitation. Pourquoi pas ? Si je décide d'acheter des écouteurs pour mes enfants qui ont une limite de capacité sonore qui abiment les oreilles c'est mon choix. C'est un élément. Les limites de vitesse sont là et le principe est que vous ne pouvez par exemple pas dépasser les 120 km/h sur l'autoroute et bientôt 100 km/h sur le ring de Bruxelles ou les zones 30. C'est la loi. Vous n'avez pas le droit de rouler à 33 km/h. On vous laisse la faculté de le faire parce que, 1 les créateurs automobiles à l'échelle européenne ne peuvent pas créer des voitures spécifiques pour la Belgique, et que 2, pour une raison ou une autre vous devez avoir le droit d'accélérer pour éviter un obstacle ou pour une autre raison. Avoir des objets comme cela ne posent donc aucun problème légal. Pratique, ben s'il y a la possibilité de trouver une application où je suis en voiture et que je ne peux pas aller plus vite que 120 km/h at je n'aurai jamais de PV, après pourquoi pas mais après je serai peut-être frustré et ça va m'ennuyer mais je n'ai pas le droit de rouler à plus de 120 km/h selon la loi, donc.... Je ne vois pas quel problème cela pourrait poser d'un point de vue éthique si l'objet en question se trouve dans un cadre légal qui est lui-même cadenassé. Est-ce que c'est pratique ? non. Est-ce que c'est frustrant ? Certainement.

Alexis : La question suivante est, quel est, selon-vous, l'avenir des objets connectés sécurisés par leur conception en Belgique ? Pensez-vous que de plus en plus d'objets de ce genre vont voir le jour dans un futur plus ou moins proche et connaissez-vous ou avez vu entendu parler d'autres projets pour ce type d'objets dans votre entourage ?

Matthieu : Alors dans un futur plus ou moins proche je ne pense pas. La seule application à laquelle je pense et dont on parle depuis 10 ans, c'est de mettre un appareil dans lequel vous devez souffler dans votre voiture pour démarrer si vous avez été condamné pour conduite en état d'ivresse. On parle de ça depuis 10 ans, il n'y a rien qui a été fait hein.

De plus en plus d'objet je ne pense pas non plus, car on va se heurter à la limite commerciale. Le marketing et l'achat d'un objet est rarement basé sur un besoin, La plupart du temps les achat sont un besoin mais majoritairement un plaisir. Je pense à certains objets qui pourraient être connectés. Il faut qu'il y ai un plaisir dans l'objet ou une véritable potentialité ou autre.

Est-ce que je vais avoir un enjeu marketing énorme à un appareil qui va être limité et bridé parce que j'aurai fait quelque chose ? Qu'on installe automatiquement, parce que la loi le dit, à tout personne qui a été condamnée à autant de conduites en état d'ivresse, ou alors elle ne peut plus conduire ou alors on installe dans son véhicule ce genre d'objet OK, mais ça c'est une sanction. Maintenant est-ce que j'ai envie d'installer ça dans la voiture de mes enfants car je n'ai pas envie qu'ils roulent en état d'ivresse alors je le fais. Mais cela reste pour cadenasser quelque chose et même ouvrir des potentialités. Est-ce que ça va être l'avenir ? Au niveau économique je ne crois pas car il n'y a aucune société qui va réussir à vendre des milliers et des milliers d'articles basés sur un principe qui nous bride. Dans ce concept là il y a ce besoin de liberté ou de plaisir lié à un objet connecté qui va être dès le départ cadenassé.

Alexis : L'interview touche à sa fin, merci beaucoup pour vos réponses et pour le temps que vous y avez consacré. J'ai encore 2 petites questions plus formelles à vous poser. Y-a-t-il des informations transmises en réponse à ce questionnaire qui sont confidentielles et ne peuvent être qu'utilisées qu'anonymement pour mon mémoire ? Et puis je vous citer dans le cadre de mon mémoire ?

Matthieu : Non aucune information n'est confidentielle et oui vous pouvez me citer dans le cadre de votre mémoire. J'aimerais juste que mon nom ne soit pas cité à propos du reproche sur certains enseignants, c'est tout. Pour la critique générale sur l'enseignement pas de problème.

Alexis : Voilà, l'interview est terminée, au revoir Monsieur Aladenise, encore merci beaucoup et je vous souhaite de bonnes vacances

Matthieu : Au revoir Alexis, je vous souhaite une bonne continuation.

4. Interview 4

Répondant : Monsieur Nicolas Dombret, Head of supply & demand in Data Office chez Ethias

Alexis : Bonjour Monsieur Dombret

Nicolas : Bonjour Alexis, je propose que l'on se tutoie, c'est beaucoup plus sympa

Alexis : D'accord, avec plaisir. Je te tenais tout d'abord à te remercier de prendre le temps de répondre à mes questions pour la rédaction de mon mémoire dont le sujet est « Au travers des quatre grandes tendances Big Data identifiées dans le secteur du Big Data électoral (personnalisation, prédictibilité, optimisation, objectivation), nous avons développé des hypothèses d'application d'une possible Data Driven Governance pour les secteurs suivants : l'éducation ; la fraude fiscale et sociale ; le trafic routier ; les objets connectés sécurisés par leur conception ».

Comme je l'ai expliqué dans mon mail, le but de ce questionnaire est d'avoir une opinion extérieure, de la part de gens qualifiés qui présentent un intérêt particulier dans le domaine, sur la possibilité ou non de mener une politique de gouvernance dirigée par les données en Belgique dans les quatre secteurs évoqués ci-dessus. Bien entendu, s'il y a des questions auxquelles tu ne sais ou ne veux pas répondre, cela ne pose aucun problème.

Nicolas : Oui effectivement, j'ai eu le temps de parcourir les questionnaire que tu m'as envoyé et j'ai moi-même quelques question à te poser à ce sujet car il y a des secteurs du styles, fraude sociale et fiscale dans lequel j'aurais plus de mal à t'aider.

Alexis : Il n'y a vraiment aucun problème si tu ne sais pas répondre à toutes les questions. En fait j'ai fait un questionnaire très général dans le but d'interviewer des gens de tous les horizons, par exemple un enseignant, un avocat, afin d'avoir différents points de vue de différentes personnes de différents secteurs afin de pouvoir faire un audit global. Car dans les secteur sur lesquels je me base il est très difficile de trouver des spécialistes dans le secteur même.

Nicolas : Ok, parce qu'en fait, j'avais peut-être une première question. Quand je prends la première question ne fusse que dans le général « comment définiriez-vous la notion de « Data Driven Governance », soit la gouvernance dirigée par les données ». En fait ici, est-ce que toi, tu as déjà une définition, est-ce que tu cherches un peu à croiser quelles sont les autres définitions des gens qui travaillent un peu la dedans ou pas ?

Alexis : Oui, c'est bien ça mon objectif. J'ai aussi pu la poser autrement et demander, qu'est-ce qui vous vient à l'esprit quand on évoque une Data Driven Governance, il ne faut pas me donner une définition académique.

Nicolas : D'accord, mais c'est comme par exemple je reprends ici la 2^{ème} question, tu vois quand je lis quelle est votre opinion concernant l'utilisation de la pratique du Big Data dans le cadre de la gouvernance en général, la première chose qui me vient à l'esprit c'est que le Big Data, enfin je trouve, c'est vraiment un postword quoi. Tout le monde en parle, et parfois c'est un peu comme le sexe et les adolescents, tout le monde en parle et finalement y'a personne qui en fait. Et je trouve que la définition, c'est comme Data Science, ces définitions ne sont pas toujours claires et les gens parlent de Big Data mais finalement, eux, qu'est-ce qu'ils mettent derrière Big Data ? J'ai l'impression que chacun a un peu sa propre définition.

Alexis : Je suis d'accord avec toi, j'ai d'ailleurs mis cela dans la partie théorique de mon mémoire, que je ne t'ai pas envoyée dans mon mail. Qu'il n'y a pas de définition unanime du Big Data. Moi je l'utilise en prenant les 3 critères de base qui sont le volume, la variété et la vitesse, Ce sont les 3 V. maintenant il y a des gens qui définissent en fonction de ce que l'on peut faire avec, de ce que c'est, de la valeur que ça peut créer, enfin, il y a plein de définitions possibles.

Donc par exemple pour la première question, qu'est-ce qui te vient à l'esprit quand on parle de Data-driven Governance .

Nicolas : Avec un peu de contexte, cela paraît déjà beaucoup plus simple d'y répondre. Je ne sais pas si je vais avoir les mots juste tout de suite. Pour moi, et je parle là de mon expérience, c'est vraiment pouvoir arriver à, dans un cas comme le nôtre ici chez Ethias, diriger la compagnie et prendre des décisions sur base de faits. Et ces faits ils viennent de données. Par exemple, est-ce qu'il y a des secteurs qui sont plus rentables que d'autres, est-ce que des clients sont plus rentables que d'autres ? Quand un commercial doit aller voir des clients, l'idée grâce aux données est de savoir quel client il faut d'abord assainir ou quels clients sont peut-être sur le point de partir et qu'il faudrait aller voir pour essayer de les retenir ? Ce genre de chose là quoi. Pour moi le Data-driven Governance c'est arriver à diriger une compagnie ou autre tant sur le plan stratégique qu'opérationnel, le plus possible sur base de faits qui proviennent des données. Que ce soit en terme de vision sur certains secteurs, en terme de cibles clients, certains produits, parce que, évidemment les faits, que ce soit sur les tendances passées ou les tendances futures, c'est ça qui va les aider à pouvoir prendre des décisions et prendre la bonne direction pour le futur. Et pouvoir également regarder ce qui se passe à l'extérieur sur le marché, pour être certain que l'écosystème dans lequel la société évolue est toujours adéquat avec ce qu'elle fait. C'est pareil sur le plan opérationnel. Je prenais l'exemple des sales qui peuvent s'aider d'un CRM de données pour mieux cibler les clients, plutôt que de faire cela au feeling ou au feeling. Parce qu'on constate dans le passé

évidemment que quand était moins Data-driven, moins Data based, les gens prenaient fatalement des décisions qui étaient beaucoup plus basées sur l'instinct, le feeling et l'expérience. J'avais un exemple assez criant de vérité, c'était un manager, c'était vraiment sur des croyances, il disait qu'il y avait probablement beaucoup plus de sinistres bris de vitres voitures à Charleroi qu'ailleurs. Et finalement quand on regarde les chiffres, ce n'est pas du tout le cas. C'est un bête exemple mais cela prouve que le Data-driven et d'être fact-based est important. C'est comme cela que moi je définis le Data-driven Governance et à quoi ça sert.

