Research report

Exploring State Practices and Uses of Big Data Technology

ADC / Asociación por los Derechos Civiles

Privacy Area

ADC / Asociación por los Derechos Civiles

October, 2015

www.adc.org.ar

All activities by *Asociación por los Derechos Civiles* were and are consistent under the Internal Revenue Code Sections 501(c)(3) and 509(a). If any lobbying was conducted by Asociación por los Derechos Civiles (whether or not discussed in this report), Asociación por los Derechos Civiles complied with the applicable limits of Internal Revenue Code Sections 501(c)(3) and/or 501(h) and 4911. Asociación por los Derechos Civiles warrants that it is in full compliance with its Grant Agreement with the New Venture Fund, dated June 9, 2015, and that, if the grant was subject to any restrictions, all such restrictions were observed.

Table of contents

Exploring State Practices and Uses of Big Data Technology*

1. Executive Summary

The use of Big Data is new in Argentina. Some incursions have been made by the scientific and technical academy, the government and companies to develop it. However, generally speaking, we may say that up to now, the concepts of "Big Data", "big data processing" and "massive use of information" are not that present in society in general, nor in public discourse or in civil society organizations. It is important to highlight that these concepts will be developed in a special section in order to introduce the reader to the subject so that they can thoroughly understand the case study.

The importance of this subject is evident, given that, despite the lack of public awareness of almost every aspect of this technology, the velocity with which it is being used (as is the case with every new technological trend) calls for its consideration. Public initiatives such as the case study chosen, as well as the current commercial practices, show that Big Data is here to stay.

This project will allow us, as a civil society that is committed to human rights protection, to collect information on this recent phenomenon and approach it from a human rights perspective. It will also allow us to anticipate Big Data's influence on the massive use of information in order to reinforce the human rights approach, especially with regards to privacy and personal data protection.

Our case study is the Tarjeta Vos (Vos Card), a multifunctional card of the network "En todo estás vos" ("You are everywhere") implemented by the Ministry of Modernisation of the Government of the Autonomous City of Buenos Aires. Even though the card

*This report was produced by Tatiana Anabel Fij, lawyer at ADC. Tatiana graduated as an attorney at the Universidad of Buenos Aires (UBA) in 2007. After that she did a LLM in Intellectual Property. Now she is finishing her thesis on Social Networks sites and its relationship with Intellectual Property. She is member of the Informatics Law and the Author Rights Institute at Bar Association of Buenos Aires (CPACF), she participated as academic collaborator in the Treaty of Informatic Law (Authors: Daniel Ricardo Altmark – Eduardo Molina Quiroga, La Ley, 2012) as well as present at conferences and teach on Informatic Law and Intellectual Property. Also she is a researcher in the project UBACyT 2014-2017 "Tensions generated by the application of Technology Information and Communications (TICs) and the rights to protection personal data, privacy and image".

implementation is recent –it was launched in November 2013–, we consider the case study will allow us to investigate the way in which the local government handles the use of massive data, that is to say: how the government collects, stores, processes, analyses, treats and utilises the data; and how this may potentially impact our democratic society and human rights.

The case study is relevant in connection with the subject of Big Data for many reasons. To start with, it is a local case study (involving the Autonomous City of Buenos Aires), but it may be replicated in the rest of the country. The Autonomous City of Buenos Aires is the capital and most important city of Argentina, as well as one of the most significant cities in Latin America. With this said, what happens in this city generally impacts other cities in the country and the region.

The case study will also allow us to investigate and carry out an exploratory analysis on how the Government of the Autonomous City of Buenos Aires handles (or intends to handle) massive data; and, by means of this first approach, will help us gain insight and acquire tools to replicate this analysis in different scenarios such as the national government practices and private companies in future research projects.

In addition to the abovementioned, it must be considered that this card programme is but another step forward made by the local government towards the implementation of a strong technology development policy designed to turn the city into a "Smart City". This is important because the latter concept has a symbiotic relationship with Big Data.

2. Goals and questions raised by this research

The goals of the proposed study case are as follows:

- To analyse the technical and regulatory structure underlying the implementation of the "VOS CARD."
- To investigate the way in which the Government of the Autonomous City of Buenos Aires collects, stores, processes, analyses, treats and utilizes the data and how this could potentially impact human rights.
- To identify and confront the relevant legal regulations of personal data protection and privacy for their application in the processing of big data.
- To identify cases of best and worst practices in the use of "Big Data."
- To analyse some possible consequences of the decision-making process resulting from large-scale data or its massive use.

These are some of the questions we attempt to answer in this exploratory study on Big Data:

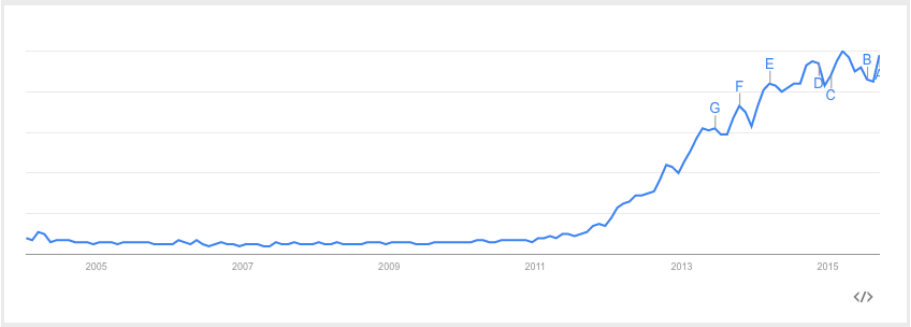
- What information is being analysed –by whom, where does it come from, to what effect, with what type of results?
- Are there examples of best and worst practices with respect to the use of large-scale data?
- What are the relevant and applicable legal regulations nowadays that can regulate the collection, analysis, processing and use of large-scale data? Are these suitable? Are they complied with?
- Does the decision-making process resulting from Big Data and the use of large-scale data jeopardise matters contemplated by laws on personal data protection and privacy? Does the regulatory framework contribute to its efficiency?

3. What is Big Data and Why is it Important

Before starting with the specific analysis of the VOS card ("you", in English), we must answer what Big Data is.

These concepts have been discussed for a long time in the technical and scientific sector of the academy (mathematics, IT, engineering, astrophysics, cosmology, meteorology) and in the fields of social sciences (economics and econometrics), but not with respect to the legal context of human rights: personal data protection and privacy. The current scenario leads us to analyse the consequences of the use of this technology, especially from a human rights perspective.

Big Data is a term that has caught on. Actually, if we search the term Big Data in Google Trends[1], we will find the result showed next.



Dan Ariely[2], Professor of Psychology and Behavioural Economics at Duke University and founding member of the Center for Advanced Hindsight, said about Big Data: "Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it".

[1] https://www.google.es/trends/explore#q=big%20data

[2] http://whatsthebigdata.com/2013/06/03/big-data-quotes/

There is no one definition ruling over the others to explain what Big Data is; yet, there is a certain consensus as regards the disruptive force presented by the great volumes of data and in the imperious need to capture, manage and process them to obtain greater benefits for organisations and companies.

> "Perceptions about Big Data's disruptive power are not confined to technology organisations; users see a new competitive weapon in the scene across industries and geographies, from businesses such as financial services and insurance, to practitioners such as postal services and governments…"[3]

Adrian Merv[4], vice-president of Gartner Consulting has defined Big Data by saying that "Big Data exceeds the reach of commonly used hardware environments and software tools to capture, manage and process it within a tolerable elapsed time for its user population". Therefore, Big Data is a convergence of technologies designed to capture and analyse great volumes of data which originate in sources of various natures at a high speed.

Judging from the definition transcribed, it becomes evident that Big Data sizes are constantly increasing and the same definition can be adapted, subject to change. The McKinsey Global Institute [5] also expands on Merv's definition by adding:

> "… we don't define big data in terms of being larger than a certain number of terabytes (thousands of gigabytes). We assume that, as technology advances over time, the size of datasets that qualify as big data will also increase. Also note that the definition can vary by sector, depending on what kinds of software tools are commonly available and what sizes of datasets are common in a particular industry. With those caveats, big data in many sectors today will range from a few dozen terabytes to multiple petabytes (thousands of terabytes)…"

IBM[6] points out that big data spans three dimensions: Variety, Velocity and Volume.

---

[3] Accenture. 2014 Report. "Big success with Big Data. Executive summary". Available on: https://www.accenture.com/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Local/es-la/PDF2/Accenture-Big-Data-POV-espanol.pdf

[4] Merv. Adrian., "Big Data", Teradata Magazine, 2011 Q1. Available on: http://www.nxtbook.com/nxtbooks/mspcomm/teradata_2011q1/index.php?startid=8#/42

[5] Manyika. James., Chui. Michael., Brown. Brad., Bughin. Jacques., Dobbs. Richard., Roxburgh. Charles., Hung Byers. Angela from The McKinsey Global Institute., "Big Data: The next frontier for innovation, competition, and productivity", May 2011. Available on: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[6] http://www-01.ibm.com/software/au/data/bigdata/

- Variety – Big data extends beyond structured data, including unstructured data of all varieties: text, audio, video, click streams, log files and more.
- Velocity – Often time-sensitive, big data must be used as it is streaming in to the enterprise in order to maximise its value to the business.
- Volume – Big data comes in one size: large. Enterprises are awash with data, easily amassing terabytes and even petabytes of information"

In conclusion, Big Data can be characterized as any kind of origin of data which has at least one of the following three characteristics:

- Extremely big volumes of information
- Extremely high velocity of information
- Very wide range or variety of information

To the model of the 3Vs, which includes the concepts of volume, velocity, and variety, we can add a fourth V, which is veracity (IBM), and a fifth one, which is value.

- Veracity – IBM explains that "One in three business leaders (directors) does not trust the information they use to make decisions."[7] The establishment of the veracity or reliability of Big Data presents a great challenge as the variety and sources of data increase.
- Value – Organisations are analysing how to obtain information from Big Data in a profitable and efficient way [8].

Big Data is not a technology per se but a combination of the last 50 years of technology evolution. As Ricardo Barranco Fragoso (IBM) explains[9]

> "…in general terms, we could refer to the trend in the progress of technology which has given way to a new approach of understanding and decision-making, which is used to describe big amounts of data (structured, unstructured, and semi-structured), which would take a lot of time

[7] IBM Global Services Route., Executive Report. IBM Institute for Business Value. "Revolution in the analytic skills and business optimization. When the new intelligence and business operations meet", November 2009. Available on: http://www-05.ibm.com/services/es/bcs/html/bao/bao-revolucion-en-las-capacidades-analiticas-y-optimizacion-de-negocio.pdf

[8] Ortoll. Eva., "Big Data is written with V", COMEN Magazine of Information and Communication Science Studies, October, 2014. Available on: http://comein.uoc.edu/divulgacio/comein/es/numero37/articles/Article-Eva-Ortoll.html

[9] Barranco Fragoso, Ricardo. What is Big Data? IBM Developer Works, 06/18/2012. Available on: https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/

> and would be too costly to load to a relational database for analysis. Therefore, the concept of Big Data applies to any piece of information which cannot be processed or analysed using traditional processes or tools…."

However, a minimum of data has not been established to define Big Data. Normally this term is used when we make reference to information Terabytes, Petabytes, Exabytes, and, currently, Zettabytes. Others say that Big Data must be used when the traditional system of database management is expensive and, therefore, Big Data techniques and/or technologies must be used.

Human beings generate and store information at every second and they are doing it in bigger quantities. Paraphrasing Ricardo Barranco Fragoso (IBM), in 2012, we could say that if all the data bits and bytes of that year were to be kept in CDs, we could build up a tower from the Earth up to the Moon and back. Just like Eduardo Galeano explained in "Open Veins of Latin America" in connection with the silver produced in Potosí (Bolivia): "Some excessively enthusiastic Bolivian writers insist that in three centuries Spain got enough metal from Potosi to make a silver bridge from the tip of the Cerro to the door of the royal palace across the ocean…"[10]. Now the new metal or gold is information and it has a BIG value. In the World Economic Forum held in June 2012 in Switzerland, the concept of Big Data was highlighted as a new economic asset comparable to gold, money, or oil. The European Commission, warned about the domination of data being the future of the European industry[11], in the words of Commissioner Kroes, has betted to its exploitation and development, facing the significant contribution it will offer to the European economy.[12]

Paul Siegele, president of the Energy Technology Company at Chevron said [13]: "Information technology is enabling us to get more barrels of each asset". Chevron's internal IT traffic alone exceeds 1.5 terabytes per day. Oil companies are using high speed communications, distributed sensors and data mining technologies for monitoring. The goal is to use data in real time in order to make wiser decisions and predict failures[14].

According to a study carried out by Cisco [15], between 2011 and 2016, the amount of

---

[10] Galeano. Eduardo., "Open veins of Latin America", page 32, Publisher Siglo XI de España Editores S.A., Spain, year 2008.

[11] The World Economic Forum., Big Data, Big Impact: New Possibilities for International Development, 2012. Available on:

http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf

[12] Kroes. Neelie, "Big data for Europe", European Commission. 7/11/2013. Available on: http://europa.eu/rapid/press-release_SPEECH-13-893_en.htm

[13] Leber. Jessica, "Big Oil Goes Data Mining", MIT Technology Review, 5/8/2012. Available on: http://www.technologyreview.com/news/427876/big-oil-goes-mining-for-big-data/

[14] Leber. Jessica., "Big Oil Goes Data Mining", MIT Technology Review, 5/8/2012. Available on: http://www.technologyreview.com/news/427876/big-oil-goes-mining-for-big-data/

[15] Cisco. Report: "Internet will be four times bigger in 2016", 05/30/2012. Available

traffic of mobile data will grow by an annual rate of 78%, and the number of mobile devices connected to the Internet will surpass the number of inhabitants of the planet. The United Nations project that the world population will reach 7,5 trillion inhabitants by 2016 which means there will be approximately 18,9 trillion devices connected to the net at a world scale. The aforementioned means that the global traffic of mobile data will reach 10.8 monthly Exabytes or 130 annual Exabytes. This volume of traffic forecasted for 2016 equals 33 trillion annual DVDs or 813 quintillion text messages.

People create mountains of data, but machines also do it. There is a communication called M2M, or machine-to-machine, which has generated the Internet of Things (IoT).

Big Data is not a linear process, but rather a complex interconnection system of components, circuits, platforms and feedback. Big Data includes a territory of predictive analysis and business intelligence that is worth exploring[16], known as Data Mining (DM).

Juan Miguel Marín Diazaraque (Head Professor at the Carlos III University of Madrid - Ph.D in Genetics from the Complutense University of Madrid (1997) and Ph.D in Statistics and I.O from the Complutense University of Madrid (1998)) has said that Data Mining refers to "a set of statistical methods providing information (correlations or patterns) when a lot of data are available (hence the name Data Mining) (…) Usually, Data Mining is the process of analysing data from different perspectives in order to summarise data in useful information segments. This information may be used to increase credits or benefits, lower costs, etc. DM allows users to analyse data from different aspects or angles, classifying them and summarising the relationships identified. With these techniques, sometimes it is even possible to show hidden relationships between events. A simple example would be finding out the relationship between diaper and beer purchases on a Saturday afternoon in supermarkets (…) Only a specialist who knows his clients will be able to interpret a raw correlation that will allow him to picture the typical case of a couple doing their shopping. Finding causal relationships leading to correlations such as the one described above may be quicker and simpler with Data Mining…"[17]

There are various areas where Data Mining is applied, with different objectives. We can include the following among them [18]:

---

on:http://www.cisco.com/web/ES/about/press/2012/2012-05-30-internet-sera-cuatro-veces-mas-grande-en-2016–informe-vini-de-cisco.html

[16] Diaz. Adam, Wampler. Dean, Smith. Tom, Langseth. Justin, Hausenblas. Michael, "DZone's Guide to Big Data, Business Intelligence, and Analytics", 2015 Edition. Available on: https://dzone.com/guides/big-data-business-intelligence-and-analytics-2015

[17] Marín Diazaraque. Juan Miguel., "Introduction to Data Mining". Available on: http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/introduccion-DM.pdf

[18] Marín Diazaraque. Juan Miguel., "Introduction to Data Mining". Available on: http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/introduccion-DM.pdf

- Commerce / Marketing: to identify consumer purchase patterns; to make associations among consumers and look for demographic characteristics; to predict responses to product mailing campaigns; to analyse consumer purchases, etc.
- Banking sector: to detect fraudulent use patterns of credit / debit cards; to determine expenses on credit / debit cards by groups; to predict the number of customers who are likely to switch Banks; to identify stock exchange rules based on historic data, etc.
- Transportation: in order to analyse freight patterns; to predict routes chosen by users, among others.
- Medicine: to associate symptoms and to conduct a differential pathology classification; to study risk factors for different pathologies; to make predictions in real time regarding medical facilities in order to optimise resources, visits, areas and rooms, among others.

In conclusion, the definitions of Big Data may vary according to the characteristics of the public using it. For some, the volume is more important; for others, it is velocity, and, finally, for another sector, there is variability.

a. Brief History of Big Data

The history of the term Big Data can be divided into two stages. The first stage (1984-2007) goes from the birth up to the expansion of the concept in the scientific and business field, its use being restricted to a first attempt of definition in the technical and academic jargon. The second stage starts in 2008, with the spread of the term from a technological and economic perspective. This benefited organisations and companies that started to investigate the technology and develop tools for the analysis of the big data to take better advantage of them.

The article which Wired [19] published in June 2008 was what triggered the explosion of Big Data. According to it: "Sensors everywhere. Infinite storage. Clouds of processors. Our ability to capture, warehouse, and understand massive amounts of data is changing science, medicine, business, and technology. As our collection of facts and figures grows, so will the opportunity to find answers to fundamental questions. Because in the era of big data, more isn't just more. More is different".

Likewise, in late 2008, the Big Data were adopted by a group of very prestigious researchers in the IT world, grouped around the Computing Community Consortium, a group which collaborates with the National Science Foundation (NSF) of the United

[19] Wired Magazine. "The Petabyte Age: Because More Isn't Just More — More Is Different", 06-23-2008. Available on: http://archive.wired.com/science/discoveries/magazine/16-07/pb_intro

States. This Consortium published an influential article (White Paper): "Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society"[20].

