# Analysis of the Big Data Strategy in Spain

Felipe Miron, Mabel Lopez, Clara Pezuela, Nuria De Lama, Juan Carlos Trujillo, Juan Luis Sobreira, Miguel Angel Mayer, Patricia Miralles, Amelia Martin, Fernando Martin, María Belen Garcia, Jesus Poveda.

*Big Data Initiative*

***Abstract* —In the last years, the phenomenon known as Big Data [1] [2] [3] [4] has been gaining the attention and interest of businesses and researchers all over the world. Many large businesses have already adopted Big Data or have started to. However, public organizations and private businesses agree[5] that the adoption of Big Data raises serious challenges and is not trivial or immediate develop and implement of successful Big Data projects. Those challenges stem from the Big Data own idiosyncrasy and main properties that, are commonly known as the 5Vs of Big Data (Volume, Velocity, Variety, Veracity and Value). Therefore, the current predictions around the Big Data phenomenon are both positive and negative at the same time. Gartner Group predicted 1 3 Big Data projects would create around 4.4 million ICT (Information and Communication Technologies) jobs all over the world. However, this information dissipates in terms of predictions that state that 85% of the business that are part of Fortune 500 will not be able to efficiently exploit the Big Data analysis for competitive purposes.**

**Nonetheless, Spain is not and hasn't been oblivious to the Big Data phenomenon. Therefore, this whitepaper intends to describe the current state of Big Data in the Spanish scenario and presents a Big Data initiative promoted by PLANETIC to establish the current status of the research, infrastructures, existing technologies and available data spaces in the Big Data context to identify the strengths and room for improvement. This document includes the data collection work carried out to date and the subsequent analysis concerning the identification of funding priorities and, at the same time, boosts and divulges the existent value.**

***Keywords*— Inter-Platform Initiative, Big Data, Spain.**

## I. THE BIG DATA INTER-PLATFORMS INITIATIVE

The Big Data Inter-Platforms Initiative, promoted and coordinated by PLANETIC, is an initiative of several Spanish industrial sectors, represented by their own technological platforms, created to provide an answer to the increasing demand and expectations generated around the data and its processing. Big Data is an environment which implies great technological and regulatory challenges, paired with the need of establishing cooperating multidisciplinary teams to cover the wide range of technologies involved (HPC, Cloud, security, semantic, interfaces, IoT, eServices, etc.). Inside the European framework which promotes the creation of a single data market, the objective is to put Spain at the forefront of the European initiative discovering and providing solutions to the current business opportunities involving Big Data.

The article is structured as follows: the main objectives of the Big Data Initiative promoted by Planetic are presented in the 2nd Section. This section is followed by the 3rd Section which describes the national strategy and Spain's positioning in Big data technologies. And, finally, the 6th Section includes and describes the most prominent national, European and international projects that focus on addressing the Big Data challenge.

### *1. OBJECTIVES*

PLANETIC's main objectives for the Big Data platform initiative are:

- To create a group of Spanish agents with skills and interests related to the Big Data environment to, in an open and coordinated manner, define a global vision and align several contexts (sectors, governments, businesses, academies, legal framework) and aspects (technological, regulatory, social, legal, etc.).
- To align the national and European strategies in Big Data that started through the PPP Big Data Value:
  - To identify the strengths and weaknesses of the national competition in Big Data.
  - To identify the I+D potential of the Big Data technological platforms.
  - To identify the needs and requirements related to training and analyze how to satisfy them.

- To recommend available research lines for the configuration of R+D funding programs.
- From this group, feed the European PPP with the national priorities and keep an open bidirectional communication line with the said PPP.

- To explore the implementation processes of similar initiatives in other EU countries to extrapolate practices that could be applied in Spain and use the previous experience of those countries.
- To encourage the Spanish participation, through the platforms, in European and national calls to consolidate research and innovation projects aligned with Big Data.
- To divulge the benefits of the use and exploitation of Big Data to the business industry and public administrations.
- To guide the members of the platform (specially SMEs) in the service exploitation of Big Data through documentation, workshops, consulting services etc.
- To identify success cases in ecosystems where the creation of sharing spaces and shared exploitation of the data means an increase in competitiveness in a certain market or sector.
- To try to replicate successful cases in future ecosystems using “lighthouse” projects or pilot projects.

## II. EUROPEAN STRATEGY

In October 2013, the conclusions of the European Council mostly focused on digital economy, innovation and the quality standard of services as an engine for business growth and job creation. The need for EU to intervene and establish the appropriate conditions and work frames for a common and single market for Big Data is identified.

In March 2016, Parliament finished shaping these new directives and published the ‘Towards a data driven economy’6resolution introduced by the Committee on Industry, Research and Energy.

The Initiative, as stated in the previous chapter dedicated to the review of goals and objectives, pursues the alignment of this European strategy within the national landscape by bringing to the table a weakness and strengths analysis in territorial entities and even by helping the development of financial programs based on the results of the analytics performed.

| **Big Data EU Observatory** |
|---|
| **2017** |
| ICT market growth rate x 6 = Big Data market growth rate |
| Market movement: 50,000M€ y generation of 3.75M jobs |
| **2020** |
| Available 16 billion Gb |
| Annual growth of 236% in the generation of data |
| Growth of 1.9% in GDP |
| **Requirements** |
| Greater control of risks in personal data |
| Safeguard and protection of consumer rights |
| Sectorial balance in Big Data operating profits |

*Figure 1. EU Observatory for Big Data (Source: European Parliament resolution on the initiative "Towards a prosperous data economy" (2015/2612 (RSP)).*

### *A. Recommendations and necessary actions*

The European Union considers that the strategy core for a unified digital market strategy should be the creation of a data-based economy due to high value and potential, helping the E.U. competitiveness recovery process in advanced sectors. It will only be possible with the existence of a proper business environment, means and data protection measures to activate digital transformation. To do so, the following actions should be undertaken:

- Development of a standardized framework to adjust access, control and ownership of big data.
- Investment in research and development focusing mainly on cloud computing, broadband and high-speed connections focusing on “weak spots” or areas with deficient level of infrastructure in order to open those market segments.
- Creation of public-private partnership associations in collaboration with research centers, universities and enterprises encouraging innovation.
- New policies should be implemented to incentivize and stimulate investments in research and development and

establish regulations to control the patent infringement of basic standards.