Alexis : Parfait, donc, cela veut dire se détacher de l'intuition humaine

Nicolas : oui c'est ça, mais parfois l'intuition reste toujours bien aussi, parce que parfois l'intuition va servir aussi parfois à challenger certains chiffres. Quand on veut faire des modèles prédictifs par exemple, on va prendre des variables explicatives et là, ne fusse qu'en amont de la constitution du modèle, c'est bien d'avoir l'intuition du Business pour pouvoir dire « oui moi je pense que pour établir si un client s'en va ou pas, on a constaté que dans les 3 derniers mois ils ne payait plus ses factures ou on avait beaucoup moins de contacts etc... ».

Cette intuition, dans certains cas, va aussi servir à nous diriger pour créer un modèle prédictif plutôt que de tracer un millions de variables brutes.

Alexis : Donc, c'est en fait émettre des hypothèses et essayer de les vérifier grâce aux données plutôt ?

Nicolas : Voilà, c'est ça. Ou alors il arrive parfois également qu'on ait des données, des chiffres et qu'une intuition nous démontrent que ces données ne sont pas correctes. Par exemple les données nous montrent qu'on a fait 100 assurances auto et un commercial nous dit que ce n'est pas possible, qu'on a dû en brasser au moins 500.

Un mix d'intuition et de données est donc nécessaire.

Alexis : J'en parle dans mon mémoire, et ton opinion rejoint ce que je dis à ce sujet. C'est parfait. Je vais passer à 2^{ème} question qui est quelle est ton opinion concernant l'utilisation de la pratique du Big Data dans le cadre de la gouvernance en général ?

Nicolas : J'ai un point de vue très mitigé sur la question. L'Etat pourrait l'utiliser mais cela dépend pourquoi. Pour pouvoir attaquer la fraude par exemple je pense que cela pourrait être une bonne chose. Maintenant je ne reste pas fan d'un Etat qui aurait les yeux sur tout et qui puisse avoir cette espèce de, comment pourrait-on appeler ça...un peu Big Brother où finalement ils pourraient voir tout ce qu'on fait et que finalement on n'ait plus aucun degré de liberté, de ça je ne suis vraiment pas fan. Je dirais donc oui mais en gardant bien entendu un gros point d'attention sur l'éthique par rapport à ce qu'on fait.

Je pense par exemple à Bruxelles, depuis quelques années, quand on va se garer en rue on doit introduire son numéro de plaque quand on veut payer le parking. J'ai toujours du mal à me positionner par rapport à ça. Dans les faits ça ne me dérange pas et ça pourrait peut-être un jour même m'aider si un jour je ne paye pas une facture je pourrais très bien leur dire, écoutez, prenez ma plaque, et vous verrez bien que sur les 150 fois où je me suis garé à Bruxelles j'ai payé 149 fois. Mais d'un autre côté, parfois, je me dis mais qu'est-ce qu'il peuvent faire si je rentre ma plaque pour payer mon parking ? Est-ce qu'à un moment donné ils vont pouvoir dire, tiens on remarque que vous avez toujours une amende impayée pour la ville, vous ne pouvez plus vous garer ici par exemple. Donc de nouveau je pense qu'il y a du bon sens, il y a sûrement de bonnes choses là dedans mais il faut faire attention à où on va trouver les limites par rapport à ça.

Alexis : Ca rejoint un peu la question suivante : quels sont les secteurs/domaines publics où vous accentueriez l'utilisation des données afin d'améliorer la politique menée par le gouvernement et pourquoi?

Nicolas : Oui, tu vois, l'idée c'est, et là je rejoins un peu ce que dit aussi dans les grandes lignes le GDPR, c'est de ne pas commencer à utiliser des données à certains buts, je reprends l'exemple d'introduire son numéro de plaque pour payer son parking, on voit que cette plaque là c'est Nicolas Dombret, on voit qu'il ne paye son parking qu'une fois sur 2 donc on va lui octroyer moins de choses à tel autre niveau. C'est vraiment là que je mets un gros point d'attention. Des choses comme de la fraude, ça me paraît faire sens. Mais je reprends toujours ce même exemple de plaque de voiture, pour moi ce n'est pas clair pourquoi on doit introduire son numéro de plaque. C'est ça aussi le problème, il y a des moments où on donne des données et où on ne sait pas ce que l'Etat en fait. Peut-être que l'Etat devrait être juste beaucoup plus clair et dire voilà, on va faire cela avec vos données, et on devrait pouvoir dire si on est d'accord ou pas d'accord. Ils doivent vraiment être plus clair avec ce que eux font des données pour que ce soit transparent pour le citoyen. Est-ce que je dois donner ma plaque pour me faire fliquer, est-ce que c'est purement statistique ? Est-ce pour faire des études sur les parking à Bruxelles ? Tout ça est beaucoup trop flou, aucun citoyen ne sait pourquoi il doit le faire.

Alexis : C'est vrai, je parle effectivement également de ça dans mon mémoire. Il faudrait ouvrir les modèles aux citoyens pour expliquer qu'est-ce qu'on va faire exactement des données.

Est-ce que les données qu'on donne restent vraiment dans le public ? Si on passe à un vrai open Data il y a un risque que des sociétés privées croisent des données par exemples avec les réseaux sociaux au final des données qui sont au départ anonymisées, à force d'être croisées, peuvent devenir des données personnelles alors que c'est à peu près interdit.

Nicolas : Complètement, pour reprendre cet exemple de plaque, l'Etat pourrait très bien dire, moi je ne demande que les plaques mais pas les noms derrière mais si un coco arrive à prendre la plaque et croiser cela avec des données d'assurances pour lesquelles ils ont des cartes vertes et donc la personne lié au numéro de plaque. C'est clair que cela devient des données tout à fait privées.

Alexis : Tu as déjà répondu à quasi toutes les questions que je voulais te poser sur le Data Driven Governance en général. Je vais donc te poser la dernière question qui est, quel futur imagines-tu pour la « Data-driven Governance » ? Penses-tu que cette tendance va plutôt s'accroître à l'avenir, ou, au contraire, que c'est un effet de mode qui va disparaître, et pourquoi ?

Nicolas : J'ai beaucoup de mal à répondre à cette question. Je pense qu'il y a un effet de mode mais ce n'est pas pour autant que cela va disparaître. Tant qu'on reste plus orienté faits, ça va de toute façon continuer d'aider un paquet de monde. Donc je ne pense pas que ça va disparaître. Je ne peux pas affirmer non plus que ça va s'accroître. Ça va peut-être même ralentir ou ça va peut-être trouver une vitesse de croisière mais je ne pense pas que ça va diminuer. Mais ça reste un feeling.

Alexis : Merci beaucoup pour tes réponses. Je vais maintenant passer aux questions concernant l'éducation. La première question est quelle est ton opinion sur le niveau global de l'enseignement en Belgique ?

Nicolas : Je vais parler de ma propre expérience. Je trouve que le niveau global est faible et quand j'ai commencé ma vie d'adulte, dans ma carrière j'ai eu la chance de pouvoir suivre des cours sur tout ce qui était leadership. J'ai été vraiment étonné qu'il y avait un paquet de choses que je découvrais en étant déjà adulte et que j'aurais pu découvrir quand j'avais entre 15 et 20 ans. Je me souviens que quand on a suivi certains cours sur le leadership, sur certains modèles de pleine conscience, de conscientisation etc, on se faisait tous la réflexion et on se disait mais mon Dieu pourquoi n'a-t-on jamais vu ça à l'école ? Pourquoi n'a-t-on pas vu cela plus tôt ? Où on peut prendre un peu conscience de plus de choses tant par rapport à soi que par rapport à l'autre et je trouve effectivement que le système aujourd'hui reste trop classique, trop vieux, par rapport à tout ce qui se fait ailleurs. Je vais prendre un exemple très factuel, c'est en cours de physique, on a encore très fort insisté sur la mécanique classique et j'ai vraiment du attendre l'université pour rentrer dans tout ce que Einstein avait fait. Alors je peux tout à fait comprendre et entendre que ce qu'Einstein a fait sur la relativité restreinte et générale est compliquée mais j'ai du mal à concevoir qu'on enseigne pas quand-même un certain nombre de choses à l'école, en humanité, pour peut-être aussi inspirer plus les jeunes au niveau de la science en dehors de ce que Newton a fait et comprendre que cette complexité fait aussi sa magie. Donc, voilà, il y a quand-même un paquet de fois où je me suis dit « mais pourquoi on n'a pas vu ça plus tôt » ? Pour répondre à la question, je trouve que le niveau du système est relativement bas. Il y a beaucoup de chose sur lesquelles

il faudrait agir, peut-être aussi au niveau des Data mais peut-être qu'il y a des choses qui sont faites à ce niveau-là, ça je ne sais pas, moi je suis sorti de l'école il y a 20 ans. Mais aussi au niveau de la préparation à notre rôle de citoyen, comme par exemple comment voter correctement, comment pouvoir gérer les taxes, éventuellement comment devenir un jour indépendant, Personnellement c'est ma vie qui me l'a appris mais pas du tout l'école.

Alexis : D'accord, merci beaucoup. Je passe à la question suivante qui est : selon-toi, est-il nécessaire pour l'Etat d'investir massivement dans l'enseignement afin d'équiper les écoles et les universités de plus d'outils numériques, qui sont un préalable à l'utilisation d'une Data-driven Governance dans le domaine de l'éducation ?

Nicolas : Clairement oui, c'est vraiment une priorité.

Alexis : Ensuite, est-ce que des applications Big Data tels que la prédiction des performances d'un élève, l'ajustement en temps réel des programmes scolaires, le repérage de données communes entre les élèves en échec scolaire, etc. pourraient-elles, selon toi, améliorer la qualité de l'enseignement et augmenter le niveau des élèves en Belgique ?