In 2008, IBM adopted the term Big Data for its marketing. In 2011, IBM introduced in Twitter #IBMBigData and in January 2012 it launched its first electronic book on Big Data technologies (Understanding Big Data)[21]. On July 29, 2012, Scott Adams' comic strip "Dilbert" included the term Big Data in its cartoons.

http://dilbert.com/search_results?page=7&terms=book+dea

b. Kinds of Data

Big Data is different data compared to that managed by traditional data sources which store structured data in relational databases. In order to understand the data sources, we must differentiate structured data (traditional data) from unstructured data (Big Data). However, the new Big Data management tools have given way to a new category within the kind of unstructured data: semi-structured data.

1. Structured Data

These are data which have a predefined format and content and we cannot get out of that format or fixed length. An example of the aforementioned is data from relational databases, spreadsheets, and files. Predefined formats make the work with those data

[20] Bryant. Randal E., Katz. Randy H., Lazowska. Edward D., "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society", Computing Community Consortium, Version 8: 12/22/2008. Available on: http://cra.org/ccc/wp-content/uploads/sites/2/2015/05/Big_Data.pdf

[21] Eaton. Chris, Deutsch. Tom, Deroos. Dirk, Lapis. George, Zikopoulos. Paul, "Understanding Big Data. Analytics for Enterprise Class Hadoop and Streaming Data, Mc Graw Hill, 2012. Available on: https://www.ibm.com/developerworks/vn/library/contest/dw-freebooks/Tim_Hieu_Big_Data/Understanding_BigData.PDF

easier. The usual formats are: date of birth (DD/MM/YY); passport (for example, 8 digits and one letter); department number (in general, 2 digits at a maximum), etcetera.

2. Semi-Structured Data

These data have a format which can be defined but its understanding is not easy for the user. This kind of information follows a kind of implicit structure but it is not regular enough to be managed and automatised as the structured information. Typical examples are the websites which follow certain common patterns and store content in HTML and metadata between labels.

3. Unstructured Data

The data which Big Data administers are the unstructured ones which cannot be stored in predefined relational data structures. The main characteristics of the unstructured data are the following: (a) the volume of data and the rate of growth of the unstructured data are quite bigger than that of structured data; (b) the origin of the data is very varied: data generated in forums, e-mails, social networks websites, data extracted from the web using semantic web techniques; (c) due to its lack of structure, we cannot use relational architecture, thus, it is necessary to work using "Big Data" tools.

Examples of unstructured data we can give are photograph, video, audio, and text data, printed documents, digital images, e-mail and text messages, messenger services (SMS, WhatsApp, Line, Viber, etc.).

Without a doubt, the data that analysts find most difficult to master are the unstructured data, but their continuous increase has brought about the birth of tools for their manipulation, as is the case with MapReduce or NoSQL databases[22].

c. Characteristics of Big Data

Big Data includes various technologies. The data entered in Big Data systems can come from several sources: social network websites, digital images, traffic flows sensors, sensors used to pick climate information, satellite images, MP3 music, bank operations, website content, administration scanned documents, biometric data (face and genetics recognition), telecommunications calls, health requests, RFID reading, GPS signals.

d. The Size of Big Data

As explained before, the megatrend of Big Data is not directly connected with the specific amount of data. A decade ago the warehouses of big companies stored from

[22] http://nosql-database.org/

1 to 10 terabytes and that was considered to be huge. Today 2 to 6-terabyte disk units can be bought in any shop for prices which range from 72.95 to 191.95 Euros[23] and many warehouses of some companies have crossed the barrier of the Petabyte.

For many experts, the controversial topic is how much Big (big volumes) involves, since the data subject is the fundamental support of such tendency. Maybe one answer adjusted to the current situation is that neither the Big part nor the Data part are the most important of Big Data. Frank pointed out that "Any of the parts is by a cat in hell's chance more important than the other one. What is important is what organisations make with big data; that's the most important thing. The analysis of the big data which your organisation makes, combined with the actions taken to improve your business is what really matters"[24].

All in all, the value in Big Data is both in Big and in Data and its final indicator depends on the analysis of data, how it will be made, and how the business will improve.

e. How have we reached the Big Data explosion?

As was already introduced, Big Data implies the convergence of tendencies which have been maturing over the last decade: social networks, mobility, applications, decrease in the cost of broadband, interconnection of objects through the Internet (M2M, machine-to-machine or Internet of Things) and cloud computing. All those tendencies have one common factor: they produce a great amount of data which needs to be captured, stored, processed, and analysed.

Companies, organisations, and governments work with thousands of digital sensors which give information of any kind to the Net: In industrial equipment, cars, planes, trains, ships, electrical appliances, in the streets, these sensors can measure and communicate the position or location, movements, vibration, temperature, humidity and even chemical changes in the air, CO_2 emissions… These situations have always existed and that has helped in decision-making processes to prevent natural disasters, detect ground motions. The difference resides in the fact that in the past the environments were controlled by structured data and now data come from everywhere and they are unstructured data.

The great volume of data comes from e-mails, videos, documents, text messages, SMS, RFID tags, photographs, digital images, networks of sensors and devices, search indexes, environmental conditions, social networks, medical explorations, governmental information, clicks history, music files, texts, online transactions, telephone incidences, together with everything which can be digitalized and transformed into data.

[23] http://www.amazon.es/Seagate-STBV3000200-Disco-externo-negro/dp/B0084LZJ1M

[24] Joyanes Agullar. Luis., "Big Data. Analysis of large volumes of data in organisations", 1st edition, Alfaomega Grupo Editor Argentino, Buenos Aires, 2013.

Many of these data need to be analysed in real time and others will be stored for years and only for specific queries. This big memory does not stop growing and it will be necessary to upgrade its intelligence.

The Network collects data of our profile (gender, age, tastes, habits, preferences, hobbies, profession). Those data are used to offer better results in searches and can help in decision-making processes or develop public policies which impact positively on society.

f. Main Uses and Applications

Big Data can be used for several activities but there are five main uses [25]:

1. Exploration of big data: To search, view, understand Big Data to improve the decision-making process, transactions, and to reduce risks. Likewise, the exploration of data contributes to reducing the risk of confidential information leakage thanks to its security mechanisms.

2. Improved user's 360º vision: Obtaining a total knowledge of the product or service user by incorporating internal data sources (e.g., user's behaviour in the company) and external ones (e.g., social networks, e-mails). This allows us to understand their behaviour and predict their future actions.

3. Extension of Security / Intelligence: Big Data allows us to decipher between big amounts of data (both internal and external), behaviour patterns, and prevent threats to safety. Likewise, it makes the discovery of frauds easier, through the monitoring in real time of the history of activity of an account, which makes it possible to reveal a user's abnormal behaviour or maybe a questionable transaction. The three main applications are:

    a. Improved vision of intelligence and vigilance: data analysis in real time and storage to discover patterns.
    b. Prediction and mitigation of cybernetic attacks in real time: by analysing the network traffic, companies can discover new threats and prevent attacks from hackers, intruders, cybernetic fraud, espionage, cyberterrorism.
    c. Prediction and prevention of crime: the ability to analyse data from the telecommunications network and social networks allows us to detect threats and act before criminals, as Philip K. Dick imagined in 1956, in his short story "The Minority Report", adapted to the cinema by Steven Spielberg[26].

[25] Lantares Solutions. "The five main applications of Big Data", Blog on Business Intelligence. Available on: http://www.lantares.com/blog/las-cinco-principales-aplicaciones-de-big-data

[26] http://www.imdb.com/title/tt0181689/?ref_=fn_al_tt_2

4. Operations Analysis: it allows obtaining, in real time, clarity of the operations, the customer's experience, their transactions, and behaviour. That is quite useful for those taking care of decisions and also increases the intelligence and efficiency of operations.

5. Increasing Data Warehouse: Integrating capacities of Big Data and Datawarehouse (DW) to increase its operational efficiency and value. Data Warehouse makes it possible to take advantage of several kinds of data to conquer new business opportunities in real time and improves the data storage structure, thus making the task easier and saving up costs.

Big Data is used in various sectors and several examples can be mentioned in each of them:

- Telecommunications: CDRs processing, social analysis, churn prediction, geo-mapping.
- Financial markets: climate's impact on the price of securities, analysis of market data in ultra low latencies.
- Transport: smart management of traffic.
- Energy: control of transaction networks, phasor monitoring unit.
- Health: newborn monitoring, systems of early detection of epidemics, remote medical care.
- Science: atomic investigation, syncrotron, detection of transitory events.
- Natural systems: control of fire, floods, and the climate.
- Defence and cybersecurity: surveillance in real time, situational consciousness, cybersecurity detection.
- Prevention of fraud in real time.
- Others: manufacture, text analysis.

g. State of the art in Argentina

In Argentina, Big Data is used in the sectors of telecommunications, agriculture, health, energy, weather, sports, entertainment, and transportation, among others.

Below, we shall mention some initiatives in each sector, based on the report drafted by Facundo Malvicino and Gabriel Yoguel (Interdisciplinary Centre of Studies in Science, Technology and Innovation - General Sarmiento National University): "Big Data: Recent Progress at an International Level and Perspectives for Local Development."[27]

[27] Malvicino. Facundo., Yoguel. Gabriel., "Big Data: Recent Progress at an International Level and Perspectives for Local Development.", Published by the Ministry of Science, Technology and Productive Innovation. Available as "Big Data" on http://www.mincyt.gob.ar/publicaciones

(1) *Telecommunications*. In 2014, Telefónica de Argentina purchased IBM's solution, PureData Systems for Analytics[28] allowing it to link traditional data with non-structured data and process information from the Internet and from social networks to understand customer preferences and habits. The convention, referred to as Big Bang Data and produced by Telefónica Foundation jointly with the Barcelona Contemporary Culture Centre, is taking place from July 2 until November 28, 2015[29].

(2) *Agriculture*. The Data Science Program of Sadosky Foundation, the Master in Exploitation of Data and Discovery of Knowledge of the University of Buenos Aires and the Argentine Association of Regional Consortiums for Agricultural Experimentation (AACREA, for its Spanish acronym) have been jointly developing activities. Multiple hacktones and talks about Big Data in Agrobusiness have been organised. In the framework of the Data Science Programme of the Sadosky Foundation the PALENQUE Project is developed, which is a Big Data platform to store, process and analyse agriculture data.[30]

Manufacturing techniques are changing the way of working in the agricultural sector, resulting in increased productivity and quality of harvests. These techniques require a technological development that is aligned with their use, including the use of GPS, various types of monitoring sensors, predictive and data analysis systems, and support systems for decision making, among other tools.

From the response given by the Ph.D. in molecular biotechnology Marcelo Soria during the interview that we will analyse below, we were able to detect that nowadays the big data retrieved, for instance, from satellite images, are used to diagnose crops development, assess humidity in the soil, and estimate performance. Producers can combine such data with the data that they generate. For a long time, most seed drills and harvesters have been equipped with GPS and computers on board computers in order to ration fertilizers and monitor crops performance. Likewise, currently, the first drones with sensors supplementing satellites are being used, thereby creating another flow of high-volume data.

(3) *Health* (medical providers, pharmaceutical industry, unions, prepaid medicine, public healthcare, etc.). Marcelo Soria has also noted that the use of digital medical records helps to create a single record to store all the patients' interactions with health services (e.g.: outpatient visits, prescribed drugs, examination ima-

[28] IBM., Press release. "Telefónica of Argentina implements IBM technology to know their clients better" Available on: http://www-03.ibm.com/press/ar/es/pressrelease/43447.wss

[29] http://espacio.fundaciontelefonica.com.ar/big-bang-data/

[30] "Palenque Project", Sadosky Foundation, 07/15/2014. Available on: http://triton.exp.dc.uba.ar/datamining/files/Charlas_y_Paneles/p4_fsadosky.pdf

ges, etc.). This makes consultations among professionals easier and significantly fosters retrospective studies.

(4) *Energy*. With regards to electric power, there is a real time consumption survey by the Wholesale Electric Market Management Company (CAMESA, for its Spanish acronym) which may be useful to determine behaviour patterns and to predict system issues[31].

(5) *Biotechnology*. Marcelo Soria has stated that more than a thousand dollars have been spent on the human genome. Production of gross data is on that threshold but if costs of the analyses are recorded, and depending on how such costs are considered, the barrier of a thousand dollars can be exceeded in a few months or years. According to specialists, this barrier determines the point from which the genome personalized sequencing will be considered an accessible study (bearing in mind that this study must not be frequently repeated).

(6) *Weather*. In the AGRANDA 2015 experience, various Big Data uses were shown in connection with the weather in order to obtain weather measurements and visualisations, predict hale using atmospheric indices, synthesise data processing of the INTA Anguil Weather RADAR, among others.

(7) *Sports*. Using Big Data in sports would improve decision making during play (football, basketball, tennis, hockey, rugby) based on game patterns or on the most optimum plays. In Brazil 2014, Germany[32] was one of the most innovative teams. Prior to and during the world cup, the German team was trained with special software of the SAP German company, which allows administering large data volumes in real time and, based on this, making play decisions. The information is gathered mainly by 2 systems: sensors used by players on their chest and legs during trainings, and passive video camera systems located around the field which record the whole match. This way, the German Football Association (DFB, for its German acronym) intends for trainers to get the best performances from their players, by optimising how they are managed. It is also useful for analysing the match once it has concluded and comparing it with prior matches.

(8) *Entertainment*. Social networks, mobile applications, games.

(9) *Transportation*. The case study is focused on the use of information potentially aimed at making changes in mobility (e.g.: improving bicycle lanes, determining origin and destination of journeys in order to learn how people move in the city).

---

[31] Malvicino. Facundo., Yoguel. Gabriel., "Big Data: Recent Progress at an International Level and Perspectives for Local Development.", Published by the Ministry of Science, Technology and Productive Innovation. Available as "Big Data" on http://www.mincyt.gob.ar/publicaciones

[32] Rua. Martina., "Big Data technology , the secret weapon of the German team in the World Cup" 07/13/2014.Available on: http://www.perfil.com/ciencia/La-tecnologia-Big-Data-el-arma-secreta-de-la-seleccion-alemana-en-el-Mundial-20140713-0043.html

4. Explanation of the Case Study – Methodology

https://twitter.com/gcba/status/593162702287147011

The case study has a specific geographical approach. The City of Buenos Aires has a permanent population of 2,890,151 people[33] and, every day, almost 1,200,000 people travel from the outskirts to work there [34].

The multi-functional card VOS enables inhabitants, regular visitors, and even tourists to make different kinds of tasks (Ecobici System, City Health Coverage, social services, access to underground in the future; and the storage of patients' medical records is on the agenda[35]) and it gives them the opportunity to buy products and/or services at a discount in several shops in the city. According to what the Ministry of Modernisation of the City of Buenos Aires informed since the commencement of the programme (November 2013), up to date, 314,008 cards have been handed out[36]. The card depends on the Network "En todo estás vos", which was started in 2012, as informed by the Ministry of Modernisation. However, we have been told that, "within the government's policies, through Resolution No. 733-MMOD-2013 the Programme 'En todo estás vos' was created"[37], which appears to be inconsistent, considering that the resolution was made after the commencement of the programme.

Resolution No. 733- MMOD-2013 states, in its relevant part, that

> "Law No 3304 established the Modernisation Plan of Public Administration of the Government of the City of Buenos Aires. Article 26 from the Law on Ministries No 4013 appoints the Ministry of Modernisation as the

[33] The latest data of the available population census (2010): http://www.censo2010.indec.gov.ar/

[34] http://www.24con.com/conurbano/nota/93755-cuantas-personas-delgba-trabajan-por-dia-en-capital/

[35] http://www.iprofesional.com/notas/189842-El-gobierno-porteo-lanz-una-tarjeta-multifuncin-para-usar-en-la-Ciudad

[36] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015. The information provided is until this date.

[37] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015.

Implementation Authority for the Modernisation Plan of Public Administration, authorised to design and implement policies on incorporation and improvement of the processes, technologies, information infrastructure and information systems, and management technologies for the Government of the City of Buenos Aires;

(…)

By virtue of the frame of the mentioned functions and for the purposes of implementing an Integral System that gives information to manage the access to Benefits, Promotions, Discounts and Services, it is necessary to create a Programme called 'En todo estás vos', centralised in a 'Smart Card'. Implementing such Programme would result in a flow of segmented information to smartly organise Benefits, Promotions, Discounts and Services that will be granted to the future users of the Programme, thus allowing them to use the services and enjoy the benefits easily, quickly and in a secure manner. In such context, shops that wish to adhere to the "En todo estás vos" Programme, in order to offer the Benefits, Promotions, Discounts and Services, shall comply with the requirements published and available on the website www.buenosaires.gob.ar/redentodoestasvos;

(…)

Therefore, by using the powers granted to it, THE MINISTRY OF MODERNISATION DECIDES TO: Article 1. - Create the Programme called 'En todo estás vos', which will offer benefits, promotions, discounts and services to users holding the smart card. Article 2.-Establish that all companies can request to be adhered to the Programme in accordance with the requirements that will be published and available on the website www.buenosaires.gob.ar/redentodoestasvos. Article 3.-The General Technical, Administrative and Legal Office of this Ministry will be responsible for the subscription of any relevant instrument. Article 4. - Record. Publish in the Official Bulletin of the Autonomous City of Buenos Aires. Communicate to all the Executive Ministries and Agencies, to all the government offices with the equivalent rank and level and to the General Office of Smart City Projects and the General Technical, Administrative and Legal Office of this Ministry. Completed, file it. Ibarra."

As indicated in the website http://www.buenosaires.gob.ar/tarjetavos "The VOS Card is a free programme of the Government of the City of Buenos Aires which allows you

to have access to services, benefits, and discounts in a wide range of shops throughout the country"[38].

Likewise, it "becomes a technological instrument for the acquisition of services (like Ecobici and, soon, the Subway System) and benefits in the City of Buenos Aires, which integrates networks for the fast and timely rendering of its services to citizens and visitors, and it also contributes to real time monitoring of the execution of public policies and the efficiency of public expense. Its use is personal and nontransferable; it is meant to become a symbol of identification with the City […] the VOS card is, for the time being, completely for free"[39].