Nicolas : Je ne me suis jamais posé cette question mais j'aurais tendance à dire que oui.

Alexis : Et quel est ton avis concernant une éducation plus « personnalisée » en matière d'enseignement, par opposition à l'enseignement « standardisé » pratiqué aujourd'hui ?

Nicolas : J'aime beaucoup l'idée et je me rappelle de cette phrase mais je ne sais plus qui l'a dite « on ne peut pas apprendre à un poisson à grimper à un arbre ». Je pense que chacun à son potentiel et que effectivement il faut savoir l'exploiter correctement.

Alexis : Et à partir de quels niveaux de l'enseignement (maternel, primaire, secondaire, universitaire) pensez tu qu'il serait le plus utile de l'appliquer et pourquoi ?

Nicolas : Honnêtement j'aurais du mal à me prononcer en disant, maternelles, primaires, secondaires. Maintenant autant je suis pour un enseignement plus personnalisé, autant je pense que par exemple on ne doit pas supprimer les math pour quelqu'un qui n'est pas bon en math, on ne doit pas non plus empêcher quelqu'un de faire de l'effort pour arriver à quelque chose. Parce que cela reste important de quand-même avoir un certain niveau, il faut peut-être trouver des niveaux minimum requis. A quel niveau commencer ? Maternelles, je trouve que c'est un peu trop tôt, primaires on voit quand-même des

chose ultra classiques pour tout le monde donc ça me paraît fort tôt. Je dirais plutôt entre les secondaires et universitaires.

Il y a également une chose que je voudrais rajouter, c'est que ce qui pourrait être amélioré dans le système éducatif c'est faire plus attention aux personnes qui ont plus de mal à écrire, à noter. Il y a des gens qui sont plus « visuels » ou apprennent mieux juste en écoutant. C'est quelque chose qui n'est pas pris en compte aujourd'hui, tout le monde doit absolument écrire. On devrait trouver des alternatives par rapport à ça également. Personnellement, moi si j'avais pu, j'aurais enregistré mes prof plutôt que devoir écrire. Ça m'aurait vraiment facilité la tâche car je suis quelqu'un qui apprend mieux en écoutant.

Alexis : OK, merci. La question suivante est Dans le cadre de la crise du COVID-19, as-tu observé/utilisé des applications orientées « Big Data » dans l'enseignement, que ce soit en tant que parent, professeur ou personne neutre ? Si Oui, lesquelles ?

Nicolas : Non, je ne suis ni parents ni enseignant et dans mes proches je n'ai pas d'exemples non plus.

Alexis : D'accord, pas de problème. La dernière question qui est en 2 parties sur l'enseignement est quel est ton avis concernant la potentielle utilisation des données des élèves (relevés de notes électroniques, journal de travail numérique, statistiques de réussite) par les universités, afin d'orienter/recruter les élèves dès l'enseignement secondaire et de leur proposer un cursus qui, de par une étude dirigée par des algorithmes, pourrait leur correspondre ?

Nicolas : Par les Universités, pour l'équité je trouve cela une excellente idée. J'imagine que ce ne serait pas une science exacte mais personnellement en sortant des études je me suis posé beaucoup de questions. Parce que je ne savais pas quoi faire et que le monde du travail c'était vraiment un grand flou pour moi. Je pense que ça m'aurait vraiment aidé d'avoir un espèce d'algorithme qui dit tiens suivant tout ce que tu as fait et là où tu sembles avoir des facilités, tu pourrais peut-être te diriger vers ça ou ça. Maintenant quelles données de mon cursus d'humanitaire on aurait pu utiliser pour dire tiens tu peux faire ceci ou cela. On aurait peut-être dit » oui tu es meilleur en math, dirige toi vers ingénieur », thanks. Il faudrait voir ce que l'on pourrait apporter de plus pointu que ça mais j'aime assez l'idée.

Alexis : Et par l'Etat, afin d'orienter les élèves vers des études qui mènent à des débouchés, en fonction de l'offre et la demande sur le marché de l'emploi ?

Nicolas : Alors, par l'Etat pour orienter vers des emplois, c'est important pour mettre les étudiants au courant car il y a des emplois qui sont tout à fait méconnus. Quand tu sors des humanités ou même de l'université, ce qui se passe dans le monde du travail est encore trop méconnu et des choses dont on a besoin dans la sphère économique.

L'Etat devrait être plus transparent tant sur les débouchés de certaines études, comme psycho par exemple. Énormément d'élèves qui ne savent pas très bien quoi faire comme études se dirigent vers des études de psycho mais il n'y a malheureusement pas beaucoup de débouchés la dedans, l'Etat devrait conscientiser les élèves à ce sujet. Mais à l'inverse l'Etat devrait également communiquer sur le manquement de certains emplois.

J'ai connu autour de moi plein de gens qui faisaient des études sans savoir vraiment ce qu'ils voulaient faire comme métier parce que de nouveau la sphère économique est très floue, on ne sait pas ce qu'on peut faire après, on veut faire un truc parce que ça nous plaît un peu. Effectivement, je pense que l'Etat peut jouer un gros rôle la dedans pour faire mieux le lien entre ce monde universitaire et le monde économique.

Alexis : Super, merci beaucoup pour tes réponses, on va maintenant passer à la première question du suivant qui est la fraude fiscale et sociale. Quel est votre opinion générale sur la fraude fiscale et sociale en Belgique ? Que pensez-vous des efforts faits par le gouvernement belge à ce sujet ?

Nicolas : J'ai l'impression, et c'est vraiment une impression, qu'il y a moins de fraude car on n'est moins capable de frauder aujourd'hui qu'il y a 30 ans. Mais il reste sûrement de la fraude et peut-être même de la très grosse fraude. J'ai l'impression, pour moi-même avoir été indépendant, que l'Etat avait quand-même quelques indicateurs, quelques alertes ou quelques croisement de données pour pouvoir s'assurer qu'il n'y avait pas des « foufrelages » à gauche et à droite. Maintenant ce que moi je regrettais quand j'étais indépendant, c'est que je pense qu'il y a un paquet de fois où tu peux frauder sans le faire exprès. On peut appeler ça un fraudeur passif, je ne sais pas. Parfois tu fais des choses, tu as l'impression que tu es dans le bon et en fait tu es tout à fait dans le mauvais. Je prends l'exemple des frais, il y a des choses qui n'ont jamais été claires pour moi comme par exemple est-ce qu'on peut mettre dans les frais des vêtements genre un costume quand on est consultant. Ce n'était jamais clair. A quel moment l'Etat considère qu'on abuse trop dans les frais de resto par exemple. Je pense que là aussi j'aurais aimé avoir quelque chose de plus transparent. Par exemple avoir un simulateur au niveau de l'Etat qui nous permettrait de dire, voilà, j'en suis là dans mes frais. Peut-être quelque chose qui facilite aussi la vue de mon bénéfice net à la fin de l'année en fonction de mes frais. Là actuellement, c'est compliqué, c'est à chaque fois tous des calculs, il faut faire appel à un comptable. L'Etat pourrait aussi arriver avec les données, à faire des choses plus dynamiques où quand à chaque fois que je peux rentrer mes frais peut-être plus instantanément vers l'Etat et qu'il puisse me dire directement via des tax-on-web, tu gagnes autant, tu paies autant à l'ONSS, tu as rentré autant de frais donc tu donnes au paies autant d'impôts. Je trouve qu'il y aurait moyen de faire quelque chose de plus dynamique et pas toujours attendre une année ou 3 mois pour faire hyper gaffe à ce que tu fais. Je pense qu'il y a encore énormément de choses que l'Etat peut faire par rapport à ça.

Alexis : Merci, je vais passer à la question suivante, Selon toi , est-ce que la lutte contre la fraude fiscale est un combat prioritaire à mener ou plutôt une fatalité dans un contexte de mondialisation et de globalisation de l'économie ?

Nicolas : Je pense que cela doit rester un combat prioritaire et cela doit s'organiser au niveau Européen ou d'un ensemble d'Etats et pas uniquement au niveau d'un pays.

Alexis : Alors, estimes-tu que l'utilisation du Big Data et des données personnelle pourrait être une arme efficace de lutter contre la fraude fiscale et sociale ?

Nicolas : Oui, en tout cas, j'en ai déjà parlé dans une de mes réponses précédentes.

Alexis : Est-ce que tu penses qu'il y a des techniques de Data-driven Governance qui sont utilisées par le gouvernement belge aujourd'hui afin de lutter contre la fraude fiscale et sociale ?

Nicolas : Certainement, moi pour l'avoir vécu, d'ailleurs ils s'étaient plantés joyeusement, je rentrais ce que je gagnais sur Taxon Web, je rentrais donc ce que j'avais facturé. A l'époque, j'étais consultant pour Ethias, j'étais indépendant. Et ce que l'Etat avait apparemment fait comme check, c'est qu'ils checkaient ce que j'avais rentré comme montant et ils recheckaient derrière avec ce que Ethias avait renvoyé à l'Etat comme montant qui m'avait été facturé. Donc ils sont arrivés chez moi en disant « il y a des différences », pourquoi, parce que ils ont vérifié mon montant qui était hors TVA avec celui d'Ethias qui était TVA incluse. C'est un truc complètement aberrant mais ils faisaient quand-même en tout cas des croisement de données à gauche et à droite. Mon comptable m'a dit qu'ils faisaient apparemment aussi des check de variations au cours des années de ta quantité de frais. D'après mon comptables, ils ont des règles qui donnent des alertes s'il y a de grosse variations de frais sur plusieurs années. Et si c'est le cas on se choppe un petit contrôle fiscal. Ce sont des check relativement basic mais ils existent.

Alexis : OK, merci, c'est un très bon exemple, je pourrai intégrer cela dans ma thèse. La question suivante est d'un point de vue pratique, légal et éthique, que penses-tu d'opter pour une politique proactive en matière de fraude, c'est-à-dire de l'empêcher et de la prévenir plutôt que de la sanctionner, comme c'est le cas aujourd'hui ? Des algorithmes prédictifs permettent aujourd'hui d'identifier les fraudeurs potentiels à priori et non plus à posteriori, penses-tu que l'utilisation par l'Etat de ce type d'instruments est réaliste et réalisable dans notre société ?