In the terms and conditions[40], in section A.2, it is set forth that "for those cases where the card is the means provided by the City of Buenos Aires for the execution of social plans in force, Subscribers may also use the benefits of the card…".

The administration of the VOS card depends on the General Office of Smart City Projects - Smart City Subsecretary, Ministry of Modernisation. The agencies of the City of Buenos Aires involved in the gathering, analysis, and processing of information are the Office of Information Systems of the City of Buenos Aires, the Operative Management of Benefits and Services to the Citizen dependent on the General Office of Smart City Projects and the General Office of Smart City Projects itself, Smart City Subsecretary, Ministry of Modernisation. The Network "EN TODO ESTÁS VOS" is transversal to every Ministry of the Government of the City of Buenos Aires[41].

The VOS card is personal and cannot be transferred and can only be used by the person to whom it was given. The aforementioned is expressly established in section B.5 of the Terms and Conditions

When using the card (before requesting the invoice or upon entering the location), users must show their card together with a document certifying their identity (a credit or debit card, the national identity document –DNI, for its Spanish acronym–, I.D. card, driver's license, etc., as indicated on the website). This is why the card states on the front, in addition to the holder's name and surname, their DNI, to facilitate a better user identification. These indications are set forth in section D.1 of the Terms and Conditions, further adding that all of this will be done "before validating the card at the "POS"terminal of said location".

VOS cards are issued in the offices of the General Office of Smart City Projects, located at Bernardo Irigoyen 272, as well as in various Community Management and Participation Centres (CGP, for its Spanish acronym), in decentralised locations such

[38] http://www.buenosaires.gob.ar/tarjetavos

[39] http://www.buenosaires.gob.ar/tarjetavos/preguntas-frecuentes

[40] http://www.buenosaires.gob.ar/tarjetavos/terminos-y-condiciones

[41] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015.

as Hospitals, Tourism Centres, and in various fairs and events. The user decides how to use the card with respect to the various benefits it offers.

In the request for information filed with the Ministry of Modernisation and answered on the 12th of August of 2015, they were asked what information and/or personal data they collected and/or analysed and/or processed through the VOS card. In regards to this, they stated that anybody who wishes to obtain the card (whether residing in the Autonomous City of Buenos Aires, the Greater Buenos Aires, the rest of the country, or even tourists) may request it for no charge through the website http://www.buenosaires.gob.ar/tarjetavos/registro by entering the following information:

- *Mandatory:* Type of document, number, name, surname, date of birth, sex, profession, nationality, province, city, e-mail address, place to collect the card.
- *Non-mandatory:* Marital status, worker identification number (CUIL, for its Spanish acronym), street, number, postal code, landline telephone number, mobile number.

We were informed that the following information is mandatory due to the following reasons: "Date of birth and sex: referential information. Occupation, nationality: required to be able to analyse information. Province and city: required due to post mail. E-mail address: information via e-mail."[42]

Once all the information is uploaded and validated, the card may be collected at the service centre of choice.

In this work, we found some inconsistencies in the Terms and Conditions. Section H.5 states that "Data of the Subscriber and data referring to transactions resulting in the use of the Cards (hereinafter, the"Information"), shall be included in a database registered with the National Registry of Personal Databases as 'VOS card' (the"Database"). The Information shall be available for use to VOS and the Participating Locations authorised by the programme, for publicity, promotional and commercial purposes, or for any other use within the legal framework in force. Data may be used to determine Subscriber interests and/or affinities so that the Benefits are tailored based on their interests, and to achieve the best performance of the Programme" (emphasis added).

The first feature of note is that databases are said to be registered with the National Registry of Databases. This caught our attention provided that, based on the jurisdiction and type of database (public base), the same should be registered with the Centre for the Protection of Personal Data – Ombudsman of the CABA (Autonomous City

[42] Reference: reply to the information access request submitted to the Centre for the Protection of Personal Data of the City Ombudsman's Office on 09/18/2015.

of Buenos Aires), the local control authority wherein this type of databases are registered. For this reason, we wanted to confirm in which registry the databases of the VOS card were kept.

Precisely, regardless of what is stated in the terms and conditions, the National Department for the Protection of Personal Data informed us that they had no jurisdiction for the registration of these databases and that we had to resort to the local registration authority. Thus, based on the access request to the Centre for the Protection of Personal Data[43] of the CABA Ombudsman, it was confirmed that the database of the programme "En todo estás vos" ("Everywhere there is you") is registered with the Database Registry of said office. The registration was conducted by Lic. Gustavo Eduardo Gazzaneo – Executive Director of the Information Systems Agency (ASI, for its Spanish acronym). The Ministry of Modernisation[44] was consistent when informing that the databases were registered in said Centre under the name of the Information Systems Agency of the Government of the Autonomous City of Buenos Aires (GCABA, for its Spanish acronym), domiciled at Independencia 653 of this City. Pursuant to the second article of the Transitory Provisions of Law 1845, and considering the information provided by the Ministry of Modernisation, the Information Systems Agency has allegedly obtained the corresponding authorisation.

With regard to recipients and categories of recipients of said files, registries, databases and/or data banks, we were informed that "Every access to the database is given after provifing a username and password authorised to that effect. Users may be found in the Information Systems Agency of the Autonomous City of Buenos Aires and in the General Office of Smart City Projects, which depends on the Ministry of Modernisation. Categories include administrator users and others defined by the General Office mentioned above"[45].

What is curious for us is that the VOS card information is available under the terms of section H.5, which means that the VOS card is based on a dynamic of exploitation of information, profiling, and recommendation systems. That is what we explained when analysing Data Mining.

Section H.6. clearly states that "the Subscriber expressly accepts and consents to the following: (a) Providing the Information requested to subscribe to the Programme, and authorising 'VOS card' to access, maintain and process the Information kept therein; (b) that each Participating Location discloses to 'VOS card' and/or to its agents or dependants Information referring to transactions made by the Subscriber for the pur-

[43] Reference: reply to the information access request submitted to the Centre for the Protection of Personal Data of the City Ombudsman's Office on 09/09/2015.

[44] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015.

[45] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015.

poses stated above;(c) that the 'VOS card' processes and/or transfers the Information on its Database to its agents and/or to the Participating Locations for the purposes stated above; (d) that the 'VOS card' processes and/or transfers the Information to the Participating Locations, so that they may send the Subscriber different information, offers and benefits they may be interested in; (e) that faced with claims made by Subscribers, or by any Administrative or Judicial authorities, the 'VOS card' uses the Information on its Database…".

In accordance with the above, the privacy policy of the terms and conditions states the following: "the information included in the databases may be used to: Process, validate and verify the card activation of 'VOS card'; authorise participation in offers, awards or online contests; offer advertisers general information about Subscribers and usage trends; develop new products and services to meet Subscribers' needs; contact Subscribers, by phone or e-mail, to survey opinions about the service, at the direction of VOS, and to provide information about VOS products and services. Subscribers may at any time notify their refusal to receive information about products and services…"

This means that VOS card accesses, maintains and processes the information, and the locations where subscribers use their benefits show VOS card information related to transactions made by Subscribers. It is made clear that VOS card processes or may transfer information included on its database to its agents and/or to the Participating Locations for publicity, promotional and commercial purposes or for any other purpose, as stated in the above paragraph„ some of which include: to determine the interests and/or affinities of subscribers so that Benefits are tailored to their interests and to achieve the best performance of the Programme. What also draws our attention is the amount of information being surveyed by the City Government, through the relationships between Government - user, user - service or product provider, and Government - service or product provider.

As reported, the data collected is added to a database designed by the Information Systems Agency (ASIN) of the GCABA, wherein the PAN code (Primary Account Number) of the assigned card and details about the authorised features of the card are added (e.g.: use of the Ecobici system).[46]

The PAN code is used in cases in which it is necessary to univocally identify the card, as informed by the Ministry of Modernisation.[47]

We've been informed that "Databases are protected in accordance with the Policy published in Resolution 177/ASINF/2013 of the Autonomous City of Buenos Aires."[48].

[46] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015.

[47] Reference: reply to the information access request submitted to the Ministry of Modernisation on 09/18/2015.

[48] Reference: reply to the information access request submitted to the Ministry of Modernisation on 09/18/2015.

Such resolution is broad as regards security standards and can be seen on the following link: http://www.buenosaires.gob.ar/asi/estandares.

Moreover, the Ministry of Modernisation has indicated that: "Information from the VOS Card system is stored in the Data Centre of the Government of the City of Buenos Aires, in state-of-the-art equipment and security controls (firewall, IPS, antiDDoS, etc.) Such Data Centre is located a Independencia 635, where the information systems for the Government are domiciled, as well as the card, disks, servers, security schemes, backup, amongst others."[49]

Furthermore, the Ministry of Modernisation has reported that the data are used to verify cardholders' identity (e.g., badge for the City Health Coverage in Hospitals Programme), and, in turn, in case the user agrees to receive regular e-mails with details about benefits and discounts associated to the card.[50]

As reported by the Ministry of Modernisation[51], the information is collected automatically from the website of the City Government wherein the card is requested by the Information Systems Agency of the GCABA. The Operations Management of Citizen Benefits and Services, dependant on the General Office of Smart City Projects, undertakes information analysis to act upon, for instance, replacement requests due to deterioration, loss or theft, or upon enquiries about the use of discounts and/or benefits.

We were also advised that information is collected in order to enable the printing of cards, the use of discounts and/or benefits associated to the Programme, and to allow identification for the use of services such as the Ecobici system or the City Health Coverage.

Following the request for information filed with the Ministry of Modernisation, on the 18th of September we were informed that: "The following technologies are used for each state of the card information. For collecting: web forms. For storage: databases. For analysis: spreadsheet forms and dashboards. For elimination: logical delete. For tracking and auditing module: Big Data." That means that the Government of the Autonomous City of Buenos Aires may now be using Big Data solutions for tracking and auditing.

However, in a reply to a second Access request, they informed that Big Data or Data Mining technologies are not used.

---

[49] Reference: reply to the information access request submitted to the Ministry of Modernisation on 09/18/2015.

[50] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015.

[51] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015.

The Ministry of Modernisation[52] pointed out that the information is stored in the servers of ASIN of the GCBA, and that the PostgreSQL8.3 platform is used to process said information. PostgreSQL8.3 is an open source relational database management system released under a BSD license.

Furthermore, as informed by the Ministry of Modernisation in the request for information replied on the 18th of September, 2015: "The card contains RFDI technology. This technology (antenna) permits reading/recording information related to the specific activity that is being worked on (…) it does not contain geolocation technology (…) The card stores information both in the RFID structure and in the magnetic strip."

With regard to the EcoBici system, it is provided that "the City of Buenos Aires offers a Public Bicycle Transportation System, now with automated stations, so you can enjoy EcoBici for free, 24 hours a day, every day of the year. Go to the designated spot at the station and place your VOS card on the reader. The System will assign you a bicycle, stating the dock number where the bicycle is located on the screen next to the card reader…"[53]

Users must register and find a bicycle at the nearest station to use the service. A form with personal data must be completed, and proof of identity and of address must be uploaded. Multiple alternatives are offered to that end, including: a web form, the mobile application BA EcoBici (for iOS and Android), and in person. An identification number and proof of address must be provided at the communal centres.

The use of VOS card to access the underground is also provided, although it has not been implemented yet. The "Annex - Terms and conditions for use of the VOS card to access the underground" provides that card Subscribers may use the same to access undergrounds in the City of Buenos Aires, pursuant to the Terms and Conditions and those set forth in the future by Subterráneos de Buenos Aires S.E. (SBASE), in their capacity as managing authority of the underground.

Being a Subscriber to the Card programme pursuant to the issuer's Terms and Conditions is an essential requirement to access VOS card Benefits. The Card must have credit on it in order to access the underground. Credit may be added to the Card at the Ticket offices and/or Self-service Terminals that SBASE enables to this end and/or at such places as set forth by SBASE.

It is important to point out that "[t]he subscriber agrees to the use by SBASE of Subscriber data and data referring to transactions resulting from the use of the Cards pursuant to the legal framework in force. The Information shall also be used by SBASE, or by the underground licensee, or by whoever is appointed by SBASE to manage the information, in order to adequately render passenger transportation services. The subscriber

[52] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015.

[53] http://www.buenosaires.gob.ar/ecobici/sistema-ecobici

knows and agrees to the following: (1) that as a consequence of the service agreement, and in furtherance thereof, SBASE and/or whoever it expressly appoints to this end may request certain information which may be considered Personal Data pursuant to the provisions of Law 25,326 (Protection of Personal Data); (2) SBASE may contract a company to render services of storage, systematisation, modification, evolution, blockage, and, in general, processing of Personal Data (…) The acceptance of the Terms and Conditions implies an express authorisation by the subscriber to SBASE to disclose data related to them or to their use of the service. SBASE may disclose subscriber data, or data related to their use of the service, whether under the law and/or by virtue of a requirement set forth by competent authorities and, in accordance with internationally recognised rules."[54]

a. Research background

The activities undertaken to conduct the work of analysis include, among others, the following:

- The main source of information was the Government of the Autonomous City of Buenos Aires. We had access to the technical and statutory structure on which the card was based by virtue of information access requests addressed to the Network "En todo estás vos" Ministry of Modernisation (the latter unified all responses) and to the Information Systems Agency. Furthermore, information access requests were made to the Centre for the Protection of Personal Data - Ombudsman (local control authority and registry of public databases in CABA) and to the National Department for the Protection of Personal Data.
- It is essential to stress that we conducted an extensive bibliographic survey on the issue, using academic documents and reports available online related to this particular technology, not only globally, but also regionally and nationally.
- Among the material survey activities, we participated in a lecture by Dr Walter Sosa Escudero (PhD in Economics, specialised in Statistics and Econometrics) on "BIG DATA: challenges and perspectives for social sciences"[55] on the 16th of June of 2015 at the School of Economic Sciences, University of Buenos Aires.
- We participated in the "Argentine Workshop on Data Science"[56] on the 19th of August of 2015, at the School of Exact and Natural Sciences (University of Buenos Aires, UBA for its Spanish acronym).

[54] http://www.buenosaires.gob.ar/tarjetavos/terminos-y-condiciones

[55] http://home.econ.uba.ar/economicas/?q=content/big-data-desaf%C3%ADos-y-perspectivas-para-las-ciencias-sociales

[56] http://www.fundacionsadosky.org.ar/workshop-argentino-en-ciencia-de-datos/

Facundo Malvicino (in charge of the project "Big Data para el desarrollo [Big Data for development]" – Interdisciplinary Centre of Studies in Science, Technology and Innovation (CIETCI, for its Spanish acronym) – Ministry of Science, Technology and Productive Innovation) participated in said Workshop with his presentation "Big Data. Recent progress at an international level and perspectives for local development."

The event was later divided into three main themes. The first was "Sector Público & Datos (Public Sector & Data)," where Natalia Sampietro (Data Operational Manager) and Cristian Reynaga (Operational Manager of Urban Sensing) presented "Open data and the Internet of things. An opportunity to implement public policies".

The second main theme was "Software & Business Intelligence," where the cases of companies BeSmart (Diego de Arriandiaga); Globant (Juan J. López Murphy and Tomás E. Tecce) and BluePatagon (Eduardo Azzola) were presented.

Finally, the third main theme was "Internet & Marketing Online" where the cases of Despegar.com (Daniel Altman); Mercado Libre (Gerardo Loureiro and Nicolás Aliberti); SocialMetrix (Gustavo Arjones) and GranData (Carlos Sarraute and Esteban Donato) were presented; and it finished with a networking session.

- Furthermore, we interviewed people in the academic area, in the business world, and officers of the Government of the Autonomous City of Buenos Aires.
- Participating in AGRANDA 2015 (Argentine Convention on Big Data - first edition) was very fruitful. It took place at the School of Exact Sciences, Engineering and Land Surveying - National University of Rosario (Argentina) on the 1st and 2nd of September of 2015, where we were able to meet scholars and technological experts on this topic, businessmen and people from governmental areas related to Big Data approaches and applications, as well as to learn and gather more useful information for our research. This conference gave us the opportunity to exchange our research and preliminary findings with expert scholars and businesspeople in order to enrich our investigation. Agranda 2015 is part of the 44th Argentine Conferences on Informatics (JAIIO 2015).
- The Workshop conducted on the 23rd of September of 2015 was extremely important, with the participation of 16 people, at the NH Buenos Aires 09 de Julio, with academic representatives, technology experts and people working for the protection of personal data and privacy in attendance. Based on this, we were able to draw our conclusions about the technology implemented in the case study, its advantages, disadvantages and challenges.
- It should be noted that these surveys, analysis and conclusions were undertaken with the aid of a specialised technical consultant (Gastón Darío Nápoli) and a junior assistant (Leandro Ucciferi).

b. Information access requests

Various information access requests have been made in order to clarify certain technical and legal aspects of the VOS card.

1. National Department for the Protection of Personal Data[57]:

The first request was addressed to the National Department for the Protection of Personal Data. As mentioned above, we confirmed with the National Department for the Protection of Personal Data that the VOS card database was not registered with said office (in spite of the terms and conditions to the contrary).

The National Department for the Protection of Personal Data replied as follows:

"In this regard, this National Department hereby advises that it is not within its statutory competence to control entities which depend on provincial and municipal governments (…)".

In this particular case, the VOS CARD is a free programme of the Government of the City of Buenos Aires that allows access to services, benefits and discounts in shops all across the country.

The Government of the City of Buenos Aires has enacted Law No. 1845, for the Protection of Personal Data, with the purpose of controlling the use of personal data related to physical persons or legal entities, recorded or destined to be recorded in databases of the public sector of the City of Buenos Aires (article 1), said databases shall be registered with the Registry of Personal Data created within the Ombudsman of the City of Buenos Aires, which is the control authority of the aforementioned local law (articles 22 and 23).