Nicolas : J'adore l'idée, je pense beaucoup aux fraudeurs soit qui fraudent inconsciemment soit qui fraudent consciemment mais qui ne se rendent pas compte fatalement des impacts. Ce n'est jamais gai

pour une personne de se prendre un contrôle ou de se prendre une sanction voire une faillite . Ce n'est ni bon pour le pays, ni bon pour les clients. Je pense donc que ce serait vachement bien d'en tout cas pouvoir prévenir. Ceux qui le font inconsciemment et qui ont un peu de conscience professionnelle vont réagir en se rendant compte qu'ils ne sont pas dans le bon. cela pourrait et leur éviter pas mal d'ennuis. Je prends l'exemple de mes parents, il y a 20 ans, qui faisaient des choses avec leurs frais qu'ils ne devaient pas du tout faire. Ils ont eu une grosse sanction après quelques années et franchement si mes parents avaient été prévenus à l'avance que ce qu'ils faisaient n'était pas du tout juste, ça leur aurait évité bien des ennuis. Donc oui, j'aime beaucoup l'idée, honnêtement.

Alexis : Voilà, on arrive maintenant à la dernière question au sujet de la fraude. Quel est ton avis concernant le système de cartes prépayées « contrôlées » utilisées aux États-Unis, qui interdit aux personnes bénéficiant d'allocations sociales certaines dépenses (alcool, salles de jeux)? Est-ce que ce système serait transposable en Belgique ?

Nicolas : Là je trouve par contre, qu'on commence à franchir une limite avec laquelle j'ai beaucoup de mal. Ici on commence avec les alcool et les salle de jeux, on pourrait dire que ça fait sens. Mais d'un autre côté, quelle serait l'étape suivante ? Plus on laisse du pouvoir, pour la bonne cause, à l' Etat, plus à un moment donné, s'il n'y a pas une éthique où on dit stop, on arrive dans des drôles de situations. Autant je pourrais dire à priori que cela fait sens, autant je n'aime quand-même pas trop l'idée.

Alexis : Merci pour tes réponses, je vais maintenant passer aux questions sur le trafic routier. Tout d'abord quel est ton ressenti par rapport au trafic routier en général en Belgique ?

Nicolas : Complicé. Je roule à Bruxelles et à Liège, c'est compliqué. Bon maintenant évidemment depuis l'arrivée du Covid, le trafic est relativement light. Le Homeworking et le Covid y jouent pour beaucoup, sinon avant ça c'était une horreur. Surtout dans Bruxelles.

Alexis : Tout le monde répond en effet la même chose et tu réponds déjà en partie à la question suivante qui est : L'utilisation intelligente des données disponibles (géographiques, météorologiques, capteurs de vitesse, etc) pourrait-elle être un moyen d'améliorer significativement la fluidité du trafic routier en Belgique selon toi ? Ou penses-tu que la congestion est plutôt d'un problème structurel (infrastructures des routes) et que d'autres solutions seraient préférables (home working le matin, transports publics, etc.) ?

Nicolas : On pourrait utiliser les 2. On pourrait utiliser les data justement pour mettre en place des solutions qui pour moi doivent être plus structurelles, comme le home working le matin, pour mettre en place dans des ville ou des communes x personnes commencent entre 9h et 10h, d'autres commencent

entre 10h et 11h et on change une semaine sur 2 ou ce genre de choses, pour arriver un peu à faire ça par paquet. Hors je sais que ce n'est pas toujours évident car il y a des gens qui ont des enfants etc. Car c'est ça aussi le problème, on aura beau mettre du home working mais il y aura toujours les écoles, et quand tu dois déposer tes enfants à l'école tu n'as pas le choix de l'heure. Mais sinon je crois qu'avec les données et un peu de changement structurel il y aurait moyen de faire beaucoup. En ce qui concerne le structurel, dans des endroits comme Bruxelles, mettre clairement bien des pistes cyclables et gérer vraiment bien les transports en commun, vont aussi permettre à des enfants de se rendre de façon plus autonome à l'école, en vélo par exemple, à un ou deux km, là où avant les parents devaient les conduire parce que le vélo n'était pas une solution possible. Je pense qu'il y a beaucoup de solution qu'on pourrait mettre en place.

Alexis : Super, merci, la question suivante est afin d'optimiser la Data Driven Governance, nous évoquons le partage des données types « Open-Data » du secteur public vers le secteur privé mais également du secteur privé vers le secteur public, ce qui est peu le cas aujourd'hui. Penses-tu que l'ouverture des données du secteur privé est un impératif pour mener à bien une Data-driven Governance dans le domaine du trafic routier, ou considères-tu cela comme une utopie ?

Nicolas : Je pense que c'est une bonne idée, mais je ne pense pas que le secteur privé ferait cela gratuitement. Il faudrait qu'il y ait une contrepartie comme par exemple des avantages fiscaux.

Alexis : Ok, merci, je vais maintenant passer au dernier sujet qui concerne les objets connectés sécurisés par leur conception. Est-ce que tu as déjà entendu parler de ce genre d'objet ?

Nicolas : non, ça ne me dit rien du tout en fait.

Alexis : C'est par exemple, des oreillettes qui ne pourraient pas dépasser un certain volume parce que ce n'est pas bon pour les oreilles. Ou par exemple des voitures qui quand elles repèrent un panneau de limitation de 120 km/h, ne pourrait plus dépasser cette vitesse. Que penses-tu de ce genre d'objet d'un point de vue pratique, éthique et légal ?

Nicolas : Là , de nouveau, je trouve que l'on touche trop au libre arbitre. Moi ça ne me va pas. Je prends l'exemple de la voiture avec le panneau de limitation à 120 km/h, je prends l'autoroute tous les jours et il y a toujours des moments où tu dois un peu pousser sur l'accélérateur jusque 135 km/h qui est pile à la même vitesse que toi sur la bande du milieu parce que quelque chose à fait que tu t'es retrouvé sur la bande de gauche et qu'une voiture s'est mise sur la bande du milieu à la même vitesse que toi et qu'à un moment donné tu dois te dégager de cette bande de gauche et donc tu piques un coup d'accélération pour arriver à te remettre à un moment donné quelque part sur la bande de droite. Si tu es bloqué à 120,

tu ne sais pas le faire. Tu n'as plus qu'à freiner et si il y a une voiture derrière toi c'est dangereux, c'est l'horreur. Pour les oreilles, c'est pareil, si à un moment donné, moi j'ai envie, et ça ne regarde que moi, de m'éclater les tympans avec un morceau que j'adore, c'est mon problème.

Ce n'est pas le problème des gens qui fabriquent la voiture ou les oreillettes par exemple. Alors c'est toujours bien d'avoir un dispositif qui pourrait mentionner la chose, par exemple un bip pour signaler qu'on a atteint les 120 km/h, mais pas qui empêchent de faire la chose car je trouve qu'il faut toujours avoir un certain libre arbitre. Sinon on n'est plus des êtres humains, on n'a plus qu'à se transformer en robots et on fait tout suivant les règles. Ça ça me pose plus un problème.

Alexis : D'accord, je passe donc à la dernière question à laquelle tu as quand-même en partie déjà répondu. Que penses-tu d'une société où énormément d'objets de ce genre feraient leur apparition ? Quel serait encore la part de liberté individuelle laissée au citoyen ? Penses-tu que par exemple, le citoyen doit parfois pouvoir enfreindre les lois pour se faire entendre (grève de l'impôt, manifestations non autorisées, etc.) ?

Nicolas : Je n'aime pas l'idée qu'on soit bloqués sur tout. Ça me fait peur même. J'aime bien garder mon libre arbitre même si à un moment donné je sais que je fais une connerie tu vois, c'est le propre de l'être humain. Si on veut toujours sécuriser tout et ne plus jamais prendre aucun risque, à quoi va-t-on ressembler dans 100 ans ? On sera complètement aseptisés. Si tu regardes un peu le passé, si tu t'intéresses un peu à la période des années 50 et 60, 70, eux avaient déjà une liberté qui était beaucoup plus grande que la nôtre sur plein de plans. Je prends la cigarettes par exemple, on pouvait fumer partout. Alors je ne dis pas que c'était mieux avant, je dis juste qu'il y avait beaucoup plus de libertés et ce que je constate même, je ne vais pas donner la cigarette en exemple car c'est un peu un mauvais exemple, que les personnes qui ont vécu dans les années 70, regrettent parfois un peu plus cette liberté. Aujourd'hui on est dans une époque où il y a de moins en moins de libertés, de plus en plus d'instantanéité et pour ces raisons je suis plutôt contre ce genre d'objet.

Alexis : Je suis tout à fait d'accord avec toi. Pourtant nous fait miroiter qu'on est dans des démocraties de plus en plus avancées alors qu'en fait on a beaucoup moins de libertés qu'avant.

Nicolas : Oui j'en ai vraiment l'impression. Tiens, je pense à quelque chose. Est-ce que tu as déjà regardé la série Black Mirror ?

Alexis : Oui, j'ai vu quelques épisodes mais pas tout.

Nicolas : Parce que c'est une série qui attaque quand-même très fort ce genre de sujet. Je pense à un épisode où tout le monde a une cote sur son smartphone, les gens se donnent des notes.

Alexis : C'est ce qu'il font un peu en Chine

Nicolas : oui exactement et dans Black Mirror, ils l'avaient un peu poussé à l'extrême et en fonction de ta cote, si tu n'avais pas un 4 sur 5 quand tu voulais louer une voiture par exemple, on ne te donnait que les vieilles voitures toutes pourries. Tu voyais que les gens vivent dans un espèce de monde aseptisé, parce que plus personne n'ose être franc, plus personne ne s'insulte. L'être humain perd un peu de son identité, de son ADN parce que tu rentres dans un espèce de monde robotisé. Une série comme black Mirror m'a fait prendre conscience un peu de ça même si c'est poussé à l'extrême. Mais en même temps tu te rends compte que cette extrême n'est pas si loin.

Alexis : Effectivement tu as raison , cela fait réfléchir. Alors j'ai juste encore 2 questions un peu plus formelles à te poser. Y-a-t-il des informations que tu m'as transmises en réponse à ce questionnaire qui sont confidentielles et ne peuvent être qu'utilisées anonymement pour mon mémoire et Est-ce que je peux te citer dans mon mémoire ?

Nicolas : Non aucune information n'est confidentielle, sinon je te l'aurais signalé au moment de ma réponse et oui bien sûr, tu peux me citer, aucun soucis.

Alexis : Voilà, l'interview est terminée, je te remercie vraiment beaucoup pour toutes tes réponses et de m'avoir accordé de ton temps. Cela me sera très utile pour mon mémoire et il y a beaucoup d'infos que je vais pouvoir intégrer.

Nicolas : c'est avec grand plaisir Alexis, c'est un sujet super intéressant, et n'hésite pas à me contacter si tu as encore des questions. Bon courage pour la suite.

Alexis : C'est super gentil, merci beaucoup, au revoir Nicolas.

		L'éducation	La fraude fiscale et sociale	Le trafic routier	Les objets connectés sécurisés par leur conception
<i>1. les compétences</i>		1	2	2	2
<i>2. Le modèle économique</i>		1	2	2	1
<i>3. Le modèle social</i>	<i>Ajustements qualitatifs</i>	2	3	3	2
	<i>Ajustements quantitatifs</i>	2	2	2	2
<i>4. La fiabilité</i>		1	1	1	1
<i>5. La culture décisionnelle</i>		1	1	1	1
<i>6. Les libertés individuelles</i>		1	2	1	1
<i>7. La nécessité</i>		3	3	3	2
<i>8. La légalité</i>		2	3	1	1

5. Interview 5

Répondant : Monsieur Stéphane Faulkner, consultant en technologies et informatique, donne également des cours à propos du Big Data à l'université de Namur.

Alexis : Bonjour Monsieur Faulkner, merci beaucoup de prendre le temps de répondre à quelques questions pour la rédaction de mon mémoire dont le sujet est « Au travers des quatre grandes tendances Big Data identifiées dans le secteur du Big Data électoral (personnalisation, prédictibilité, optimisation, objectivation), nous avons développé des hypothèses d'application d'une possible Data-driven Governance pour les secteurs suivants : l'éducation ; la fraude fiscale et sociale ; le trafic routier ; les objets connectés sécurisés par leur conception »

Stéphane : D'accord

Alexis : Le but de ce questionnaire est d'avoir une opinion extérieure, de la part de gens qualifiés qui présentent un intérêt particulier dans le domaine, sur la possibilité ou non de mener une politique de gouvernance dirigée par les données en Belgique dans les quatre secteurs évoqués ci-dessus. Je vais d'abord vous poser quelques questions plus générales dans chaque domaines et ensuite quelques questions plus spécifique.

Stéphane : D'accord, allons-y

Alexis : J'ai également d'abord une question spécifique pour vous. Pourriez-vous m'expliquer votre parcours et pourquoi vous enseignez le Big Data à l'Université ?

Stéphane : J'ai fait des études d'ingénieur civile en électromécanique, ensuite j'ai travaillé pendant 2 ans dans le privé et j'ai fait une spécialisation en informatique. J'ai eu l'opportunité de recommencer une thèse de doctorat à l'UCL, et cela dans le domaine du Data. Il faut savoir que l'informatique à l'UCL se trouve en pré technique, sauf qu'à l'époque, le domaine du Data avait été mis à la LSM. Et donc le professeur qui avait été engagé par l'école polytechnique de l'UCL pour le Big Data était à l'époque, Monsieur Alain Pirotte, était localisé à la LSM. Donc, moi j'ai fait ma thèse au niveau du Data avec Alain Pirotte et c'est comme ça que l'équipe a été lancée à l'UCL. Et alors par la suite j'ai beaucoup travaillé avec Manuel Kolp qui revenait du Canada où il avait travaillé dans une équipe de recherche très prestigieuse et où il a travaillé sur une méthodologie de développement logiciel par agent qui était hyper innovant et qui reste aujourd'hui une des méthodologie les plus utilisées dans le monde. Cela s'appelle Tropos, J'ai travaillé et terminé ma thèse avec Manuel Kolp.

Alexis : Ok, merci, je vais passer aux question générales. La première est comment définiriez-vous la notion de « Data Driven Governance », soit la gouvernance dirigée par les données ?

Stéphane : Moi je la définirait comme étant aujourd'hui, plus une suite ou une application de ce qu'on connaît dans la programmation informatique, c'est-à-dire pouvoir se baser et pouvoir exploiter de plus en plus de données sous format numérique, ces données proviennent vraiment de différentes sources qui sont hétérogènes, donc des données sous différents formats. Aujourd'hui évidemment on bénéficie de beaucoup plus de sources qu'au paravent puisqu'aujourd'hui on a des sources avec des bases de données institutionnelles mais on a aussi toute une série de sources qui sont non institutionnelles, comme les réseaux sociaux, des collectes de données consommateurs et ainsi de suite. Et donc sur base de ces données numériques on parvient à structurer une information qui est nettement plus vaste et qui a nettement plus de dimension, et permet sur base de ça, de pouvoir mieux décider. Et donc à partir du moment où les décisions sont meilleures dans un environnement de données qui est exploitable, tel que celui-là, et bien, effectivement, la gouvernance qui s'appuyait auparavant sur d'autre processus, va s'appuyer de plus en plus sur un processus dans lequel on a véritablement de l'exploitation de données. Et donc il y a une évolution technologique qui fait en sorte que la décision, ou en tout cas le spectre sur lequel s'appuie la décision, est différent, et je trouve que quelque part quand il y a décision, il y a gouvernance, et donc aujourd'hui la gouvernance peut être dirigée par les données. Alors je n'aime pas tellement le terme « dirigée par les données » parce que je crois quand-même que le décideur reste libre de sa prise de décision. Mais en tout cas les données contribuent de plus en plus fortement à la prise de décision. Et c'est ce que je vois moi, typiquement dans ma vie de tous les jours quand je suis en entreprise, je constate que les décideurs, que ce soit pour des décisions opérationnelles ou stratégiques, s'appuient de plus en plus sur une multitude de données qui se trouvent dans des tableaux de bord structurés, construits à travers le temps.

Alexis : OK parfait merci beaucoup pour votre réponse. Ensuite quelle est votre opinion concernant l'utilisation de la pratique du Big Data dans le cadre de la gouvernance publique en général ?

Stéphane : C'est toujours la question de l'éthique et de la propriété de la vie privée. Je veux dire, si les données sont exploitées en respectant principalement les règles de RGPD ou en respectant un certain nombre de règles éthiques moi je crois que c'est une bonne chose. Je crois que ça peut contribuer de manière à améliorer les choses, à mieux connaître un certain nombre d'objets, dans le monde de l'éducation, dans le monde de la santé et ainsi de suite, qui permet d'apporter des évolutions. Mais encore une fois, pour moi tout ça doit être bien contraint par des règles strictes et des règles législatives telles que RGPD. Il y a aujourd'hui beaucoup d'interprétations, les choses ne sont pas encore très claires à ce niveau-là. Donc, moi, je suis quand-même favorable à l'utilisation de ce Big Data.

Alexis : Merci, la question suivante est où situez-vous le niveau de la Belgique en particulier, de façon générale, quant à l'utilisation du Big Data dans le domaine de la gouvernance publique (comparé à d'autres pays Européens) ?

Stéphane : Moi j'ai une vision assez claire, c'est une vision personnelle mais elle est assez claire. Alors, dans le Big Data, la première dimension du Big Data, celle qu'on a mis publiquement en avant, c'est la gestion des masses de données. La gestion des masses de données nécessite véritablement d'aborder l'analyse de données sous d'autres angles. Pourquoi ? Parce qu'on a des masses de données qu'on appelle du calcul distillé avec la, du calcul mathématique et ainsi de suite. Cette dimension-là, la dimension sur la base de données, en Belgique je n'y crois pas. A ma connaissance, il n'y a pas d'environnement. Il y a le souci de mettre véritablement toutes les techniques du Big Data en avant pour gérer les quantités de données. Je pense qu'on peut continuer avec des processus classiques de traiter des données telles qu'on les a en Belgique. Parce qu'il n'y a pas suffisamment de masse de données, on n'a pas des sociétés comme Google, Amazon et ainsi de suite qui produisent des millions et des millions de données en streaming. Par contre, où le Big Data prend tout son sens, c'est pour moi sur une autre dimension des données. C'est-à-dire qu'à côté de la dimension masse de données, il y a une dimension hétérogénéité des données. Ce que je veux dire c'est que de plus en plus de sources d'informations, qu'on a envie d'interroger, d'exploiter vont se retrouver sous différents formats. On va retrouver des données très structurées, comme on a toujours eu dans les bases de données, dans les fichiers Excel ou des choses comme ça. On va trouver des bases de données semi-structurées par exemple sur [XXXX.net](#), qui se retrouvent sous forme d'une base de données mais malgré tout qui gravitent autour du Web, et alors il y a les données complètement non-structurées qu'on va retrouver sur les réseaux sociaux, des capteurs, des vidéos et ainsi de suite. Ça veut dire que cette multitude, cette hétérogénéité de données fait que on a dû changer dans le domaine de l'informatique, nos techniques de manipulations et d'exploitation de données. Et c'est là où le Big Data a pris tout son essor. Pour moi la vraie dimension du Big Data dans les premières années de son utilisation dans le public, dans une grande partie du monde, elle est là. Principalement on utilise le Big Data comme un outil d'exploitation de sources de données diverses avant de l'utiliser comme un outil d'exploitation de données de masse. Pourquoi ? Parce que je peux avoir de grandes quantités de données, comme dans le passé, que j'aurais pu manipuler avec des techniques PI, sauf si je suis contraint à devoir manipuler ces données complètement hétérogènes. Et donc sur cette base-là, j'utilise les techniques du Big Data.

Alexis : D'accord, parfait, ensuite, est-ce qu'il y a des secteurs, selon vous pour lesquels l'utilisation des données a vraiment une qualité et au contraire des secteurs pour lesquels vous pensez qu'il ne faut pas le faire ?