Consequently, the request issued to this National Department should be redirected to the Government of the City of Buenos Aires, specifically to the Ombudsman of the City of Buenos Aires, wherein the Centre for the Protection of Personal Data operates…"

2. Centre for the Protection of Personal Data - Ombudsman (authority for the control and registry of public databases of the CABA).[58]

At the same time we sent the information access request above, we forwarded the same questions to the Centre for the Protection of Personal Data – Ombudsman, on the 16th of July of 2015. Said information access request was answered in a very

[57] Reference: reply to the information access request submitted to the National Department for the Protection of Personal Data on 07/21/2015.

[58] Reference: reply to the information access request submitted to the Centre for the Protection of Personal Data of the City Ombudsman's Office on 09/09/2015.

untimely manner. The response should have been sent by the 30th of July of 2015, possibly extended until the 13th of August of 2015, but it was received on 09/09/2015.

The first question in the request was the following: "State if there are files, records, databases and/or data banks entered in your registry related to the VOS card within the network "En todo estás vos"made by the Government of the Autonomous City of Buenos Aires and/or any other entity, Ministry and/or Division thereof. If so, please provide the access link to said information and/or the entry number of the same".

The Centre for the Protection of Personal Data has stated that the database of the programme "En todo estás vos" is registered with said Database registry pursuant to the provisions of article 23, Law No. 1845. The registration was conducted by Lic. Gustavo Eduardo Gazzaneo – Executive Director of the Information Systems Agency (ASI) - Decree 37/GCBA/2012, who, as stated, met the requirements defined by said Department. Furthermore, the database in question is entered under the Single Identification Key No. 194.

It remains unclear whether the VOS card (this is being analysed) is also included under said registration. And we also requested a link to said registration. We only obtained the single key. None of the data provided by the Centre could be confirmed since the search website for public database registries has been under maintenance for an extended period of time[59].

The second question was "What information is collected in said files, registries, databases and/or data banks?"

In relation to this, they stated that "according to the information provided by the Executive Director of the ASI, through the registration form, we can report that the database does not contain sensitive data (Art. 3 of Law No. 1845), nor does it contain data related to health nor to criminal or misdemeanour records. It only contains identification data…" However, it is not made clear what kind of identification data is contained in said database.

One of the questions (question 8) enquired the following: "Is the Government of the Autonomous City of Buenos Aires and/or any other entity, Ministry and/or Division thereof registered as a service provider in connection with the VOS card within the network "En todo estás vos¿"

The Centre for the Protection of Personal Data has noted that "according to the database manager, the base does not contract service providers for data processing."

Question 8 was designed to find out whether the Government of the Autonomous City of Buenos Aires and/or any other entity, Ministry and/or Division thereof was registered as a service provider in relation to the VOS card in the network "En todo estás vos," and not whether the database contracted service providers.

[59] http://www.cpdp.gob.ar/index.php?option=com_content&view=article&id=47&Itemid=58

Questions 10 through 13 were as follows:

> 10: "Have any warnings, recommendations, reminders and/or suggestions been made to the Government of the Autonomous City of Buenos Aires and/or to any other entity, Ministry and/or Division thereof with regard to the VOS card of the network 'En todo estás vos' ?"
>
> 11: "Have any judicial and/or disciplinary proceedings been suggested or started against the Government of the Autonomous City of Buenos Aires and/or any other entity, Ministry and/or Division thereof with regard to the VOS card of the network 'En todo estás vos' ?"
>
> 12: "Have data holders been represented to enforce their rights to access, amend, suppress, and update files, registries, databases and/or data banks of the Government of the Autonomous City of Buenos Aires and/or any other entity, Ministry and/or Division thereof with regard to the VOS card of the network 'En todo estás vos' ?"
>
> 13: "Have data holders received any assistance in proceedings undertaken resulting from the use of information of the VOS card in the network 'En todo estás vos' by the Government of the Autonomous City of Buenos Aires and/or any other entity, Ministry and/or Division thereof?"

The Centre for the Protection of Personal Data has reported that: "In relation to questions 10 through 13, this Department hereby states that all answers are negative. There has been no judicial or administrative proceeding, neither at the request of a party nor on its own motion, related to the Vos Card of the network "En Todo Estás Vos."

3. Ministry of Modernisation

It should be noted that the same information access request has been sent to the Ministry of Modernisation and to the Network "En todo estás vos", the request being handled by the Ministry of Modernisation, which is the entity that set the card in motion. Most of the responses were mentioned in chapter "4. Explanation of the case study – Methodology."

As already mentioned, the network "En todo estás vos" started in 2012, although it was later disclosed, in the answer to the question regarding what the legal framework was, the policies adopted, and the grounds related to the goals of the network, that the network was created by Resolution No. 733-MMOD-2013. This shows that the Resolution creating the Network "En todo estás vos" came one year after the reported

start date. Delivery of Cards started in November of 2013, and up until the response date of the access request (08/12/2015), a total of 314,008 cards were handed out.

The administration of VOS cards depends on the General Office of Smart City Projects, Smart City Subsecretary, Ministry of Modernisation. The Network "En Todo Estás Vos" is cross-sectional to all the Ministries of the Government of the City of Buenos Aires.

Anybody who wishes to obtain the VOS card (whether residing in CABA, the Greater Buenos Aires, the rest of the country, or even tourists) must enter the following data on the website http://www.buenosaires.gob.ar/tarjetavos/registro: Type of document, number, name, surname, date of birth, sex, profession, nationality, province, city, e-mail address, location to collect card.

This response raises the following concerns: Why is so much data required? What is the purpose of requesting, for instance, information about profession? A census, with targeted advertising purposes?. We will go back to this in our conclusions.

The following question was included in the access request: Is the personal data and/or information collected subjected to automated processing and incorporated to automated databases of personal nature? Is there an algorithmic use for data collection and use of data at a large scale?

The answer was: "Data collected is incorporated in a database designed and secured by the Information Systems Agency of the Government of the City of Buenos Aires, wherein the PAN code of the assigned card and details about the authorised features of the card are added (e.g.: use of the Ecobici system once the user has completely filed the documents required in this specific case by the Transport Department to loan bicycles to citizens for no charge). The data are used to verify cardholders' identity (e.g., badge for the Hospital City Health Coverage Programme), and, in turn, in case the user agrees to receive regular e-mails with details about benefits and discounts associated with the card."

The abovementioned is stated in the following terms in the terms and conditions - privacy policy: "Personal Data shall be subject to automated processing and incorporated in the corresponding automated databases of personal nature for which VOS shall be the holder and responsible party (hereinafter, the"Databases"). To this end, VOS shall provide Subscribers with adequate technical resources so that they may, in advance, have access to this notice about Data Protection Policies, or to any other relevant information, and so that they can give their consent for VOS to proceed with the automated processing of their Personal Data. Except indication to the contrary, answers to questions about Personal Data are voluntary, the absence of reply does not entail an impairment in the quality or quantity of the applicable services, unless otherwise specified…"

It draws our attention that this privacy policy within the terms and conditions is not accessible by the subscriber at the time of registration. The policy may be reached by

entering the words "terms and conditions vos card" in the search engine, but not on the website of the GCABA. This leads us to question the effectiveness of a privacy policy that is not known by the subscriber. We shall reconsider this matter in our conclusions.

The PAN code mentioned above is used at the financial level. This interbanking account number is printed and/or embossed on plastic payment cards. Section B4 of the Terms and Conditions indicates that cards "have a magnetic band enabling them to operate in point of sale terminals ("POS") located at the participating locations of the 'VOS card' programme and it will allow Suscribers to register the use of Benefits, in any sale, contracting or use made in the locations or events participating in the Programme."The relationship between the card issuer and the store is governed by the terms and conditions accepted by the user who requests to be included in the programme. The point of sale terminal POS is Prisma Medios de Pago S.A[60], as informed by the Ministry of Modernisation in response to the information access request.

Regardless of the above, section B.8 in the Terms and Conditions states the following: "It is expressly clarified that the Cards are neither credit nor purchase or debit cards, they are not useful as authorised payment methods unless so decided by 'VOS card' and they are the exclusive property of 'VOS card'…" We clearly found an inconsistency between the terms and conditions and the actual use of the card.

Even more so considering that section C.1 states that "[t]he Subscriber may access the Benefits of the 'VOS card' through the purchase of products or services determined by VOS, in the locations participating in this Programme (…) which shall be announced on a case by case basis through the Website and through any other means that VOS deems fit, and they shall be subject to availability and validity in the manner and as prescribed by the Programme." Although, section C.3 states that "under no event nor circumstance can the Benefits of the 'VOS card' be traded for money in cash or for any other goods".

As indicated in the reply provided by the Ministry of Modernisation in the information access request, the GCABA obtains no financial benefits. Businesses do obtain a bigger flow of customers and transactions based on the discounts they offer.

With respect to the question: To what end and with what type of results is the information collected and/or analysed and/or processed? The Ministry of Modernisation replied the following: "The information is collected in order to enable the printing of VOS cards, the use of discounts and/or benefits associated with the Programme, and to allow identification for the use of services such as the Ecobici system or the City Health Coverage."

In regards to who is in charge of information collection and/or analysis and/or processing, the Ministry of Modernisation replied the following: "The information is collected automatically from the website of the City Government wherein the card is requested

[60] http://www.prismamediosdepago.com/

by the Information Systems Agency of the GCABA."

When asked where the information collected is stored, the Ministry of Modernisation replied "in the servers of the Information Systems Agency of the GCABA."

With respect to the conditions of organisation, operation and procedure applicable to said files, registries, datasets and/or data banks, the Ministry has been reluctant in their reply. It has informed that these are used when a new card is requested in order to avoid card duplication and to disable the former plastic in theft or loss cases as well as to replace the previous ones. The new cards are delivered in person and handed out upon proof of identity given by the applicant.

Concerning the safety measures of the databases, the Ministry of Modernisation was evasive when replying that: "at an application level, users who access the databases are identified with username and password. In addition, an auditing log is being developed which registers changes made in databases (former value – current value) with timestamp and username. At an operation infrastructure level, the databases are located ASIN's data center."[61].

In the privacy policy of the terms and conditions it is provided that "VOS has adopted the protection security levels of the Personal Data as required by law, and has installed all technical means and measures available to it in order to avoid loss, misuse, alteration, unauthorized access and theft of the Personal Data disclosed to VOS."

Furthermore, and regarding whether the collection of a greater type of information is expected in later stages, they answered no for the time being.

Based on the answers given by the Ministry of Modernisation, a new information access request was submitted to this Ministry and to the Information Systems Agency, dated August 25, 2015. The request was answered on the 09/18/2015 by the Ministry of Modernisation.

Questions and answers

Question 3: What is the purpose of the development of an audit log to record changes made to databases (former value - current value) with timestamp and username?

> "The purpose of an audit log is to record any change made on the VOS card system, registering the user and date of the change/ event, any necessary information for revision/management tasks, etc."

Question 5: What type of information do you collect or do you plan on collecting in regards to the use of the VOS card?

[61] Reference: reply to the information access request submitted to the Ministry of Modernisation on 08/12/2015.

> "Any information presented in the locations where the benefit is implemented or obtained will be taken into account, provided the shop of the network of benefits or points of contact meets the conditions to process the card in the available 'POF'."

Question 8: Is the VOS card user registration the only source of information managed and analysed by you?

> "Any information that was taken into account when the user registered is regarded as user's information. User's data are only registered if they had been entered into the registration."

Question 9: Is there any additional mandatory information required by another division? If so, what is the content and purpose of said information?

> "The card supplies other services for which specific information needs to be added. If the card is used as a COPS (City Health Coverage) plan card, data from the healthcare centre where the person is assisted are added as a reference."

Question 10: What benefits and/or services are provided by the VOS card to users?

> "Apart from allowing them to use the Ecobici system and subscribing the City Programmes, users can get benefits by showing the card when purchasing at shops that they can check at http://www.buenosaires.gob.ar/tarjetavos."

Question 11: What is the use and how does the VOS card work in relation to the Ecobici system? What type of information is collected through the use of the VOS card with this service and what for? Does this information include georeferencing data?

> VOS Card allows users to take bikes from the system automatically. They can ask about how it works on the same website that was previously mentioned because the Ecobici System is thoroughly explained there or directly on the Ecobici website: http://www.buenosaires.gob.ar/ecobici."

Question 12: What is the use and how does the VOS card work in relation to the City Health Coverage Programme? What are the benefits users can access using the VOS card in relation to said programme?

> "With regard to the City Health Coverage Programme, VOS card works as an identity card for the user and the benefits of the programme can be checked at: http://www.buenosaires.gob.ar/areas/salud/cops.php."

Question 13: What is the use and how does the VOS card work in relation to social benefits services and access to public transportation? What are the social benefits that users may access using the VOS card? What public transportation systems may be accessed? Is the fee always the same or does it vary depending on how frequently the user uses his card?

> "As regards social benefits, at the moment VOS card is related to the City Health Coverage Programme, as stated on the previous question. Right now, regarding access to public transportation for the public in general, VOS card only allows users to access the Ecobici system, which is referred to in the answer to question 11."

Question 16: Is Data Mining, A/B testing, machine learning, genetic algorithms and/or Big Data, Cassandra, Cloud Computing, Hadoop, ETL, HBase, SQL, Mapreduce or any other technology used to process the current volume of information generated by the use of the VOS card? Please specify.

> "It is not used".

Question 17: In case the previous answer is negative, would the Ministry of Modernisation and/or the Information Systems Agency consider that the current or future volume of information generated in the programme of the VOS card requires the current or potential use of Big Data technologies?

> "This is the standard of the Government of the City for these > technologies, it is the platform used with the suite of the SAS > products."

Question 19: Is the analysis of information conducted in real time or with collection purposes? What is the purpose of said collection and analysis? Who undertakes said analysis?

It's worth mentioning that the Ministry of Modernisation did not answer the first part of the question that asked if the analysis of information was conducted in real time or with collection purposes. However, it replied the following:

> "Any information requested to the citizen in order to obtain the card, apart from differentiating the citizen when printing the cards, allows identification for the use of bikes from the Ecobici system, thereby protecting them from wrongful removals."

Question 20: Does this person ensure that the information collected is encrypted and/or transported encrypted and/or anonymously? In what way?.

"SSH certificates are used."

Question 21: What are the safety measures and standards and the best practices policies of the Software Engineering implemented throughout the entire lifecycle (collection, storage, processing, analysis, removal) of information to prevent it from being breached, both in the case of internal and external incidents?

> "VOS card system works in accordance with the standards and processes established by ASI (http://www.buenosaires.gob.ar/asi/estandares)."

Question 22: Do you conduct any penetration tests? Please specify which ones.

> "Yes, penetration tests are conducted as well as security assessments."

Question 24: Is the VOS card aligned with national and international standards in matters of personal data protection and computer safety? Please specify which ones.

> "Personal databases of the Vos card are encrypted in the Centre for Protection of Personal Data, under the control of Ombudsman Office of the City of Buenos Aires and such databases have not been submitted to obtain other national and/or international standards."

Question 25: In regards to shops participating in the programme, what is the exchange of information existing among them, and what is its purpose? Does this information include georeferencing data? If it is not currently in use, would the Ministry of Modernisation and/or the Information Systems Agency consider using Big Data for this type of analysis?

> "No information should be exchanged with the shops adhered to the Programme and no Big Data analysis should be conducted considering the purpose of the Programme is to offer benefits and discounts instead of collecting/using information for another end."

Question 26: Do you conduct or do you plan on conducting an analysis of use patterns of the VOS card and/or its associated services? If so, what for?. Do you use or would you use Big Data solutions for this?

> "Patterns of the use of VOS card are not analysed and no information has been obtained so far in order to do that."

Question 27: What safeguard copy policies have you implemented? How would they be affected should any Big Data technologies be applied?

The Ministry of Modernisation was a bit evasive when replying:

> "Safeguard policy is within the general Policy of information safeguard for the Data Centre of the Information Systems Agency." No information of the safeguard policy was attached nor was it informed where information on such policy could be found.

c. Interviews

Interviews were aimed at personalities in the academic and business world. The questions and a summary of the answers are transcribed below.

1. *Do you consider that the current or future volume of information generated by the VOS card programme, or any other variable resulting from said programme, will merit the use of Big Data or Data Mining techniques and/or technologies, if they are not in use today? For you, what would be the direct and indirect benefits of using them?*

Gustavo Arjones (Degree in Information Science, MBA in Services and Marketing and Co-founder and CTO of Socialmetrix) answered:

> "Considering current amounts of cards (350,000) and considering that each citizen makes 6 transactions per day, we would have ~2.1MM/a day, this could be classified as a problem where Big Data (BD) techniques are necessary. However, the use of BD tools brings about an additional complexity that reduces the velocity of development and delays getting benefits from the use of data. In this context, I recommend starting with a more traditional model and then moving to BD tools according to specific needs and cases of known use."

Ernesto Mislej (Co-founder and Director of the Lab of Data Science of 7Puentes and Professor of the Master in Exploitation of Data and Discovery of the Knowledge of the School of Exact and Natural Sciences of the University of Buenos Aires) was certain about his answer:

> "Yes, without a doubt." For Mislej, the direct and indirect benefits would be: "To better know the behaviour of users to be able to offer better and new services. Not only limited to the ones mentioned but also including

new ones. To do better what is being done and use the data and insights to offer better labour services."

Djamel Abdelkader Zighed (d'Excellence Scientifique award, director of the Human Science Institute in Lyon- France, vice-president of the International Association of "Mining and Knowledge Management EGC") has said:

> "In terms of technology, I believe that using great volumes of data and data mining are necessarily required to collect, manage and use such data in a convenient way. The reasons are obvious given that this multi-function card will be used several times a day by all users and this will result in dozens of data for each transaction (geolocation, type of transaction, transaction terminal, user, etc.) Data will be very heterogeneous and they will grow exponentially with the launch of the card, its use and the proliferation of users and points of sale."

Juan Martín Pampliega (Data Engineer at Jampp and Assistant Professor at "Big Data Certification" at Buenos Aires Institute of Technology- ITBA) affirmed:

> "… In the case of medical records, several times, data are unstructured and maybe because of this it is considered a problem for"Big Data". Big Data tools provide tools that make the processing of this type of information easier than with traditional bases, such as PostgreSQL. In order to assess this, it is convenient to make a difference between the operational use of a database (operating the system on a daily basis, recording every detail) and the use for data analysis. The second use is relevant for Data Mining technology and Big Data tools. Both are certainly useful to deduce new useful information on the use of the card and people using it."