Stéphane : Je pense que pour tous les secteurs, à un moment donné, un environnement de données peut aider à prendre de meilleures décisions. A partir du moment où on a des données qui sont exploitables et où on a surtout des gens qui sont suffisamment compétents pour rendre ces données potables et pour les exploiter. Les rendre exploitables cela veut dire qu'ils sont susceptibles de pouvoir analyser ces données et la qualité de ces données. Et là c'est quand-même toute une question aujourd'hui qui taraude encore toute la recherche dans les données. Il faut savoir qu'il y a beaucoup de thèses de doctorat, moi je suis occupé par exemple avec 2 doctorant à travailler sur cette dimension de qualité de données par rapport à des secteurs tels que la médecine. Il faut des gens qui soient susceptibles de pouvoir contrôler cette qualité de données à la source mais également après, lorsqu'elle ont été retravaillées, reformatées pour l'analyse. A partir du moment où on a cette plus-value de compétence, tous les secteurs devraient bénéficier effectivement d'une information complémentaire provenant d'un environnement de données.

Alexis : OK, merci, je passe maintenant à la dernière question générale. Quel futur imaginez-vous pour la « Data Driven Governance » ? Pensez-vous que cette tendance va plutôt s'accroître à l'avenir, ou, au contraire, que c'est un effet de mode qui va disparaître, et pourquoi ?

Stéphane : Je pense vraiment que ça va s'accroître. Les prochaines générations seront de plus en plus confrontées à l'exploitation de leurs données, leurs données de consommation et ainsi de suite. Je pense qu'on va de plus en plus trouver des applications de la vie de tous les jours, des systèmes de recommandations qui ne pourront fonctionner qu'à partir du moment où on capture des données sur l'individu. Donc je crois qu'on est juste au début de ce processus, qu'il va encore s'amplifier dans des applications de tous les jours. On parlait des voitures autonomes, des applications qui vont jusqu'à des recommandations alimentaires au niveau de la gestion d'un frigo et ainsi de suite. Il faut savoir aussi une chose, c'est qu'on parle beaucoup d'intelligence artificielle. Il faut savoir que 80% de ces applications d'intelligence artificielles ont été développées sur base de données.

Alexis : Merci pour vos réponses. On va maintenant passer aux questions sur l'éducation. La première question est assez générale. Quelle est votre opinion sur le niveau global de l'enseignement en Belgique ?

Stéphane : Alors, je pense que malgré tout on garde un bon niveau. Alors je sais que dans les sondages et les études qui sont fait au niveau européen on voit que chaque année on perd des places. Mais je reste assez optimiste par rapport au niveau global. Je vais d'abord parler au niveau francophone. Je pense qu'au niveau francophone je crois qu'il y a de plus en plus d'homogénéité au niveau des universités. Il y avait par le passé encore des différences philosophiques, des différences de niveau avec des facultés qui accentuaient plus l'aspect scientifique dans leurs études par rapport à des études en sciences humaines et tout ça. Aujourd'hui avec justement, des référentiels, des référentiels compétents, il y a de

moins en moins de différences. Donc j'ai l'impression que le niveau est en train de s'homogénéiser au maximum. Moi ce qui m'interpelle plus au niveau des études aujourd'hui, c'est la différence que j'observe entre les parcours d'études en Flandre et les parcours d'étude en Belgique francophone. Je pense réellement que les études en Flandre sont plus compliquées, plus sélectives. Ils ont quand-même un peu une formation plus solide. D'ailleurs il suffit de le voir aussi dans le cadre par exemple de Smart data, quand on voit par exemple au niveau de la recherche qui se fait, les plus grosses start-up qui décollent vraiment au niveau Belge sont plutôt du côté néerlandophone. C'est peut-être une question de moyens mais derrière ces moyens se cache de plus en plus une école à 2 vitesses avec un enseignement néerlandophone, qui prend, je trouve personnellement, le pas sur l'enseignement francophone.

Mais de manière générale, en Belgique, pour moi le niveau est quand même respectable. La preuve c'est que, nos étudiants en Erasmus, on n'a aucune crainte de les envoyer dans les meilleures universités européennes. Même vers le Nord, où on dit souvent qu'il y a plus de rigueur, ce sont des Erasmus plus compliqué où il faut beaucoup travailler mais ils réussissent.

Alexis : Merci beaucoup, c'est très intéressant. La question suivante est selon-vous, est-il nécessaire pour l'Etat d'investir massivement dans l'enseignement afin d'équiper les écoles et les universités de plus d'outils numériques, qui sont un préalable à l'utilisation d'une Data-driven Governance dans le domaine de l'éducation ?

Stéphane : Cela c'est clair car en fait le problème qu'on va avoir aujourd'hui c'est qu'on va avoir de plus en plus de génération qui vont être éduquée à l'intuition de l'usage du numérique parce qu'ils ont des smartphone, des jeux et qu'ils passent énormément de temps là-dessus. Mais mon souci à moi c'est de faire sortir les jeunes de cet esprit ludique C'est-à-dire qu'il y a effectivement une informatique ludique, celle qu'on utilise nous au travers de nos applications tous les jours, qui nous permettent effectivement de bénéficier d'un ensemble de services. Pour les jeunes ce sont des jeux, pour des adultes ça peut être des application où on commande son repas du soir par exemple. Mais au-delà de ça il y a une informatique professionnelle. Donc pour moi il va devenir difficilement envisageable pour une grande partie de la population de pouvoir être compétent sur le marché du travail sans avoir accès à des outils informatiques. Et c'est là où l'éducation doit jouer son rôle. C'est à dire que le système éducatif doit permettre de sortir l'enfant de cette sphère ludique de l'informatique vers une sphère plus éducative basée sur l'acquis de compétences et ainsi de suite. Pour moi c'est très important.

Alexis : D'accord, merci, je passe à la question suivante. Des applications Big Data tels que la prédiction des performances d'un élève, l'ajustement en temps réel des programmes scolaires, le repérage de données communes entre les élèves en échec scolaire, etc. pourraient-elles, selon-vous, améliorer la qualité de l'enseignement et augmenter le niveau des élèves en Belgique ?

Stéphane : Je XXX sur le point de l'automatisation. De tout processus d'automatisation, de l'interaction, homme/machine, ici système éducatif/élève. Moi je pense concrètement qu'un système éducatif qui s'appuierait principalement sur une diminution de l'interaction entre les professeurs et les étudiants mais beaucoup plus sur la machine des étudiants, je n'en veux pas car pour moi l'interaction humaine est très importante sur l'aspect social. Donc si l'école ne remplit plus son rôle de sphère sociale où l'enfant va découvrir la diversité, où l'enfant va découvrir l'autre et ainsi de suite alors je pense qu'on rentre dans une société qui est très dangereuse. On parle beaucoup de e-learning, on en vient quand-même à dire que la plus part des études à ce sujet, démontrent qu'à partir de plus de 2 heures par jour, l'e-learning devient tout à fait inefficace par rapport à l'apprentissage en présentiel. Donc il faut faire très attention à ça.

Alexis : OK, merci beaucoup, c'est typiquement le genre d'information pertinente pour mon mémoire. La question suivante est Quel est votre avis concernant une éducation plus « personnalisée » en matière d'enseignement, par opposition à l'enseignement « standardisé » pratiqué aujourd'hui ?

Stéphane : Là on rentre dans ce qu'on appelle, l'optimisation des parcours scolaires par les **hybridation**. On peut effectivement jouer avec différentes modalités pédagogiques. On a des modalités pédagogiques qui peuvent s'appuyer sur du présentiel, on peut passer sur du distanciel avec du numérique. Un des avantages qu'on tire de là, c'est d'aller jusqu'à l'individuel. C'est-à-dire qu'au travers d'un parcours scolaire, la plupart des enfants vont travailler à un certain rythme et certains vont se démarquer par rapport à ce rythme. Certains vont vouloir aller plus vite, d'autres vont vouloir aller moins vite Et donc, effectivement ça, car le professeur ne peut pas se démultiplier, on pourrait le travailler à l'aide d'outils numériques. Et donc là effectivement c'est intéressant, parce que c'est là que l'outil numérique prend tout son sens pour moi. Dans l'apprentissage, cela veut dire qu'on peut compléter l'approche d'un professeur qui travail en présentiel ou en interaction globale avec ses étudiants par un outil qui permet d'avancer sur l'approche individuelle, donc une approche singulière. Je trouve que ça c'est intéressant tout en gardant des garde-fou comme laisser un pourcentage de présentiel et un pourcentage de distanciel. Il y a une autre garde-fou, on est quand-même dans une société où les parents ont tendance à projeter énormément de choses sur leurs enfants. Je suis quand-même étonné de voir autour de moi le nombre d'enfants surdoués que j'ai. C'est-à-dire que la plupart de gens que je connais, que je fréquente, on tous repéré dans leurs enfants des surdoués. On est dans une société où on a tous l'impression que nos enfants sont surdoués. Donc on aurait tendance dans un mécanisme comme ça, s'il n'y a pas de règles claires, de pousser l'enfant à se battre, à apprendre par lui-même, d'avancer trop vite. Il existe un rythme scolaire, et pour moi, 99% des gens doivent s'inscrire dans ce rythme scolaire. Et de temps en temps il y a effectivement des gens qui peuvent sortir de ce rythme scolaire mais ça ne fait pas partie de la norme. Et ce que j'ai peur avec un système hybride comme ça, c'est qu'on surdose ça et que ça peut

faire des catastrophes. C'est un peu comme quand on parle de dopage dans le sport, ça pourrait être du dopage. Et à un moment donné on pourrait avoir des gens qui explosent de toute part et qui n'en peuvent plus.

Alexis : D'accord, ça rejoint un peu la question suivante qui est en admettant l'hypothèse d'un enseignement « personnalisé » pour les élèves, à quels niveaux de l'enseignement (maternel, primaire, secondaire, universitaire) pensez-vous qu'il serait le plus utile de l'appliquer et pourquoi ?

Stéphane : Pour moi on peut le faire à tous les niveaux mais je pense que le système éducatif doit rester maître de ça. C'est-à-dire que si on donne cet outil là uniquement dans les mains des parents ça va être une catastrophe. Je pense qu'on doit vraiment avoir un système éducatif alors avec des comités de pilotage, des comités de pédagogues, qui évaluent, qui donnent la recommandation et qui à la limite même donnent accès aux ressources parce que sinon on risque d'avoir un emballement.

Alexis : Parfait, je vais maintenant passer à l'avant-dernière question sur l'éducation. Dans le cadre de la crise du COVID-19, avez-vous observé/utilisé des applications orientées « Big Data » dans l'enseignement, que ce soit en tant que parent, professeur ou personne neutre ? Si Oui, lesquelles ?

Stéphane : On a utilisé des applications numériques mais pas Big data en tant que tel. J'ai accentué l'outil numérique ça oui. J'ai énormément appris aussi, moi, je suis prof d'informatique et j'ai énormément appris sur l'usage de ces outils. Je pense qu'aujourd'hui, effectivement, je suis capable de mieux donner cours en distanciel. Je pense que pour l'après Covid je resterai avec mes étudiants universitaires dans un mode hybride mais où le présentiel aura toujours sa place. Le numérique ne va pas remplacer le présentiel mais ce sera un complément.

Alexis : Donc la plateforme d'examens Testwe qui détecte , en analysant le regard des élèves pour voir s'ils trichent, vous ne l'avez pas utilisée ?

Stéphane : Nous à Namur, on ne l'a pas utilisée. A Namur, on a pris la décision de ne pas sortir les étudiants des outils habituels qu'ils connaissaient pour ne pas les perturber durant l'examen. Notre philosophie a été de dire, voilà, nos étudiants ont appris durant leurs cours à utiliser 2 environnements numériques, d'une part Moodle et d'autre part Teams, et donc les examens doivent se faire selon ces modalités-là. Alors évidemment, et on le sait bien, la peur de l'enseignant par rapport à ça, c'est de se dire oui il va y avoir des zones d'ombre et peut-être des zones de triche. C'est-à-dire qu'on va plus baser cela sur 2 dimensions, la confiance et puis une autre dimension c'est l'évaluation au travers de l'outil tel quel qui doit être réfléchi pour éviter au maximum la triche. Ou bien il y a effectivement le système

sur lequel différentes institutions se sont appuyées, comme Testwe. Il existe aussi d'autres applications de ce types-là.

Alexis : OK, merci, la dernière question au sujet de l'éducation est : Quel est votre avis concernant la potentielle utilisation des données des élèves (relevés de notes électroniques, journal de travail numérique, statistiques de réussite) par d'une part les universités, afin d'orienter/recruter les élèves dès l'enseignement secondaire et de leur proposer un cursus qui, de par une étude dirigée par des algorithmes, pourrait leur correspondre ?

Stéphane : Pour moi c'est très dangereux car cela peut mener à 2 clivages. Ça pourrait mener à un problème social. Il ne faut pas se le cacher. Ça veut dire encore une fois des modèles à la française où les grandes écoles ont pris le dessus sur les universités et où le concours d'accès devient tellement difficile que si vous n'êtes pas inscrit dans les bonnes options, dans le bon environnement et dans la bonne école dès vos 12 ans, et bien la chance de pouvoir rejoindre ces écoles-là est très faible. Et donc là je pense plutôt à un clivage social qui ferait que certains enfants qui ne seraient pas armés parce que les parents n'auraient peut-être pas les moyens de les mettre dans des toutes bonnes écoles ou n'auraient peut-être pas les moyens de les éduquer rapidement à être dans les bonnes filières et ainsi de suite, ne puisse pas un jour pouvoir exploiter tout ce qui est exploitable. Le vrai danger serait de voir grandir un gap entre des gens qui ont des facilités à rejoindre rapidement un bon niveau d'enseignement et les autres. J'aime beaucoup le système belge où finalement on laisse la possibilité à l'étudiant d'étudier, même à certains qui ont peut-être eu des parcours plus compliqués ou qui n'ont pas été préparés. Ils vont peut-être mettre plus de temps à devenir performant mais à un moment donné ils arrivent à être performants et parfois plus performants que d'autres.

Alexis : d'accord, et l'autre partie de la question c'est l'utilisation de ces données par l'Etat, afin d'orienter les élèves vers des études qui mènent à des débouchés, en fonction de l'offre et la demande sur le marché de l'emploi ?

Stéphane : Là par contre c'est une question qui est très compliquée parce que aujourd'hui effectivement on a une tendance à dire, et bien voilà, les enfants font des études universitaires, ça nous tranquillise et ainsi de suite. Mais il y a une vraie tendance, certaines études universitaires aujourd'hui ont très peu ou pas de débouchés. Moi je vois effectivement des gens qui ont fait 5 ans d'université et qui se retrouvent sur le carreau. Ils ne se retrouvent pas sur le carreau un an ou deux ans mais 3 ou 4 ans et qui sont pré-décrochage par rapport à ce pourquoi ils sont sur le marché de l'emploi et qui finalement prennent des boulots alimentaires et qui se trouvent sur une voie sans issue. Là je trouve quand-même qu'il y a un rôle important mais qui est un rôle plus d'information. Je suis toujours étonné finalement que l'information qu'ont certains parents par rapport aux débouchés des études pour lesquelles ils viennent.

Nous dans le domaine de la gestion, en ingénieur de gestion, moi je dis souvent qu'il n'y a pas trop de problème car on est dans des études dont les débouchés ne se posent même pas car on est polyvalent. Mais il y a quand-même d'autres études où c'est beaucoup plus compliqué. Il faudrait avoir l'honnêteté d'informer les parents à ce niveau-là.

Alexis : Très bien, merci beaucoup pour toutes vos réponses à propose de l'éducation.

Stéphane : Je peux vous demander une chose ?

Alexis : Oui

Stéphane : Est-ce qu'on pourrait faire la suite de l'interview à un autre moment car je n'ai plus que 2 pourcent de batterie ?

Est-ce qu'on peut se rappeler à un autre moment de la journée ?

Alexis : Oui, on peut se rappeler dans la journée, pas de problème,

Stéphane : Et bien on se rappelle dans la journée

Alexis : OK merci beaucoup, Au revoir, à tout à l'heure

Stéphane : Oui au revoir.

Stéphane : Bonjour Alexis, voilà, je suis sur la route, là

Alexis : On peut continuer l'interview ? Vous vous souvenez où on en était arrivé hier ?

Stéphane : Oui, on avait terminé le sujet de l'éducation je pense.

Alexis : Oui c'est bien ça. Je vais maintenant passer au sujet suivant qui est la fraude fiscale et sociale. La première question est plus générale, quelle est votre opinion générale sur la fraude fiscale et sociale en Belgique ? Que pensez-vous des efforts faits par le gouvernement belge à ce sujet ?

Stéphane : Je n'ai pas beaucoup d'avis là-dessus, si ce n'est que je pense que le gouvernement est en train de mettre en place depuis quelques années des outils qui sont de plus en plus performants et qui croisent de plus en plus de données. Je pense qu'ils commencent à avoir des bases de données structurées qu'ils partagent entre organismes, que ce soit au niveau de l'impôt des personnes physiques, comme des

sociétés, tout ce qui est ONSS et ainsi de suite. Ils commencent à avoir des données qui sont de plus en plus précises et complètes sur chaque individu.

Alexis : D'accord, la question suivante c'est selon-vous, est-ce que la lutte contre la fraude fiscale est un combat prioritaire à mener ou plutôt une fatalité dans un contexte de mondialisation et de globalisation de l'économie

Stéphane : Je pense que c'est un des combats à mener parmi d'autres. Je veux dire l'obsession de la lutte fiscale, pour moi c'est plus un symbole politique et c'est plus une promesse politique qu'autre chose. Je crois qu'il y a une multitude de combats à mener, celui-ci en fait partie, maintenant je pense qu'un Etat qui serait obsédé par la lutte fiscale serait un Etat qui se fermerait la porte de pas mal d'opportunités économiques. Cela ferait peur, je pense, à pas mal d'entreprises.

Alexis : D'accord, merci. Ensuite, estimez-vous que l'utilisation du Big Data et des données personnelle pourrait être une arme efficace de lutter contre la fraude fiscale et sociale ? Si oui, pourquoi et comment ?

Stéphane : Oui, ça j'en suis certain par contre. La plupart des comportements économiques peuvent être tracés au niveau data. L'Etat la déjà fait hein, le plus possible pousser des gens à consommer des produits, qu'au travers de paiements en ligne et de paiements numériques. Je pense que c'est déjà le cas aujourd'hui puisqu'on a une limite de 3.000 EUR pour les paiements en liquide. Donc on va commencer à tracer les comportements économiques de chaque individu.

Alexis : D'accord, très bien merci. La prochaine question est d'un point de vue pratique, légal et éthique, que pensez-vous d'opter pour une politique proactive en matière de fraude, c'est-à-dire de l'empêcher et de la prévenir plutôt que de la sanctionner, comme c'est le cas aujourd'hui ? Des algorithmes prédictifs permettent aujourd'hui d'identifier les fraudeurs potentiels à priori et non plus à posteriori, pensez-vous que l'utilisation par l'Etat de ce type d'instruments est réaliste et réalisable dans notre société ?

Stéphane : Pour moi il doit y avoir 2 dimensions. Il doit effectivement y avoir la dimension de sensibilisation et de proactivité et quand-même une dimension offensive parce que je suis certain que si on oublie cette dimension offensive on va oublier un certain nombre d'agents qui ont les moyens de faire de la déductibilité fiscale très avancée et qui passeront les mailles du filets de tout ce qui est algorithmes, de prédiction et de classification à grande échelle de fraudeurs potentiel.

Alexis : OK, merci. La dernière question sur le sujet est quel est votre avis concernant le système de cartes prépayées « contrôlées » utilisées aux États-Unis, qui interdit aux personnes bénéficiant

d'allocations sociales certaines dépenses (alcool, salles de jeux)? Est-ce que ce système serait transposable en Belgique ?

Stéphane : Sans doute, maintenant je pense qu'on vit dans un modèle économique et social qui est différent. Je pense qu'on croit plus en la responsabilité de tout un chacun que d'imposer cette responsabilité comme on le fait aux États-Unis. Je crois qu'en Belgique on ne verrait pas le jour de ce genre de choses. Pour moi on est dans une culture qui a une approche sociale complètement différente.

Alexis : D'accord, merci beaucoup. Je vais maintenant passer aux question du sujet suivant qui est le trafic routier.

Stéphane : Ok allons-y

Alexis : Quel est votre ressenti par rapport au trafic routier en général en Belgique ?