Marcelo Soria (Ph. D. molecular biotechnologist, Professor of the Master in Exploitation of Data and Discovery of Knowledge in the School of Exact and Natural Sciences of the University of Buenos Aires and professor in the School of Agronomy of the University of Buenos Aires) indicated:

> "The volume of data gathered by the programme of the VOS card will clearly be very big in the future. This, in addition to the diversity of information gathered, creates the necessary environment for techniques of analysis of Big Data and data mining to be used, which we can include within the discipline of the data science. The application of methodologies of the data science will help to discover use patterns of the services offered by the card. That way, the offer of services may be adjusted, user groups

> may be characterised according to use patterns, and a base of knowledge may be created to offer new services and discover potential frauds. Starting from a certain volume of data, which varies with the applications, it is very hard to extract useful information by only making inquiries on the basis of relational databases or with simple viewing and analysis methods."

Adriana Echeverría (Co-director of the Master in Exploitation of Data and Discovery of Knowledge in the School of Exact and Natural Sciences of the University of Buenos Aires) pointed out:

> "I think that whoever has the possibility of using those data may discover interesting information for 'their business.' The benefits will depend on how the information is used, on the reason why new and valuable information will be discovered."

The Mg. Cecilia Ruz (head of the Adviser Commission of the School of Exact and Natural Sciences and professor of the Master in Exploitation of Data and Discovery of the Knowledge of the School of Exact and Natural Sciences of the University of Buenos Aires) answered that:

> " 'Big Data' is commonly defined through 3 characteristics: Variety: this problem already counts with a moderate variety of sources of information. It has access to bikes, to social services, to the underground, etcetera. With the incorporation of the medical record, the kind of data treated is going to increase. Velocity: at least judging from what I know of the problem, it does not seem that this aspect is going to be significant. Velocity is commonly associated with the information coming from sensors or from interaction in social networks. Volume: it depends on the diffusion, on the amount of cards of this kind which are handed out and on their use. Considering the amount of inhabitants in the City of Buenos Aires, it is likely that it is only a matter of time until the use of this card requires the use of Big Data techniques. If the volume grows enough without Big Data-specific techniques, processing of these data will be impossible. Without a doubt, facing problems of this kind in the context of the State is a source of opportunities for the local development of this kind of technologies."

She also added that:

> "Some of the uses that, to my view, are more important are: Fraud detection: through the analysis of the behaviour, both when a card has been stolen and when one is being used to attempt to commit some kind of

> fraud can be detected. Crossed promotions; through a technique which is called association rules it can be detected that people who have a preference for the kind of service A may also have a preference for the service B; then, if there is a person who 'consumes' A and not B, B can be offered to them. It is the kind of technology which is used in the newspapers when it is said 'those who read this article also liked this other one.' Detection of communities, through the cards which consume together with a certain frequency. Detection of patients at risk, potential candidates who may suffer a certain kind of accident can be detected by using the history of patients who had same kind of serious health incident and such information can also be used to take preventive measures."

Mg. Eduardo Poggi (professor of the Master in Exploitation of Data and Discovery of Knowledge in the School of Exact and Natural Sciences of the University of Buenos Aires) answered the following:

> "I am not familiar with the volume that it manages so I cannot say. Anyhow, the concept of Big Data is more associated to flow than to volume. There is a variety of versions as to what Big Data is or is not, but if it were only a volume issue, it would fall within the common practices of Data Mining. The essential feature of Big Data is velocity, both of data arrival and collection as of generation of validated hypothesis based on the former to make decisions. I do not know the scope of the transactions generated by the use of the VOS card to give an opinion on its use. On the surface, it looks more like a traditional Data Mining issue than a Big Data one, I don't know what"immediateüse may be applied in this case".

Carlos Diuk, Data Scientist at Facebook Inc., pointed out that

> "If all transactions are actually going to be stored (e.g., all underground rides), it will be necessary to mount a data warehousing structure on top of the PostgreSQL8.3 relational base. The volume of data and a non-productive infrastructure to access it would allow for all kinds of data analysis to provide insights to shape public policies (from transportation to health)." It's worth mentioning that all the answers given by Carlos Diuk reflect the interviewed professional's opinions and that Facebook, as a company, has no connection with his answers.

Ana Silvia Haedo, PhD in Statistics and director of the Master in Exploitation of Data and Discovery of Knowledge in the School of Exact and Natural Sciences of the University of Buenos Aires), suggested "yes, considering it allows to pull interesting information without preconceptions."

Dan Rozenfarb, Head of Research and Development at Keepcon company, explained to us that "there is no general consensus on what Big Data is. This is why it is always being defined in every presentation. Otherwise, there would be a consensual definition. Big Data is a trending term. It is not a concept in itself. It is more associated to techniques, technologies, but it is not a concept or a discovery, and it depends on how much information is gathered. Big Data is a large volume. In the case of the VOS card, Data Mining would be more applicable. An exploitation of information, profiling, is being undertaken, which translates into recommendation systems."

Esteban Feuerstein, PhD in Informatics and one of the managers of the Data Science Programme at the Sadosky Foundation, explained to us that:

> "Since I am a resident of this City, and I have not heard of the card until now, I imagine that the current volume of data generated in the programme is not exceedingly large. It will surely grow in the future. In any event, I don't think it is worth looking for a precise answer to this question. It is not even clear what techniques may be individually classified as Big Data, they are more a group of techniques constantly evolving and enabling a change of paradigm. Since it consists of a set of data to be analysed, it will surely be beneficial to use storage, processing and visualisation techniques that will allow for their efficient use."

Carlos Sarraute, R&D Director at GranData, told us that "with the delivery of 300,000 cards and the data collected through those cards, there might be talk of Big Data and automated processing of information."

2. *To what end and with what type of results do you think that the information is collected and/or analysed and/or processed by the Government of the City of Buenos Aires?*

Marcelo Soria pointed out:

> "I do not know the specific uses that the Government of the City of Buenos Aires makes, but some of the possible uses are the ones I mentioned in the previous question. It must be kept in mind that even when information is anonymised for its analysis, it is possible to enrich the data obtained from the database of the VOS card with geographical data, such as the approximate location from where people log into the system, crossing the demographic information of the users (age, gender, neighbourhood where they reside) with grouped census data or with data of surroundings to hospitals, schools, means of transport (like underground stations), etce-

tera. This means that the information gathered by a system can be combined with the one gathered by another system and in that way a bigger and more detailed basis can be obtained."

Juan Martín Pampliega answered that:

> "Collecting all this information allows the Government of the City to have a better and more detailed view on how those services are used. Analysing such information helps find opportunities to improve the use and supply of services."

Cecilia Ruz said that:

> "Since they are just getting started, I figure exploitation is very limited. New uses will appear as they go further."

Eduardo Poggi answered that:

> "I would imagine it is done to adjust the characteristics of the service based on the use made by users. I suppose it must have a secondary use, which is to know the behaviour of users, especially if the VOS data are compared with other data of people/businesses involved."

Ana Silvia Haedo stated it might be to "obtain information that will allow to improve health, mobility and garbage collection services, etc."

Eduardo Feuerstein answered:

> "I do not know what the purpose is, although I can imagine different potential uses for this type of information, particularly to detect behavioural patterns which could lead to better public policies in each area or at a higher level."

Carlos Sarraute answered:

> "When you use the card, there is a purchase record. This could be useful for predicting future purchases of users. With the use of bicycle lanes we can determine the origin and destination of journeys, and thus make estimates on how people move in the city".

Gustavo Arjones answered: "Understanding the movement of people and which products are consumed to optimise city services…"

Djamel Abdelkader Zighed has pointed out:

> "I think that using a data card of this kind may be very interesting for the different uses, marketing, security, traffic (Smart City), etc. However, guards are necessary to reassure users about privacy and use. This card has to include terms and conditions related to the appropriate use of data (…) I believe the display has to aim mainly at well-defined services, where the use of this card allows users to save time and money."

3. *Do you believe that the manner in which the local government collects and/or analyses and/or processes data might potentially affect human rights? If so, specify which.*

Ernesto Mislej informed:

> "I don't know. Managing data about personal behaviour is an issue to be dealt with, not only in the case of the Government of the City of Buenos Aires. Unambiguous identification and eventual unethical use of data may take place whenever data about personal behaviour are available. Providing data anonymisation and generalisation tools is essential."

Marcelo Soria replied:

> "Availability to personalized data of third parties can affect people's privacy. In a deeper level, collecting and analysing data makes it possible to"develop a picture.ºf those using the VOS Card system. But, again, I don't know about data security protocols. An analyst may be allowed to work with anonymised data and, in this way, the possibility of getting individualised data can be reduced to the maximum. Therefore, establishing protocols to protect information and authorization systems to access to data and tools is important. Ensuring compliance with those protocols and regular audits is vital, too."

Juan Martín Pampliega pointed out:

> "We can't determine how secure the access to data is or if processing such data could have an impact on human rights."

Adriana Echeverría said:

> "It depends on what is done by those involved in this issue. They may do it or not. They should take into account the rules on data management that are already established, for instance, in the European Union."

Cecilia Ruz indicated:

> "In this new era in which everything is connected, governments have a great flow of data about what citizens do and this may be used for good or evil. Data can be used either to create a kind of "big brother.°r to improve people's lives. An example could be the one of health prevention I mentioned in the previous question. A detailed tracking of compliance with vaccination plans could be implemented, too. Opportunities are almost endless."

Eduardo Poggi was adamant in his reply:

> "Of course, positively and negatively. Big Data breaks boundaries, therefore one must be very careful with the unexpected impact it may have and act accordingly. The State as user and as controlling entity must provide for the protection of citizens/users at the same time it must use technology to provide better services."

Esteban Feuerstein stated that

> "potentially, yes, it may have an impact, like most of these projects. Mainly, in regards to privacy. Unexpectedly in most cases, the accumulation of large amounts of data from various sources, although allegedly anonymised, in combined manner allows for the disclosure of private information. Based on this, the use given to said information is, potentially, a risk for the involved individual."

Gustavo Arjones answered:

> "This type of information is quite sensitive because it can be used to identify people and groups, the management of data security is the key to avoid risks and misuse."

Ana Silvia Haedo stated that: "Surely, the access and disclosure of information and data as well as the results of their analysis."

4. *How would you rate, on a scale of Excellent - Very Good – Good – Regular – Bad – DK/NA, the processing that has been/is being applied by the Government of the City of Buenos Aires to this massive data? Could you justify your answer?*

Answers reflected ignorance on this issue.

However, Cecilia Ruz pointed out:

> "I'd say with Positive Perspectives, due to the interaction I've had, through the master's degree at the University of Buenos Aires, with people from the Ministry of Modernisation. They are clearly devoting a lot of effort to improving this area."

Marcelo Soria answered:

> "I don't know about the particular case of the VOS Card, but the Government of the City of Buenos Aires has a very good team of professionals in charge of data analysis."

Finally, Ernesto Mislej said:

> "I don't know as regards massive data. When it comes to open data, the Government of the City of Buenos Aires has really made progress towards the availability of public data, actions to increase and strengthen the community of players, etc."

5. *The VOS card is currently supported by a PostgreSQL 8.3 platform, a relational open source database management system released under a BSD license. Would you consider this technological framework as the most suitable for processing massive information?*

Marcelo Soria answered:

> "It's a valid option. PostgreSQL is a scalable tool that will adapt to the increase in the volume of data. If at some point, it becomes insufficient, the tool can be combined in an ecosystem with other NoSQL systems to allow managing other data massively. Given the nature of data and the

need to do complex operations on such data, using a relational database system is necessary. Such database makes it possible to control and ensure the correct functioning of all transactions related to data (adding, deleting and updating several tables simultaneously.) So far, traditional transactional systems cannot be replaced by systems of NoSQL high-volume data, but they supplement each other. Since it is an open-source software system, once vulnerabilities are detected, they are usually corrected quickly and these events are made public. Besides, it is a very popular system, so there is a great deal of professionals who can make it work efficiently and in a safe manner."

Cecilia Ruz indicated that:

> "Actually, massive data analysis is not performed with traditional data storage systems (in this case PostgreSQL), so this information has little relevance when carrying out the analysis. The relevant aspect is that if information is stored in a database, we can assume that it is relatively organized, which is an advantage when facing projects of massive analysis".

Eduardo Poggi replied that

> "PostgresSQL is a relational database of generic use. This type of repositories was intended for transactional use, not for massive queries such as those required by Big Data. From this point of view, it would not be the most adequate tool. In any event, it depends largely on the work volume; if used as a prototyping tool, it can be very useful and low-cost."

Gustavo Arjones indicated:

> "I think it is an interesting platform to support the current volume, probably in the future they will have to change it but starting with a relational model is appropriate and then evolve according to specific needs."

Carlos Diuk answered that

> "A relational database of this type would be able to support a productive system, but not the long-term storage of data and its availability for analysis. If the goal is to apply Big Data techniques to analyse this data, I would recommend installing a NonSQL data warehouse (Hive or similar)."

Juan Martín Pampliega said:

> "… the PostgreSQL version used is quite old and, besides, this type of database is not appropriate if stored data are expected to be massive (Data Terabytes) and if data are to be analysed including great time periods…"

Esteban Feuerstein stated

> "I am not a supporter of strict classifications, which would be the case if I were to say that PostgreSQL is a Big Data tool or not. It is a traditional database, and if that is enough for what is intended to be done with the data, it is all right. I do not think that moving to a more modern database, the so-called Non-SQL, would automatically result in benefits."

6. *What are the safety measures and standards and the best practices policies of the Software Engineering that should be implemented throughout the entire lifecycle (collection, storage, processing, analysis, removal) of information to prevent it from being breached, both in the case of internal and external incidents?*

Ernesto Mislej replied:

> "There are established quality and security standards of information for each stage of the development, including information encryption, security access, data anonymisation of identifiable information, directing the insights towards statistical and not individual purposes, etc."

Marcelo Soria said:

> "There are standard rules for these situations based on which a set of ad-hoc rules can be designed for each situation. These information protection policies are included in a relatively new discipline called"Data Governance". Such discipline includes different but supplementary aspects to those dealt with by information security. Information security focuses on protection against intrusions (hacking), strength of information systems, including physical security, among other related aspects."

Cecilia Ruz mentioned that:

> "There is an international association DAMA (www.dama.org) that has published a book (DAMA Guide to the Data Management Body of Knowledge) that contains a recompilation of the best practices and principles for data management in all aspects and includes specific chapters on data security and operating data. The book is a very good reference to start defining a life cycle for"data". In my opinion, a very important part, and frequently forgotten, is the one related to"meta data", that is to say, the meaning of data and how they are represented. For instance, recognizing if an amount is in pesos or dollars, when such amount started being stored and if it went through any change since it was stored."

Carlos Diuk answered that

> "If I understand correctly, and the intention is to include health data, there should be very strict safety standards. Ensuring the privacy of this information is key, and, to be honest, I am slightly concerned about the ability of the CABA Government to ensure it."

Juan Martín Pampliega pointed out:

> "There are different standards that could be followed as a guide to ensure data security, where possible. PCI DSS (Payment Card Industry Data Security Standard) could be taken as an example."

Carlos Sarraute told us that

> "we are faced with a delicate process. The information should travel encrypted, and penetration tests such as, for instance, simulating a hacker attack, would have to be conducted for sensitive data."

7. *According to your knowledge and experience, is the VOS card aligned with national and international standards in matters of personal data protection and computer safety?*

Interviewees were not able to answer this question since they were unfamiliar with the legal context. Marcelo Soria said: "I don't know if policies on data governance were implemented for the VOS Card."

8. *What standards and policies would apply to large scale data processing? What do you think would be the adequate legal framework for the use of Data Mining and/or Big Data tools in the Argentine Republic?*

Marcelo Soria replied:

> "Actually, data protection and privacy policies are the same, whatever the volume of data."Big Datarefers to a series of technologies and methodologies designed to deal with large volume of data. This task with existing tools had become difficult or directly impossible, but the final goals and the kind of questions have not changed much compared to what happened some years ago. The difference is that integrating types and sources of very different data in only one study is easier nowadays. I couldn't say which the legal framework is in particular for data protection and privacy in Argentina. In practice, either because this legal framework is not appropriate or because its compliance is not guaranteed, data protection and privacy seem to be poor. If bigger excesses don't take place, it is because those who try to use data for malicious purposes lack the required software, hardware or expertise."

Adriana Echeverría pointed out: "The international standards and legal rules that are being used in the European Union, for instance."

Cecilia Ruz replied:

> "As regards the more technical aspects I'd take as a reference the book I mentioned before (DAMA Guide to the Data Management Body of Knowledge). With respect to legal aspects, I'm not an expert at privacy issues but we should take into account the Personal Data Protection Law (Law 25,326) and the efforts of the UN and European Community towards this matter."

Ana Silvia Haedo told us they should be "the same than for any data processing and analysis, regardless of the database size."

9. *Determine some potential consequences resulting from the use of Data Mining techniques and/or Big Data technologies for decision making on the use of large scale data for the VOS card.*

Ernesto Mislej distinguished two big consequences:

"Licensing public information. Making information available only for a group, the one that pays for it or the one that thinks in a certain way, etc. Not protecting unambiguous, identifiable personal information. I don't mind having third parties using public information to do business, as long as that information is available for many third parties to do business."

Marcelo Soria affirmed:

"Apart from ethical and legal considerations, using methodologies of data science in the analysis of high-volume data and complexity gives the possibility of discovering patterns, its variations and anomalies that are useful to know the behaviour of systems' users with great detail, predict their future behaviours and design interventions or policies highly focused on particular groups."

Cecilia Ruz replied:

"Once behaviour patterns and fraudulent behaviour are detected, the allocation of resources can be improved. New possibilities will be identified while the system is being used. Given my professional experience, I know that as you"surf"data, you find more things that catch your attention and that results in new action lines."

10. *State some advantages, disadvantages and challenges from a social and technological point of view in the implementation of the VOS card.*

With regard to this question, there was near unanimous agreement about the possibility of a negative impact if data privacy is not safeguarded.