Stéphane : Écoutez, moi je suis bruxellois, et je peux vous confirmer que le trafic routier à Bruxelles deviens de plus en plus difficile. Je pense c'est une vraie catastrophe. Je pense vraiment que les politiques de mobilité sont très très mal pensées en Belgique. On nous fait croire qu'il y a des stratégies, mais derrière tout stratégie, pour moi, il y a toujours une opérationnalisation de la stratégie, et j'ai l'impression qu'en Belgique on a des stratégies qui partent à chaque fois dans tous les sens en fonction de la couleur politique, et derrière ça, avec des modèles d'opérationnalisation complètement différents. Pour ne citer que ça, on voit effectivement à Bruxelles, que d'année en année on perd des places de parking, on perd des voies de mobilité soit disant au profit des transports en commun mais je constate quand-même qu'on n'a pas de vraie solution en terme de transports en commun au niveau Bruxellois. Mes enfants, qui doivent parfois prendre 2 transports en commun différents pour se rendre à l'école, peuvent des fois perdre trois quart d'heures parce qu'il n'y a pas suffisamment de métros, parce qu'il y a une congestion des métros et ainsi de suite. Je trouve que c'est dramatique.

Alexis : C'est en effet ce que beaucoup de gens ont répondu. Donc la question suivante c'est l'utilisation intelligente des données disponibles (géographiques, météorologiques, capteurs de vitesse, etc.) pourrait-elle être un moyen d'améliorer significativement la fluidité du trafic routier en Belgique selon-vous ? Ou pensez-vous que la congestion est plutôt d'un problème structurel (infrastructures des routes) et que d'autres solutions seraient préférables (home working le matin, transports publics, etc.) ?

Stéphane : Je serais plus nuancé. Il y a un vrai problème structurel en Belgique. Alors c'est très bien, on peut essayer de valoriser le home working, c'est peut-être une des voies de solutions. Mais moi je ne crois pas à une solution exclusive, je crois effectivement à un vecteur, un spectre de solutions qui,

combinées entre elles, vont commencer à diminuer significativement les problèmes de congestion à long terme de mobilité. Oui home working OK, je pense que effectivement, l'analyse basée sur des données avec des chiffres de recommandation en terme de trafic pourrait aussi contribuer à améliorer ça. Et donc pour permettre effectivement de bénéficier d'un réseau de transports en commun beaucoup plus élargi et beaucoup plus performant au niveau Bruxellois aussi, peut-être jouer aussi, plus sur la taille des véhicules. Il faut peut-être penser mobilité différente au niveau des véhicules. Aujourd'hui on a quand-même le luxe de pouvoir rouler dans des véhicules très grands, très agréables, très spacieux mais peut-être que ce n'est plus la mobilité de demain. Je pense plus à un ensemble d'actions qui vont contribuer à l'amélioration du trafic, qu'à une action unique.

Alexis : D'accord, donc la question suivante c'est afin d'optimiser la Data Driven Governance, nous évoquons le partage des données types « Open-Data » du secteur public vers le secteur privé mais également du secteur privé vers le secteur public, ce qui est peu le cas aujourd'hui. Pensez-vous que l'ouverture des données du secteur privé est un impératif pour mener à bien une Data Driven Governance dans le domaine du trafic routier, ou considérez-vous cela comme une utopie ?

Stéphane : Je pense qu'effectivement, il va falloir trouver des partenariats avec les entreprises. C'est du win-win, c'est ce qu'on appelle des contrats PPP (Private Public Profits ou Partnership). Et donc en fait l'idée c'est effectivement le win-win parce que en permettant une meilleure mobilité dans Bruxelles, on permet aussi aux entreprises d'avoir des atouts parce qu'une entreprise qui est bloquée par la congestion du système routier et bien je pense qu'il n'y a aucune entreprise qui est preneur de ça, et donc je crois qu'il faut aller trouver des solutions win-win qui permettent à la fois à l'Etat de faire de meilleures recommandations et aux entreprises d'en bénéficier au travers d'une mobilité plus fluide.

Alexis : Donc vous pensez que les entreprises privées seraient prêtes à partager leurs données ?

Stéphane : Je pense qu'une entreprise privée, à partir du moment où on développe un projet, et qu'au travers de ce projet on sait montrer que l'entreprise est gagnante, elle sera toujours prête.

Alexis : Et vous pensez que les entreprises feraient ça sans contrepartie réelle comme par exemple des avantages fiscaux ?

Stéphane : Moi je pense qu'il ne faut même pas aller jusque-là. Je pense qu'en Belgique, on a trop l'impression d'avoir une image des entreprises qui doivent toujours tout financer, monnayer, moi je ne pense pas. Je peux en parler car je suis moi-même fondateur de plusieurs entreprises, moi si on vient me trouver avec des bons projets, je n'ai pas besoin qu'on monnaye ces projets à partir du moment où j'y

trouve un bénéfice. Ce bénéfice peut être de l'ordre du financier mais il peut aussi être de l'ordre, justement, de la qualité de travail. Ici je parle plutôt en terme de qualité de travail.

Alexis : Très bien, je suis assez étonné de votre réponse car personnellement, je pensais que les entreprises privées étaient plutôt réticente à partager leurs données.

Stéphane : Tout dépend de quel types de données, mais encore une fois, si on vient avec des projets concrets, des projets où on sait montrer qu'on est dans un partenariat win-win, sur certaines données les entreprises ne seront pas contre. Alors bien évidemment je parle de certaines données. Il ne faut pas aller chercher derrière des données de la mobilité d'autres données qui permettraient de faire des calculs fiscaux pour les entreprises. C'est toujours de ça que les entreprises ont peur.

Alexis : D'accord, merci. On va maintenant passer au dernier sujet qui est les objets connectés sécurisés par conception qui sont des objets qui, de par leur nature, empêchent l'utilisateur d'enfreindre la loi. Prenons le cas des voitures : certaines voitures sont aujourd'hui capables de limiter la vitesse lorsqu'elles détectent, grâce à des capteurs, des panneaux de limitation de vitesse. Que pensez-vous de ce genre d'objets (d'un point de vue pratique, éthique et légal) ?

Stéphane : Je dirais que pour toute avancée numérique, objets connectés ou d'autres technologies, c'est de se dire par exemple, quelque par cette application limite mon indépendance par rapport à ma vie privée, soit par rapport à un certain nombre d'actions. On est dans une société où effectivement on partage de plus en plus de données avec des applications qui vont nous rendre des services. Alors il y a des gens qui disent , ça me fait peur d'utiliser ces applications car je ne veux pas partager mes données privées, mais dans ce cas je leur répondrais il ne faut non plus pas utiliser de GPS, ni Uber par exemple si vous n'êtes pas prêts à ça. Il faut jouer le jeu car on est dans une société où on est dans un partage de données avec des applications qui vont nous rendre des services donc nous avons ou pas des actions où nos données privées vont forcément être transmises, c'est logique.

Alexis : Je ne vous entends pas très bien, il y a souvent des coupures mais j'ai quand-même pu comprendre une partie de votre réponse. La question suivante est quel est, selon-vous, l'avenir des objets connectés sécurisés par leur conception en Belgique ? Pensez-vous que de plus en plus d'objets de ce genre vont voir le jour dans un futur plus ou moins proche ?

Stéphane : Je pense que les grosses évolutions de services sur un certain nombre d'appareillages ou l'évolution technologique de la plupart des appareillages vont se faire au travers des objets connectés. Donc oui, je pense que toute l'évolution de produits aujourd'hui va se faire au travers de la technologie

qui va se faire au travers de l'appareillage d'objet connectés et de la mise en réseau du produit. Je pense que ces objets vont être de plus en plus présents.

Alexis : D'accord, et avez-vous entendu parler d'autres projets pour ce type d'objets dans votre entourage ? Si oui, serait-il possible de les décrire et de les expliciter ?

Stéphane : Il y a par exemple des systèmes de régulation cardiaques aujourd'hui qui permettent de suivre un patient et qui permettent à un moment donné de déclencher automatiquement des électrodes, des ondes de choc au niveau cardiaque pour agir sur la tension et ainsi de suite.

Alexis : OK, parfait, je vais maintenant passer à la dernière question. Que pensez-vous d'une société où énormément d'objets de ce genre feraient leur apparition ? Quel serait encore la part de liberté individuelle laissée au citoyen ? Pensez-vous que par exemple, le citoyen doit parfois pouvoir enfreindre les lois pour se faire entendre (grève de l'impôt, manifestations non autorisées, etc.) ?

Stéphane : C'est plus une question d'Etat. Je pense que aujourd'hui on est dans une société où les citoyens ont la liberté d'échapper à ces technologies mais en échappant à ces technologies ils s'excluent du jeu d'un certain nombre de services qui sont rendus. Et donc la seule chose que j'espère c'est que dans les années futures on laissera quand-même un libre choix au citoyen, de se dire OK je suis prêt à m'exclure d'un certain nombre de services. Mais les services élémentaires comme la santé, l'alimentaire, l'éducation devront absolument être protégés de cette dépendance des applications des objets connectés. Cela veut dire que je peux envisager que dans ces environnements-là, il y en ait, mais je dois envisager aussi qu'on soit libre de les utiliser ou de ne pas les utiliser en fonction de sa volonté personnelle.

Alexis : Donc laisser la liberté au citoyen de se « marginaliser » ?

Stéphane : Par rapport à certains environnements, oui.

Alexis : OK, merci j'ai encore 2 petites questions plus formelles. Pourriez-vous m'indiquer quelles informations transmises en réponse à ce questionnaire sont confidentielles (et ne peuvent être qu'utilisées anonymement pour mon mémoire) et dans l'affirmative, lesquelles ?

Stéphane : Il n'y en a aucune.

Alexis : Et puis je vous citer dans le cadre de mon mémoire ?

Stéphane : Oui oui, bien sûr, avec plaisir

Alexis : Merci c'est super gentil. Je vous remercie en tout cas pour toutes vos réponses à mes questions et d'avoir pris du temps pour y répondre. Cela m'aidera beaucoup pour mon mémoire.

Stéphane : Avec grand plaisir et bonne continuation. Au revoir

Alexis : Au revoir