Cecilia Ruz said:

"The biggest technological challenge is not collecting data but knowing how to use them. From a social point of view, the challenge is bigger, especially if you try to make it a massive product. This results in some barriers that must be carefully analysed. One of them is people's age, because"going digital"is more difficult after certain age. Another barrier is that of the so-called "digital illiterate", young people who have had little exposure to technology and, therefore, are reluctant to use it. These groups will find it more difficult to find out about the benefits because they are usually advertised through electronic media."

Gabriel Baum (Bachelor of Science in Informatics, professor at the National University of La Plata and investigator at the Research and Training in Advanced Computing Laboratory) posited

> "it would be very beneficial if correctly used. People would have more capabilities, which would result in an empowering of society, but the downside are the risks. In this regard, he said that"having the medical record is very convenient if I faint on the street and they can treat me…"but always maintaining the protection of personal data and privacy.

However, Eduardo Poggi pointed out some positive advantages such as, for example: "Anticipating needs. Managing logistics more efficiently. Implementing alternative control methods."

Carlos Diuk said that

> "it should also be defined who has access and to what end. For example, to what extent can law enforcement (Metropolitan Police, Federal Police), Intelligence or judicial agencies access data about the movement of people in public transportation? There has to be a clear legal framework in that respect. There can be a positive impact if the information is used to inform public policies.

Carlos Sarraute said the challenge would be to "build a solid infrastructure." The advantages would be evidenced by understanding the city, which would be beneficial feedback for society in general. Including business transactions together with medical record information would constitute a disadvantage. He affirmed that "it does not make a lot of sense putting everything into a single card."

11. *Would you use Big Data in your projects? If yes, for what type of projects/solutions and what would be the direct/indirect benefits?*

Marcelo Soria was very passionate about this question:

> "I'm a professor at the master's degree in data mining and discovery at the University of Buenos Aires. Therefore, because of my teaching activity, I'm directly involved in teaching data science methodologies. In the master's degree, we see that young professionals are interested in being trained on these disciplines. Most of our students come from the private sector. It's also worth mentioning that interest in this studies programme is not just local. We have a percentage of students coming from other Latin American countries, especially Colombia and Ecuador."

Apart from that, my research tasks are focused on data science applications in biological issues, with particular focus on applications in agriculture and biocomputing. In some specific areas we are living a data mining situation and an exponential growth of its production, for instance, in agriculture. Nowadays, using a large amount of data retrieved from satellite images to diagnose crops development, assess humidity in the soil and estimate performance is common. Besides, producers can combine such data with the data that they generate. Over the past few years, seed drills and harvesters have been equipped with GPS and computers on board in order to ration the use of fertilizers and monitor crops performance within a few meters. Currently, the first drones with sensors supplementing satellites are being used, thereby creating another flow of high-volume data.

A similar revolution and maybe with even larger amount of data is being seen in medicine, with respect to the use of digital medical records and the development of new lab technologies and biocomputing applications. The use of digital medical records allows to create a single record to store all the patients' interactions with health services, from visits to external doctors' office to prescribed medicines and examination images (X-ray, MRI, etc.) This makes consultations among professionals easier and significantly fosters retrospective studies.

Furthermore, more than a thousand dollars have been spent on the human genome. Production of gross data is on that threshold but if costs of the analyses are recorded, and depending on how such costs are considered, the barrier of a thousand dollars can be exceeded in a few months or years. According to specialists, this barrier determines the point from which the genome personalized sequencing will be considered an accessible study (bearing in mind that this study must not be frequently repeated.)

These last points, on biomedical data collection and processing are vital in projects such as VOS Card. Medical information is highly sensitive because if it is spread without authorization, people can suffer negative consequences and such information must be considered private. Which kind of data related to the use of health services in the City will be gathered? Which policies and agency will be in charge of its custody, spreading and eventual destruction?

Cecilia Ruz energetically replied:

> "Of course! I work in international trade and there is a world of possibilities. For instance: integrating all the information related to a product, from logistics movements since it leaves the exporting plant in the country of origin until it reaches its destination, including descriptions, opinions and quotations of the product on the web and other background information of the seller. This kind of integration would lead to better purchases and tracking. The real time analyses of warehouse movements together with the backup 'IT operations' make it possible to identify inconsistencies."

Gustavo Arjones pointed out:

> "Yes. We actively use Apache Spart processing frameworks and NoSQL (Cassandra, MongoDB) Databases. They are used as bases for our analytic social-networking products. The benefit is having a scalable platform that can support clients and data growth. This also allows having the results available in a time window that is appropriate because it takes a few hours to process all the reports and not days as was the case with the previous solution.

Juan Martín Pampliega said:

> "I have worked on projects that use Big Data tools to process large volumes of data since 2011. I worked on data analysis projects to optimize stock management, social networking sites data analysis to assess the impact of business marketing campaigns and data analysis to optimize programme advertising campaigns on the Internet and mobile devices. I also teach Big Data tools in Big Data Certification at Buenos Aires Institute of Technology."

Eduardo Poggi answered positively: "Of course."

Carlos Diuk pointed out that he uses Big Data at work every day.

Ana Silvia Haedo stated that: "In fact, it is used in the courses and works by Master students for all types of analysis, and applied to multiple fields."

Feuerstein answered:

> "Yes, of course. Except for the use or non-use of Big Data for individual tools or applications, it is clear that the paradigm implies a different view which may bring about benefits in absolutely every field of human knowledge and activity. Examples abound in different sciences or productive areas."

12. *What would be the legal and technical points to be considered by the public sector for the implementation of projects where Big Data is used?*

Marcelo Soria pointed out:

> "In principle, legal requirements and technology practices for the protection of data are those that protect any data in a digital format. The difference is that in the case of more complex systems that are likely spread in several locations that may be governed by different rules, more regulations, controls and audits to deal with these cases are necessary. For example, is executing data analyses programmes on cloud service providers, such as Amazon, Google or Microsoft acceptable? These analyses need the transfer of data to a private third party whose servers are physically located abroad. Another example would be the situation of a city administration that needs AFIP [Federal Administration of Public Revenue] data to supplement its data analyses on touristic development in its jurisdiction. What would be the process for relating the data? Who authorizes it? What kind of data can be shared if any data can be shared at all without violating fiscal confidentiality?

Adriana Echeverría replied:

> "Hiring multidisciplinary teams managed by attested qualified professionals, not by relatives or friends that lack appropriate professional training."

Juan Martín Pampliega indicated:

> "Firstly, I think teams that will develop these projects should be trained on the variety of available tools and in which case they are used. Secondly, during the implementation phase, many of the precautions established in most of the rules or international standards should be taken into account. These precautions refer to data anonymisation, access control, audits and data security."

Cecilia Ruz answered:

> "From the technical point of view, we have to take into account that we're working with brand-new technology, which brings about some aspects related to the lack of maturity of the tools. Regarding the legal point of view, privacy is the key issue, that is to say, how to combine data analysis with data protection."

Gabriel Baum said that "We need to have a State that is capable of respecting civil rights and of building a reliable political system."

Carlos Diuk stressed that

> "In addition to the promise of a log system for database modifications, it has to be clear which the monitoring system is for those logs and what protocol to follow in case of alarm. The system should not only have a log of database modifications, but also one of queries. For instance: every time someone with access to the system makes a query about a specific user, the reasons for said specificity should be monitored. There should also be a legal framework defining the rights of people over data. For instance, can a card user request that all his information related to health be deleted from the system? If health data are involved, a suitable role model might be US HIPAA (http://www.hhs.gov/ocr/privacy/)."

Ana Silvia Haedo said that "national and international regulations about privacy and access to information should be followed."

13. *Do you think that the State would improve its operation through the use of Big Data? If so, in what cases and/or areas of interest? Do you know any examples of Big Data being used by the State? Please specify.*

Marcelo Soria said:

> "The use of data science methodologies, whether Big Data technologies are used or not, allows us to learn more about the behaviour of the people who are being analysed, develop predicting models with more resolution and, thus, design new policies and interventions and assess their impact. These policies and interventions may be highly focused on particular sectors of the population."

Adriana Echeverría mentioned that: "It depends on how and why this technology is used. There are examples in very different real, natural, social and government environments".

Cecilia Ruz said:

> "Of course, there are worldwide examples for these improvements in almost all of these areas of the Estate. The uses are really different, ranging from transit control in a city to the analysis of terrorist groups behaviour.

> In Argentina, I know basically PAMI [Comprehensive Medical Attention Programme] and the Government of the City of Buenos Aires have initiatives in regards to these areas."

Gabriel Baum answered that "improvements to public policies could be made with anonymised information through the inference of profiles, trends."

Eduardo Poggi answered: "Of course, and it is being done in the City: weather sensors, sensors in garbage collection trucks, traffic monitoring, 114 monitoring, etc."

Carlos Diuk answered

> "Yes. For example, with detailed data on transportation, the public transportation system could be improved. We could know where people are travelling from when they need to go to a public hospital, understand where there are needs, etc."

Juan Martín Pampliega pointed out:

> "Countries like the United States are really advanced as regards the use of tools to manage large amounts of data and there are even open-source software tools that fall within the category of Big Data tools that have been developed by government agencies, such as Apache Nifi, which was created by the NSA. The United States was also the first government that appointed a CDO (Chief Data Officer), a person in charge of processing digital information, especially information related to Big Data. These tools are mainly used for analysing transit data and optimising services network, development and research within the fields of genetics, fiscal fraud detection and others."

Ana Silvia Haedo answered: "Like every tool, it could be used well or misused; if it is correctly used, it naturally contributes to improving the management of information and resources; I know of the Pami case."

Esteban Feuerstein affirmed:

> "Yes. The Palenque platform that we are developing at the Sadosky Foundation is an example. There are applications or case studies already under development in different public offices of the scientific technological system (INTA, CONAE, to name a few) or of public services (SUBE card, for example)."

14. *Do you think that a generalised use of Big Data by the State would influence its adoption by the local private market?*

Gabriel Baum answered that

> "The Argentine State does not have technologies. There is a Big Data Centre in Benavídez for use, but there is no political will or trained people. In the United States, for instance, it is a defence issue."

Adriana Echeverría answered:

> "It depends on the level of influence, on the need of the local market to suffer that influence, on the visible results in the private market, the behaviour, changes, improvements achieved in the State."

Cecilia Ruz pointed out:

> "Yes, although in the case of Argentina, the private sector has a clear advantage. A good example of this delay is that there are no employees of the public sector among students in the master's degree."

Following this path, Marcelo Soria replied:

> "I think the trend is the other way round. The use of data science methodology generally starts in the private sector and then moves to the public sector."

Eduardo Poggi said: "Yes, but I believe that the market is slightly ahead of the State."

Gustavo Arjones affirmed: "Yes, together with the spread of successful cases and more experts who know the technology and its problems."

Ana Silvia Haedo said: "In fact, data mining is already being used in the private sector: Despegar, Mercado libre, Hospital Italiano, etc."

Esteban Feuerstein answered

> "The State can always take the lead in the adoption of some technologies, and, in particular, it can help to boost the growth of technology researchers and providers that would otherwise find it difficult to achieve"critical mass."

Finally, Juan Martín Pampliega affirmed:

> "The implementation of projects using tools for the government to operate Big Data and considering such projects important can help spread the importance of processing information within any kind of organisation. Also, it is important to spread the fact that nowadays there are a lot of tools available for free and open-source tools. Such tools make it possible to process large data amounts successfully to find useful information for decision-making, process optimization and cost reduction."

d. Experience in AGRANDA 2015

The experience in AGRANDA 2015 within the 44th JAIIO (Argentine Conference on Informatics), which took place in Rosario, was very beneficial. The first conference we attended was "Tell me who you are talking with and I will tell you how old you are (using Big Data and Machine Learning)" given by Jorge Brea of GranData Argentina, in the Argentine Symposium on Artificial Intelligence (ASAI).

Then we participated in a Workshop on "Use of open source tools for processing large data volumes" by Lic. Ernesto Mislej.

Jorge Brea presented the work "The city Pulse of Buenos Aires" by Carlos Sarraute, Carolina Lang, Nicolás Ponieman and Sebastián Anapolsky, all of them from Grandata Labs.

Later, Camilo Melani presented the work "An end-to-end big data case: data analysis in transportation and its use in business," a work prepared by the presenter jointly with Juan V. Echagüe, Joaquín Torre Zaffaroni and Daniel Yankelevich, all of them from Pragma Consultores.

Jorge Brea returned with the work "Social Events in a Time-Varying Mobile Phone Graph," a joint work with Carlos Sarraute and Javier Burroni of Grandata Labs; Klaus Wehmuth and Artur Ziviani (Laboratório Nac. de Computação Científica) and J. Ignacio Alvarez-Hamelin (University of Buenos Aires –National Scientific and Technical Research Council [CONICET, for its Spanish acronym]).

Gabriel Tolosa of the National University of Luján presented "Using Big Data Analysis to Improve Cache Performance in Search Engines," made jointly with Esteban Feuerstein (UBA).

Cristian Antuña presented the work "A scalable and intelligent Real Time Bidding platform," made jointly with Juan Martín Pampliega and Claudio Freire, from the company Jampp.

Esteban Feuerstein presented the work by Facundo Malvicino (Interdisciplinary Centre of Studies in Science, Technology and Innovation) "Discovering Big Data in Argentina.

2014 Digital Survey".

Then, there was a conference by Dr. Bárbara Poblete (University of Chile – Prisma Research Group) "Understanding real world events through social media."

Then, the work "Weather Measurement and Visualisation with Azure-IoT" was presented by Adrián Fernández and Ariel Aizemberg, both from the Buenos Aires Institute of Technology.

Later, the work "Synthesis of a Data Processing Architecture of the INTA Anguil Weather RADAR" was presented by María de los Ángeles Martín, said work was made jointly with Mario José Diván and Guillermo Lafuente (School of Engineering of the National University of La Pampa); Yanina Bellini, María Laura Belmonte; Juan Marcelo Caldera, from the Anguil Agricultural Experiment Station of the National Institute of Agricultural Technology (INTA, for its Spanish acronym).

Continuing with this topic, Santiago Banchero, of the INTA - National University of Luján, disserted on "Hale forecasts using atmospheric indices," a work made jointly with Marcelo Soria (School of Agriculture - UBA) and Romina Nair Mezher (INTA).

It was then the turn of Federico E. Martínez of the National University of Quilmes with the work "Graph databases, a comparative analysis" prepared with Diego A. Aizemberg (Buenos Aires Institute of Technology).

He was followed by Gustavo Arjones of Socialmetrix with his work "Creating a Modern Architecture for Big Data Analytics."

Then Milton D. Pividori presented "Cluster Ensembles for Big Data Mining Problems," a team work made with Diego Milone and Georgina Stegmayer (the three from sinc(i) – School of Engineering and Water Sciences, National University of the Littoral).

The last work presented was "Discovering network relations in big time series with application to bioinformatics;" jointly made by Mariano Rubiolo (Research and Development Centre for Information Systems Engineering - National Technology University, Santa Fe Region School); Diego Milone and Georgina Stegmayer, and finished with a roundtable debate.

e. Workshop Experience

The workshop was attended by: Alejandro Baranek (Data Scientist); Laura Brestolli (Lawyer and student at the postgraduate course in Law and Technology at the University of San Andrés); Agustina Callegari (Bachelor in Communications – Centre for the Protection of Personal Data of the City Ombudsman's Office); Emanuel Calvo (Technical leader in Pythian and IT Consultant); Pablo Deymonnaz (Bachelor in Informatics– co–founder of ReProted Project[62]); Esteban Feuerstein (Ph.D. in Informatics – Sadosky

[62] http://www.reproted.com/

Foundation); Luciano Gandolla (Lawyer specialised in personal data protection); Mariana Impallari (Lawyer specialised in IT law – student at the Specialisation studies on IT Law– University of Buenos Aires); Eduardo Molina Quiroga (Ph.D. in Law, Coordinator of the Specialisation studies on IT Law – University of Buenos Aires); Javier Raimo (Lawyer – Centre for the Protection of Personal Data of the City Ombudsman's Office and student at the Specialisation studies on IT Law – University of Buenos Aires); Cecilia Ruz (Bachelor in Informatics, Data Mining expert, co-head of the Master in Exploitation of Data and Discovery of the Knowledge – University of Buenos Aires); Jorge Velázquez (Lawyer – student at the Specialisation studies on IT Law – University of Buenos Aires); Valeria Milanes (Director of the fields of Privacy, Freedom of Expression and Information Access at ADC); Vanina Mona (Director of Institutional Affairs in ADC); Leandro Ucciferri (Lawyer, researcher in privacy matters at ADC) and Tatiana Fij (Lawyer in the fields of Privacy, Freedom of Expression and Information Access at ADC).

The experience was very rewarding. The event took place on September 23, 2015 in the North Room at the NH Hotel, Buenos Aires, on July 9, located at Cerrito 154/156. It started at 9 am with the registration of all the participants. At 9.30 am, the round of presentations, knowledge and passions on the Big Data topic was opened.

The guests were divided into three tables with an expert/moderator in each of them. Esteban Feuerstein was the moderator in the first table (academic/technical table); the second table was run by Agustina Callegari (table on personal data protection) and the third table was run by Tatiana Fij (legal table). Each table discussed possible definitions from 10.15 am to 10.30 am, but in the plenary, carried out from 10.30 am to 11.10 am, they concluded that there is not a unified definition of Big Data. Therefore, the term is defined at the beginning of every talk, conference and Workshop.

Table 1 explained that they had not reached a new definition. The Big Data issue is something adopted by society. Some people are interested but they do not know to which extent the issue is a concern for society. Moreover, they analysed the Big Data topic with an example. In that respect, they said that nowadays, for instance, free Wi-Fi can be given to users while they are at the supermarket and that data can be used for data records and crossovers. In any case, volume increases. When free Wi-Fi is given, they measure how much time people spent at the supermarket and what they purchased there. However, according to Feuerstein, there has been a change (technological revolution) over the last few decades. Such technological revolution has developed and exceeded some thresholds. This caused the quantitative change to shift into a qualitative change, apart from mentioning that before things were made in a different way and now they can be bigger. There are some things that were unthinkable before and now they can be done thanks to Big Data. So, BIG became a qualitative change.

The talk led to one of the most interesting aspects for this project. That aspect is related to the social/human use of Big Data. Currently, there are more sides where

privacy may be affected than before. In that regard, Feuerstein stated that "…there are two-fold challenges: more data makes things more difficult. Protecting kilos of data is more difficult than protecting little data. More people participate; there are more channels through which information can leak, but this is also more complicate; it is a different challenge not only because of Big but also because there are so many things in that bag that more possibilities of combinations arise and put everything at risk…".

From the social point of view, they reaffirmed that the issue is something adopted by society. People are becoming interested but most of them prefer having videos and Internet for free even if it is not for free and they are risking their privacy to pay for that.

As mentioned before, the second table did not reach a definition either. With respect to that, they said that Big Data is not all the same and how we analyse the concept itself depends on where we are standing. Big Data is related to the storage of large volume of information and terms such as Data Scientist or Data Mining are connected to the analysis of such data. In general, when it comes to Big Data, we say "we use Big Data for X project." When the phrase "we use Dig Data" is said, we are including not only the storage but also the analysis of data.

In large volumes of information, there are structured, semi-structured and unstructured data. Currently, these techniques explore unstructured data (audio, video, image, etc.) that cannot be analysed with traditional databases. Finally, they said that training programmes are necessary. The Centre for the Protection of Personal Data – City Ombudsman's Office offers training programmes on online behavioural track[63]. The National Direction of Protection of Personal Data has the "Con vos en la Web"[64] (With you on the Web") programme, guides, videos and tutorials.

Emanuel Calvo indicated that Big Data is something that has already started, it is not over and nowadays we have a great possibility of processing information.

In table 3 a definition was not reached either but they did mention that Big Data includes large volumes of data analysed at a very fast velocity and with a variety of information and categories that were unthinkable in the past.

Furthermore, they pointed out that the issue is not the quantity but the way in which such data are analysed. There is not a minimum stage establishing what is big. In the past, they tried to find a sample as accurate as possible and show it to a bigger and total result. Nowadays, with these new techniques/technologies, maybe even more accurate results are achieved. Big Data presents a problem that could not have been analysed with traditional databases.

---

[63] http://www.cpdp.gob.ar/index.php?option=com_content&view=article&id=48&Itemid=61 and https://habeasdatacpdp.wordpress.com/2014/08/12/conectate-seguro-llego-a-las-escuelas-de-la-ciudad/

[64] http://www.convosenlaweb.gob.ar/

Law 25326 on personal data requires accuracy in collected data and Big Data does not necessarily complies with such requirement because there are data that are collected even from places where you cannot know whether data are real or not. Anyway, with the large volume of information that is being managed, a more accurate result can be achieved, maybe even more accurate than in the past. We face a problem, though. Some of these techniques/technologies may not be developed if data accuracy is demanded. The abovementioned steps apply if the study is done on personal data given that there are other issues that do not contain personal data: weather, traffic or data generated from dissociated information, cases in which the Law on the Protection of Personal Data would not come into play.

The purpose stated by the Law on the Protection of Personal Data is another issue that brings about a contradiction. Information collected many years ago with a different purpose or collected before Law 25326 came into force (year 2000) is sometimes used without consent from the data holder. There are companies that do not even include the Terms and Conditions in order to get the user's consent.

Furthermore, there is a lot of information being managed. The person managing the information can take it (case of the unlawful employee in commercial law). Confidentiality agreements must be entered into and they must be adapted to each case. Also, sometimes the value of data cannot be measured. There are companies, for instance, that keep historic records that they do not use. Sometimes, the importance of technology is overlooked.

From 11.30 to 11.50, Tatiana Fij presented the case study with a Power Point presentation. From 11.50 to 1 pm, random questions were answered while each table explained the conclusions they had reached to exchange views on the challenges and advantages of the implementation of Big Data techniques/technologies in projects proposed by the Governments.

One of the tables had to answer if they considered that the current or future volume of information generated by the VOS card programme, or any other variable resulting from said programme, will merit the use of Big Data or Data Mining techniques and/or technologies, if they are not in use today. Furthermore, they had to point out some advantages, disadvantages and challenges from a social and technological point of view in the implementation of the VOS card.

In this regard, they indicated that the current volume of the VOS card allows us to talk about Big Data and that it can be used to analyse people's behaviour: how they move around the city, for instance. The use is at the discretion of who manages the information. Respecting privacy and protecting personal data will be the advantages and challenges.

Another table had to answer if they believed that the manner in which the local government collects and/or analyses and/or processes data might potentially affect human

rights.

With respect to that, they talked first about the advantages that the VOS card may bring about, such as the promotion of public policies. The risks are the impact on privacy, intimacy, the so-called 'Big Brother'.

f. Relevant Laws

Relevant laws refer to Law 1845 on Personal Data Protection and National Law 25326 also on Personal Data Protection known as "Habeas Data Law". Given that this is a local study case (Autonomous City of Buenos Aires) and that the databases are registered in the Centre for Personal Data Protection – City Ombudsman Office, we will focus on analysing Law 1845 in depth.

Some definitions are given in Article 3 of Law 1845. One of the definitions indicates that sensitive data are those "personal data disclosing race or ethnical origin, political opinions, religious or moral beliefs, union membership, information related to health or sex life or any other data that may lead to, due to its nature or context, some kind of discriminatory treatment to the data holder."

According to information given by the Centre of Personal Data Protection in the access request, the database apparently only deals with identifiable data, not with sensitive data or data related to health or data on criminal or minor offences record. However, considering that medical records are intended to be included in the card, it could be said that sensitive data may indeed be stored in the future.

The Ministry of Modernisation was asked the following question:

Question 14: Is the medical record of hospital patients currently stored, or is it intended to be stored, in the VOS card? Is the medical record of patients from any other dependency stored or going to be stored?

The answer was:

> "Medical records of card holders are not stored in such cards, nor is it currently expected to store them, whether for hospital patients or patients from any other dependency." However, there is an inconsistency between this answer and the proposal of implementing the Vos card to store medical records, which was one of its original objectives[65].

The database of the 'En todo estás vos' network falls within the definition of the Law that states: "Files, records or databases or databanks: They indistinctively refer to the

[65] http://www.infobae.com/2014/06/21/1574726-en-todo-estas-vos-la-tarjeta-multifuncion-que-lanzo-el-gobierno-la-ciudad

group of organised personal data that is managed in an automated manner, regardless of its method or way of collection, storage, organisation or access, including both automated and manual." Personal data are managed in an automated manner and included in a database designed and protected by the Information Systems Agency of the Government of the City of Buenos Aires.

We reaffirm that said statement can be seen in the terms and conditions – privacy policy that establishes that "Personal Data will be automatized and included in the relevant automatised databases of personal information. VOS will be the holder and responsible for such databases (hereinafter, the 'Bases')".

With respect to that, data management is included in the definitions pursuant to Article 3: "Data management: any operation or group of operations, made or not made through automatised processes that permit collecting, preserving, classifying, storing, changing, relating, assessing, blocking, destroying, recording, organising, preparing, removing, using, comparing, deleting and, in general, processing personal data, as well as its assignment to third parties of any kind of communication, query, interconnection, transfer, dissemination, or any other way that gives access to such data".

Pursuant to article 3, the data holder ("Data holder: physical individual or legal entity whose data are being processed") is the one registering in the programme.

The person responsible for the file, record, database or databank, pursuant to article 3, is "any physical individual or legal entity from the public sector of the City of Buenos Aires that may be the holder of a file, record, database or databank."

As found in the replies to the access request, the Ministry of Modernisation and the Centre for the Protection of Personal Data informed that the Information Systems Agency, which collects the information, is responsible for the database. As regards the person in charge of the processing, article 3 states that it is "a physical individual or legal entity, public authority, dependency or body that, alone or together with others, manages personal data on behalf of the person responsible for the file, record, database or databank." The Information Systems Agency would be responsible for this, too.

With respect to the data user, article 3 establishes that such users are "individuals that, while performing their job and complying with their specific tasks, have access to personal data included in any file, record, database or databank of the public sector of the City of Buenos Aires." That is to say, the user is the one establishing the purposes of using the information. In this case, those users would be the Ministry of Modernisation, the Information Systems Agency and even the shops participating in the programme, granting discounts and benefits.

Article 4, section 3 of Law 1845 establishes "Rules on the creation, modification or deletion of files, records, databases or databanks belonging to public bodies shall be published in the Official Bulletin of the City of Buenos Aires and indicate:

a. Characteristics and purpose of the file;
b. People from whom data is intended to be obtained and the optional or mandatory nature of such information supply;
c. Process for obtaining and upgrading data;
d. Basic structure of the file, whether automated or not, and the description of the nature of personal data it will contain;
e. Estimated assignments, transfers or interconnections;
f. Body responsible for its filing, specifying hierarchical dependency in such case;
g. Dependency in which citizens can exercise the rights granted by virtue of this law.

Regarding that issue, the Ministry of Modernisation was asked, through the access request for information, if rules on the creation, modification or deletion of files, records or databases had been published in the Official Bulletin of the City of Buenos Aires. They were reluctant to answer this question by saying that "as way of example, Law 1845 on 'personal data protection' was published in the Official Bulletin No 2494."

Said article 4, section 5 points out that: "In the case of personal data collected through the Internet, interactive sites of the City of Buenos Aires shall inform the personal data holder about the rights granted by this law and the national law through a privacy policy placed in a visible part of the website." As analysed before, this is not the case, because we consider that the terms and conditions cannot be accessed easily and, in addition, they are not accepted before clicking 'submit request' so that the card is requested in the service centre chosen for picking up the card.

It is known that in most cases users do not read the terms and conditions carefully, but in this case such terms and conditions are not available for the user to read them. Therefore, the user accepts compulsorily and consents to the terms and conditions and to the privacy policy that are unilaterally established (adhesion contract) without knowing their content.

The terms and conditions and the privacy policy should be available to the user and, in this case, a system should be established to prevent the request from being sent if the user does not click on the box stating that the terms and conditions have been read. Usually, people click on that box without having read the terms and conditions. Thus, a scroll system should be implemented so that the Government of the Autonomous City of Buenos Aires ensures that the user has, at least, scrolled from the beginning until the end of the document

It is important to take into account the provisions on Article 7 of Law 1845 about the consent granted by the data holder:

"Article 7.-Consent:

1. The management of personal data is considered unlawful when the holder has not granted, by writing or in another similar manner according to the context, his free, informed and express consent. Such consent granted with other statements shall be clearly expressed and emphasised, prior notice given to the data holder, according to his social and cultural level, about the information referred to in article 18, section b) of this law.

2. Consent may be declared void by any means and at any moment. Such declaration shall not have retroactive effects.

3. Consent shall not be necessary when: personal data are collected for performing functions of the powers of the City of Buenos Aires or by virtue of a legal duty; personal data are collected from unrestricted public access sources; personal data is related to the health of the individuals and their processing is necessary due to public health and emergency reasons established by the relevant authority and dully grounded; lists whose data are restricted to name, ID, tax or pension identification, profession, date of birth and address."

Article 13 sets forth information rights to access, correct, update or delete data in accordance with Law 25326. Section a) establishes: "Information right: Any person may request any information related to the existence of files, registries, databases or databanks to the monitoring body from the public sector of the City of Buenos Aires as well as *their purposes, identity and address of the person responsible for such data*" (emphasis added).

It is worth mentioning that the right to access indicates that: "After verifying their identity, data holders are entitled to request and obtain information related to their personal data included in the files, registries, databases and databanks of the public sector of the City of Buenos Aires. Furthermore, in the same way, the data holder may demand information about the identity of the people who may have been assigned data related to such data holder, about the origin of any data included in the file, registry, database or databank and *about the rationale used when data has been subjected to automated processing*" (emphasis added).

The Ministry of Modernisation did not answer the question with respect to what the abovementioned article of Law 1845 establishes about the rationale used when data are subjected to automated processing and in the process of automated decision-making of data. The reply was "data entered by the user are subjected to a rationale prior to the control of consistency and if the result is negative (for example, if an e-mail address lacks the symbol '@' ") the applicant is informed about this situation in

an automated manner so that such data are corrected online prior to completing the registration."

With respect to the right to correct, update or delete data, article 13, section c) states that: "Any person is entitled to have his personal data corrected, updated and, where appropriate, deleted or subject to *confidentiality*" (emphasis added).

In the access request filed with the Centre for the Protection of Personal Data we asked: "Which are the guarantees for enforcing the data holder rights, in accordance with article 13 of Law 1845.

The Centre answered that "The office before which data holders may enforce their rights (access, confidentiality, corrections, updates and deletion) is the General Office of Smart City Projects located in Bernardo de Irigoyen 272, second floor, of the Autonomous City of Buenos Aires, telephone number 4323-9300, Extension 8120259, e-mail address: dgpciudadinteligente@gmail.com." This is inconsistent with the privacy policy of the Terms and Conditions that states that:

> "Subscribers are entitled to the rights of access, deletion, correction and objection as well as the right to be informed about access requests that are submitted, by contacting VOS through the Customer Service Centre 147…"

The access request for information submitted to the Centre for the Protection of Personal Data- City Ombudsman's Office asked about question 3: "Who are the recipients and the categories of recipients of such files, registries, databases or databanks?" and question 4 "What are the conditions on the organisation, functioning and procedures applicable to such files, registries, databases or databanks?"

As a reply, they have pointed out that: "With respect to question in point 3 and pursuant to article 18 of the Law on Data Protection, the person responsible for the database must inform data holders about the cases you mentioned. If users do not get an answer, they can file a request through our Office. The same applies to question in point 4."

Law 1845 establishes, in article 18, that the person responsible for files, registries, databases or databanks must: "(…) b) Inform the data holder, expressly and clearly, and subject to annulment, prior to gathering information about such holder, about: the existence of the relevant file, registry, database or databank, in a digital or any other format, and the identity and address of the responsible person; the purpose for processing such data and who their recipients could be or the categories of recipients…" but article 24 states that:

> "Article 24. - Any person will be able to know about the existence of files, registries, databases or databanks containing personal data, *their purpose*,

the identity and address of the responsible person, *recipients and categories of the recipients, conditions on the organisation, operation and applicable procedures, security standards*, guarantees to enforce the data holder's rights as well as any other registered information.

The *monitoring body shall verify, upon request by an interested party* or by law in case of a suspected illegal act, *compliance with legal and regulatory provisions regarding each of the following stages of the use of personal data:*

a. Legality of the collection or taking of personal information;
b. Legality of data exchange and legality in the transfer to third > parties or in the connection among them;
c. Legality of the assignment itself;
d. Legality of internal and external control procedures of the file, > registry, database or databank"(emphasis added).

The Centre has not answered the fourth question. The Centre for the Protection of Personal Data should have replied the question about the conditions on organisation, operation and applicable procedures as regards files, registries, databases and/or databanks.

Questions 5, 6 and 7 from the access request for information submitted to the Centre for the Protection of Personal Data asked about: "5. What are the security standards for files, registries, databases and/or databanks?", "6. State if VOS CARD from the network 'EN TODO ESTÁS VOS' complies with the characteristics of security, confidentiality and discretion provided in Law 1845 from the Autonomous City of Buenos Aires and/or Law 25326, and finally"7. Has the person responsible for the files, registries, databases or databanks obtained any relevant authorization?"

In this sense, they have pointed out that:

"As regards points 5, 6, and 7, the original law established in Article 23:"Registry of personal data. Create the Registry of Personal Data within the scope of the Ombudsman's Office of the City of Buenos Aires. Such registry shall have the following purposes:

Authorise the creation, use and operation of the files, registries, databases and databanks of the public sector of the City of Buenos Aires in accordance with the provisions of this law.

Establish the requirements and procedures that shall be complied with by the files, registries, databases and databanks of the public sector of the

City of Buenos Aires as regards the processing of data collection, general design of the system, including any necessary security and control procedures, technical equipment, procedures taken in order to guarantee the rights to access, delete, correct and update as well as any other relevant end".

It's worth mentioning that the highlighted parts in bold and italics were declared void by Decree No 1.914/05, Official Bulletin No 2351 on 4/1/06. Therefore, these are not prior requirements in order to subscribe to our Registry of Databases4

On the other hand, this Agency requires, in order to determine who the person responsible for the base is, the appointment of the public officer and the rules on the creation of the base to be registered".

Question 5 is not part of article 23 of Law 1845 that was declared void *but of article 24 of said Law* in the following part highlighted in bold:

"Article 24. - Any person will be able to know about the existence of files, registries, databases or databanks containing personal data, their purpose, the identity and address of the responsible person, recipients and categories of the recipients, conditions on the organisation, operation and applicable procedures, *security standards*, guarantees to enforce the data holder's rights as well as any other registered information…" (emphasis added).

Question 6 is not part of article 23 of Law 1845 that was declared void *but of article 20 of said Law* in the following part highlighted in bold:

"Article 20. - Duties of data users. Any individual that acts, works or provides services of any nature in or for bodies from the public sector of the City of Buenos Aires shall only deal with personal data contained in files, registries, databases or databanks from the body for or in which they perform their job upon authorisation granted by the person responsible for such file, registry, database or databank or by virtue of a legal duty.

They are subject to, as well as any person in charge of processing, the same responsibilities and duties imposed to the person responsible for the file, registry, database or databank, *both with regard to maintaining the appropriate confidentiality and discretion about collected information as well as respecting and complying with the general principles on the protection of personal data*.

> The data user may only assign personal data subjected to processing by following express instructions given by the person responsible for such processing"(emphasis added).

Security aspects are part of the general principles on protection of personal data and of the provisions established in article 24 of Law 1845.

Finally, question 7 is not part of article 23 of Law 1845 that was declared void *but of a SECOND Temporary Provision* stating as follows:

> SECOND: In a period of a hundred and eighty (180) days as of the approval of the provisions, *any person responsible for the files, registries, databases or databanks from the public sector of the City of Buenos Aires performing tasks governed by this law, shall obtain the relevant authorisation and its subsequent registration in the Registry of Personal Data*"(emphasis added).

Finally, and taking into account that the Centre for the Protection of Personal Data– City Ombudsman's Office is the monitoring body of Law 1845, it is worth establishing that, in accordance with the purposes set forth in article 23, the Centre has not given any warning, recommendation, reminder and/or reply to the Government of the Autonomous City of Buenos Aires with respect to VOS card.

The Centre has not suggested or started disciplinary and/or legal proceedings as regards said card nor has the Centre represented data holders to enforce the rights of access, correction, deletion and updating of the bases from the Information Systems Agency as regards the VOS card from the network 'En todo estás vos'. Lastly, the Centre has not assisted data holders in trials carried out due to the use of the information contained in the VOS card.

5. Obstacles, Challenges and Conclusions: V for Vulnerability

The advantage of the suggested methodology throughout the Project is that ADC has shown knowledge and experience in this type of investigation. According to an IDC report [66] there are four main obstacles to be faced when implementing Big Data technology:

1. *Lack of specialised experts due to the novelty of the subject*. It is necessary to point out that in Argentina the Master in Exploitation of Data and Discovery of Knowledge of the University of Buenos Aires annually receives more people who wish to enrol and it has graduates in different areas. Furthermore, the Master has been

[66] Joyanes Agullar. Luis., "Big Data. Analysis of large volumes of data in organisations", 1st edition, Alfaomega Grupo Editor Argentino, Buenos Aires, 2013.

organising different workshops and conferences to discuss the topic of Big Data in different areas (public sector & data, Software & Business Intelligence, Internet & Online Marketing, health, agriculture) with both academic and business representatives and State public officers. There is also a 'Big Data Certification' at Buenos Aires Institute of Technology (ITBA) containing five modules where students are taught Data Warehousing and OLAP; techniques and tools for Big Data; Data Mining for Business Analytics, data visualisation. Said certification finishes with a final project.

2. *Lack of budget*. With regard to this issue, there is a possibility of using software that is distributed and developed freely (open source), thus saving up costs.

3. *Lack of integration with business processes and, in our case, with the State organisation*. In order to be integrated, Big Data must include at least the 4 dimensions: volume, variety, velocity and value (IDC refers to the 4-V model.) We consider that it must include at least the V for velocity, volume and variety.

4. *The fourth issue is that of the quality of data*. The speed in the implementation of Big Data technologies varies according to their features, with especial influence of the size and the activity where they are developed.

ADC adds two more problems: lack of legal provisions on the subject and, additionally, the issues of privacy and personal data protection.

Therefore, the privacy of information could be violated by Big Data phenomenon if appropriate security measures are not taken and if data are not anonymised and encrypted. Within the 4 or 5 Vs defined in this exploratory study, ADC adds one more V: the V for Violation of rights if an appropriate legal framework is not taken into account when analysing the topic.

Another problem is the novelty of this specific topic and the lack of background related to it. Whether to use Big Data technologies or techniques or not is a matter of public policy because the budget issue can be settled. Moreover, in Argentina we are not advanced in the issues of personal data protection and privacy to be able to deal with them. Upgrading the Laws on Personal Data Protection is necessary to face this new disruptive phenomenon that Big Data represents, preventing the laws from affecting technology development.

It has been shown that decision-making based on Big Data and the use of large-scale data compromise issues included in laws on both personal data protection and privacy. As it has been seen in the case study of the VOS card, there are databases, personal data, so consent must always be granted, unless there are especial exceptions. Apart from what has been stated, the VOS card is optional. This issue would be different if the card were mandatory (e.g. for underground use, in medical record storage.) The

question that arises is which the purpose is of collecting data in this card. Some information contained in the databases (e.g. medical records) shall be processed separately and, likewise, the prohibition to store sensitive data except in the case established in article 8, section 3 of Law 1845 shall be taken into consideration: "The creation of files, registries, databases or databanks that store information disclosing sensitive data directly or indirectly is prohibited, unless otherwise expressly provided for in this law or in any other law or upon prior, free, express and written consent granted by the data holder." Then, if the purpose is making things easier, available and accessible, the benefits/discounts shall state the deletion of such data once promotions have finished.

In this study we have seen that too much data are requested when registering with the card and that the VOS card accesses, keeps and processes information and the participating locations where subscribers use their benefits disclose to VOS card the information related to transactions made by Subscribers. Furthermore, the terms and conditions establish that the VOS card processes or may transfer information contained in its database to its agents and/or Participating Locations for advertising, promotional and commercial purposes, or for any other purpose (e.g. to determine interests and/or affinities of subscribers so that the Benefits are tailored to their interests, and to achieve the best performance of the Programme.) What also draws our attention is that far too much information is being disclosed through the relationships between Government - user, user - service or product provider, and Government -provider-service and product. Such information shall be protected and consent shall be granted.

With respect to data confidentiality, security measures against hacking or denial of services shall be taken into account. The Ministry of Modernisation informed that they are carrying out a penetration tests and Security assessments. Furthermore, they indicated that information is stored in ASI's Data Centre, in state-of-the-art equipment and security controls such as firewall, IPS, antiDDoS. Such Data Centre also keeps the information systems for the Government as well as the Card, disks, servers, security schemes, backup, among others. Data shall always be anonymised and encrypted. Anyway, anonymising data, as pointed out in the Workshop on September 23rd, does not mean that you cannot go back. There is a possibility that the person dealing with the information may use it for any other purpose.

In the future, there may be a conflict with the assignment of databases. This may happen in the case of the storage of medical records and the use of the card as a means for accessing the underground. For such purposes, the Ministry of Health in the former case and the Ministry of Transport in the latter would be granted an assignment of the databases. It is necessary to point out that in this case article 10, section 1 of Law 1845 states that: "Personal data subjected to processing may only be assigned for compliance with the purposes directly related to the lawful interest of the assignor and the assignee and upon prior consent of the data holder who must be informed about the purpose of the assignment, identifying the assignee or the elements that allow identifying the assignee."

The Law on Personal Data Protection, both at a national and local level, establishes that data shall be true, appropriate and relevant and this implies a conflict with the type of information that is collected.

As it has been mentioned before, there is a big problem related to the fact that the terms and conditions and the privacy policy are not visible for the subscriber. That makes us think how effective they are for the user if they are not able to know them. In general, people are annoyed or indifferent about reading them but a solution, from the Government of the Autonomous City of Buenos Aires, shall be given so that the user accepts and understands such terms.

The model of provisions in Laws on Personal Data Protection sets limits, prohibitions and plays an important role in the effectiveness of protecting personal data and privacy. Such laws shall adapt to these new techniques for collecting, storing, processing, analysing, managing and using massive data.

As shown by IDC, it is necessary to define how generated data shall be collected, classified and stored, how to build architectures with high scalabilities, how to create new database models that are able to collect information in real time and with the purpose of collecting[67]. For such end, it is necessary to have equipment with a lot of RAM memory, disks, data centres, servers to store all that amount of data.

Big Data has made it easier to discover some things which could have been discovered earlier but in a much longer period of time. Therefore, implementing a solution towards Big Data means integrating different elements and projects that constitute the essential ecosystem to analyse large amounts of data[68].

6. References

a. Academic papers and magazine articles

- Accenture. 2014 Report.,"Big success with Big Data. Executive summary". Available on: https://www.accenture.com/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Local/es-la/PDF2/Accenture-Big-Data-POV-espanol.pdf
- Barranco Fragoso, Ricardo. What is Big Data? IBM Developer Works, 06/18/2012. Available on: https://habeasdatacpdp.wordpress.com/2014/08/12/conectate-seguro-llego-a-las-escuelas-de-la-ciudad/
- Bryant. Randal E., Katz. Randy H., Lazowska. Edward D., "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society",

[67] Joyanes Agullar. Luis., "Big Data. Analysis of large volumes of data in organisations", 1st edition, Alfaomega Grupo Editor Argentino, Buenos Aires, 2013.

[68] https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/

Computing Community Consortium, Version 8: 12/22/2008. Available on: http://cra.org/ccc/wp-content/uploads/sites/2/2015/05/Big_Data.pdf

- Cisco. Report: "Internet will be four times bigger in 2016", 05/30/2012. Available on: http://www.cisco.com/web/ES/about/press/2012/2012-05-30-internet-sera-cuatro-veces-mas-grande-en-2016--informe-vini-de-cisco.html
- Curto. José. "Executive Summary- Big Data: an emerging market", sponsored by: EMC, JasperSoft, Microsoft y Sybase, an SAP Company, June 2012. Available on: http://www.portalidc.com/resources/ponencias/big_data12/Resumen_Ejecutivo.pdf
- Diaz. Adam, Wampler. Dean, Smith. Tom, Langseth. Justin, Hausenblas. Michael, "DZone's Guide to Big Data, Business Intelligence, and Analytics", 2015 Edition. Available on: https://dzone.com/guides/big-data-business-intelligence-and-analytics-2015
- Eaton. Chris., Deutsch. Tom., Deroos. Dirk., Lapis. George., Zikopoulos. Paul., "Understanding Big Data. Analytics for Enterprise Class Hadoop and Streaming Data, Mc Graw Hill, 2012. Available on: https://www.ibm.com/developerworks/vn/library/contest/dw-freebooks/Tim_Hieu_Big_Data/Understanding_BigData.PDF
- IBM., Press release. "Telefónica of Argentina implements IBM technology to know their clients better" Available on: http://www-03.ibm.com/press/ar/es/pressrelease/43447.wss
- IBM Global Services Route., Executive Report. IBM Institute for Business Value. "Revolution in the analytic skills and business optimization. When the new intelligence and business operations meet", November 2009. Available on: http://www-05.ibm.com/services/es/bcs/html/bao/bao-revolucion-en-las-capacidades-analiticas-y-optimizacion-de-negocio.pdf
- Bankinter Innovation Foundation. Report: "Big data the power of data" Available on: https://goo.gl/Bx46fa
- Galeano. Eduardo., "Open veins of Latin America", Publisher Siglo XI de España Editores S.A., Spain, (2008).
- Joyanes Agullar. Luis., "Big Data. Analysis of large volumes of data in organisations", 1st edition, Alfaomega Grupo Editor Argentino, Buenos Aires, 2013.
- Kroes. Neelie., "Big data for Europe", European Commission. 7/11/2013. Available on: http://europa.eu/rapid/press-release_SPEECH-13-893_en.htm

- Lantares Solutions. “The five main applications of Big Data”, Blog on Business Intelligence. Available on: http://www.lantares.com/blog/las-cinco-principales-aplicaciones-de-big-data
- *Leber. Jessica., “*Big Oil Goes Data Mining“, MIT Technology Review, 5/8/2012. Available on: http://www.technologyreview.com/news/427876/big-oil-goes-mining-for-big-data/
- Marín Diazaraque. Juan Miguel., “Introduction to Data Mining”. Available on: http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/introduccion-DM.pdf
- Merv. Adrian., “Big Data”, Teradata Magazine, 2011 Q1. Available on: http://www.nxtbook.com/nxtbooks/mspcomm/teradata_2011q1/index.php?startid=8#/42
- Malvicino. Facundo., Yoguel. Gabriel., “Big Data: Recent Progress at an International Level and Perspectives for Local Development.”, Published by the Ministry of Science, Technology and Productive Innovation. Available as “Big Data” on http://www.mincyt.gob.ar/publicaciones
- Manyika. James., Chui. Michael., Brown. Brad., Bughin. Jacques., Dobbs. Richard., Roxburgh. Charles., Hung Byers. Angela from The McKinsey Global Institute., “Big Data: The next frontier for innovation, competition, and productivity”, May 2011. Available on: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- Ortoll. Eva., “Big Data is written with V”, COMEN Magazine of Information and Communication Science Studies, October, 2014. Available on:http://comein.uoc.edu/divulgacio/comein/es/numero37/articles/Article-Eva-Ortoll.html
- “Palenque Project”, Sadosky Foundation, 07/15/2014. Available on: http://triton.exp.dc.uba.ar/datamining/files/Charlas_y_Paneles/p4_fsadosky.pdf
- Rua. Martina., “Big Data technology , the secret weapon of the German team in the World Cup” 07/13/2014.Available on: http://www.perfil.com/ciencia/La-tecnologia-Big-Data-el-arma-secreta-de-la-seleccion-alemana-en-el-Mundial-20140713-0043.html
- Mayer Schönberger. Viktor., Cukier. Kenneth., “Big Data: A Revolution That Will Transform How We Live, Work, and Think”, An Eamon Dolan Book, Houghton Mifflin Harcourt, Boston, New York, 2013.
- Salvador. Fabrizio., “Big Data: the route or the destination?, Advanced Series Foundation, 2004. Available on: http://www.ie.edu/fundacion_ie/Comun/Publicaciones/Publicaciones/Big%20Data%20ESP%207.pdf

- The World Economic Forum., Big Data, Big Impact: New Possibilities for International Development, 2012. Available on: http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf
- Wired Magazine., "The Petabyte Age: Because More Isn't Just More — More Is Different", 06-23-2008. Available on: http://archive.wired.com/science/discoveries/magazine/16-07/pb_intro

b. Websites

- https://www.google.es/trends/explore#q=big%20data
- http://dilbert.com/search_results?page=7&terms=book+deal
- http://nosql-database.org/
- http://www.imdb.com/title/tt0181689/?ref_=fn_al_tt_2
- http://espacio.fundaciontelefonica.com.ar/big-bang-data/
- http://www.censo2010.indec.gov.ar/
- https://twitter.com/gcba/status/593162702287147011
- https://www.youtube.com/watch?v=d9NJt4DBb-I&feature=youtu.be
- http://www.wired.com/insights/2013/04/entering-the-exabyte-era-new-ways-to-support-mobile-datas-rapid-growth/
- http://tecnologia.elpais.com/tecnologia/2015/05/08/actualidad/1431077958_176608.html
- http://whatsthebigdata.com/2013/06/03/big-data-quotes/
- http://www.24con.com/conurbano/nota/93755-cuantas-personas-delgba-trabajan-por-dia-en-capital/
- http://www.infobae.com/2014/06/21/1574726-en-todo-estas-vos-la-tarjeta-multifuncion-que-lanzo-el-gobierno-la-ciudad
- http://www.iprofesional.com/notas/189842-El-gobierno-porteo-lanz-una-tarjeta-multifuncin-para-usar-en-la-Ciudad
- http://politicacomunicada.com/smartcities-el-proximo-negocio-digital/
- http://www.buenosaires.gob.ar/tarjetavos
- http://www.buenosaires.gob.ar/tarjetavos/preguntas-frecuentes

- http://www.buenosaires.gob.ar/tarjetavos/terminos-y-condiciones
- http://www.buenosaires.gob.ar/ecobici/sistema-ecobici
- http://home.econ.uba.ar/economicas/?q=content/big-data-desafíos-y-perspectivas-para-las-ciencias-sociales
- http://www.fundacionsadosky.org.ar/workshop-argentino-en-ciencia-de-datos/
- http://www.cpdp.gob.ar/index.php?option=com_content&view=article&id=47&Itemid=58
- http://www.prismamediosdepago.com/
- http://www.reproted.com/
- http://www.convosenlaweb.gob.ar/
- http://www.cpdp.gob.ar/index.php?option=com_content&view=article&id=48&Itemid=61
- https://habeasdatacpdp.wordpress.com/2014/08/12/conectate-seguro-llego-a-las-escuelas-de-la-ciudad/

7. Research Contributors

- Valeria Milanes, Freedom of Expression, Access to Information and Privacy Areas Director from April 2015. Valeria is an experienced lawyer, graduated from the Faculty of Law and Social Sciences of Buenos Aires University in 2002. Curious and passionate for IT & Internet issues and their legal impact. She is first cohort of IT Law Specialization Degree of Faculty of Law and Social Sciences of Buenos Aires University. She integrates the IT Law Institute of City of Buenos Aires Bar Association. She is a researcher of legal aspects related to ICTs and Internet as well as regular speaker in thematic events.
- Gastón Darío Nápoli (specialised technical consultant): Currently integrating the staff of the area of R + D + i of VATES Software Engineering SA (http://www.vates.com/), dedicating the last years to the research and implementation of new technologies, among others, Big Data solutions being certified professional in this area in IBM tools for Big Data. It has nearly 20 years experience in the IT industry, during which he has served in multiple areas and seniorities, developer, analyst, presales, technical leader, architect, project manager and trainer among other positions, both for local client and abroad.
- Leandro Ucciferi (junior assistant): Leandro Ucciferri is a lawyer, graduated from the Faculty of Law, Political and Social Sciences of the University of Morón (UM)

in 2015. He is part of the teaching team of the subject "Informatic Law" of the Bachelor of Information Systems in Organizations in the Faculty of Economics (UBA). Passionate about technology and the Internet, he writes about it professionally since 2012 and advocates for the defense of privacy in the digital age.

8. Acknowledgements

To Gastón Nápoli, Valeria Milanes, Vanina Mona and Leandro Ucciferri for their teamwork collaboration in this exploratory work.

To Esteban Feuerstein, Diego Bendersky, Facundo Malvicino, Cecilia Ruz, Gabriel Baum, Eduardo Poggi, Marcelo Soria, Adriana Echeverría, Ana Silvia Haedo, Ernesto Mislej, Djamel Abdelkader Zighed, Carlos Diuk, Dan Rozenfarb, Carlos Sarraute, Juan Martín Pampliega, Gustavo Arjones, Francisco Fernández, Gabriel Taubin, Gustavo Brey, Nicolás García Dutriez, Santiago Gonzalez Venzano, Gabriel Infante-López, Walter Sosa Escudero, Germán Stalker, Mariano Tamburrino, Nicolás García Dutriez, for their collaboration and academic input in the project.

To Esteban Feuerstein and Cecilia Ruz again, Eduardo Molina Quiroga, Agustina Callegari, Javier Raimo, Luciano Gandolla, Pablo Deymonnaz, Laura Brestolli, Mariana Impallari, Jorge Velázquez, Emanuel Calvo, Alejandro Baranek and Valeria Milanes, Leandro Ucciferi and Vanina Mona for their predisposition, participation and great efforts in the Workshop that took place on 09/23/2015.