- Shared research to address the lack of interoperability, need of a common interface to connect hardware and sensors, real and virtual world and improve open data availability.

The scope of this strategy strongly calls for a specific regulatory framework which encourages the creation of new enterprises in terms of equality and free market avoiding unreasonable regulations and granting the compliance with privacy and data protection current regulations.

In addition, the EU stresses the need for the implementation of a digital entrepreneurship strategy and the creation of Digital Economy Centers specialized in the manipulation of Big Data. The promotion of Innovation Spaces and a better coordination process with the Educational System is also needed. ICT needs to be made interesting for students and they need to be attracted to it.

In order to do so, community participation is absolutely necessary. To guarantee this, parliament asked for initiatives able to increase collective consciousness and encourage public debate about ICT benefits, especially in those segments close to the digital breach.

This route is the strongly recommended to the state members in order to speed up e-government adoption.

### *B. Big Data legal framework in the EU*

Within the business environment and considering that personnel data management has acquired a significant economic importance in the current market and particularly in the big data segment, the main goal of unifying the European standards regarding data protection is to encourage new business opportunities and innovation. Among all different measures proposed, the following stand out: Regulations that will establish a unique set of rules that will make it easier and cheaper for companies to do business within the European Union.

- Enterprises will only have to go through a single point of supervision.
- Companies outside the EU will have to apply the same set of rules as the companies within the EU in case they decide to do business within European territory.
- Regulations will be based on account risk factors for each company.
- Regulations will grant data protection to be integrated into product and services from early stages of development.
- Small and medium sized companies will benefit from significant drops in prices and less bureaucracy in the business process. The goal is to eliminate the requirement to notify the supervisory authorities. This will encourage the collection of fees if excessive data requests are made. It will even remove the obligation to appoint a data protection officer if data processing is not the main activity. Impact assessments will also be avoided unless there is a high risk.

### *C. Strategic agenda: PPP and funding calls*

BDVA (Big Data Value Association)[7] is a nonprofit organization created under Belgian law and a private partner called Big data value PPP (Public-private partnership). Th public part of this association is represented by the European commission. The PPP goal is the creation of a functional data market within the Union by trying to grant it a relevant role in the management of large amounts of data in the global market.

Big data is a critical economic asset in the achievement of competitivity, growth and employment. It has a high impact potential and is a horizontal & sectoral profit facilitator. The ownership of value creation from big volumes of data will be the cornerstone of future development and social welfare.

Europe is not playing the role it should be playing in the global market. Only two in twenty of the top-notch companies which are actually changing lives and making money thanks to the efficient processing of big amounts of data are European. In order to revert this situation, Europe has to strengthen the complete data value chain, so it can present an efficient ecosystem of innovation in the big data segment and present business models accordingly. It implies the commitment of people and organizations in data management whatever their role: production, analysis or value and knowledge creation from the very data.

The creation of BDVA comes from the alignment of the Commission with the European industry (medium sized and big companies), research centers and universities into a public-private partnership to cooperate and research big data fields. The goal of such an organization is to build a prosper community around data analysis and exploitation. Contractual arrangement was signed on October 12, 2014 and is currently being applied in the 2016/2020 period through different European announcements for industry leadership as part of the

H2020 framework regulation through the following instruments:

- Wide scale Demonstrators or Lighthouse projects for industrial sectors in order for them to earn greater profits from big data exploitation via transport, bio-economy, agriculture, etc.
- Experimentation and integration actions of data or innovation spaces. Innovation spaces are environments where technology providers and final users meet to identify services, skills and business models.
- Technical project facilitators and those that imply an advance in knowledge acquisition and methodology usage.
- Networks, communities and support in the definition of policies in a way that will help to create an active union between main actors working with a common social and organizational purpose.

, Current calls for financing in the "Big Data and Open Data" portfolio, created by the European Union in the framework of this initiative are:

- ICT 14 2017. Data integration and intersectoral experimentation in several languages. The Big Data Inter-Platform Initiative brings together this wealth, due to the heterogeneous nature of the platforms involved, covering sectors such as Health, Industrial, Healthcare and new technologies.
- ICT 15 2017 Big pilots in sectors that have profited from data-oriented innovation. Create wide scale solutions to set new standards that can be easily duplicated.
- ICT 16 2017 Research oriented to technological challenges in data economy where partners with big datasets are rewarded.
- ICT 15 2017 Support, expertise in industry, benchmarking and evaluation. support CSA in community building and PPP and big data governance.

### *C. Other European initiatives*

The situation in the Member States of the European Union is very heterogeneous. While in countries like Germany, Austria and France there are programs with a relevant focus on data-associated technologies and a very high level of investment, other countries have barely begun to think about their transformation strategy towards a data-based economy; this is the case in many of the so-called Eastern countries. The Big Data PPPs aim to coordinate efforts made at European and national level to increase the potential impact of these initiatives and above all to align positions on key issues such as standardization or adoption of European regulations. The BDVA project contributes specifically to these objectives within the framework of this PPP.

The Big Data initiative created by PLANETIC is a good example of creating an ecosystem of innovation around data, with a clear base of the existing strengths and weaknesses in Spain. In other countries we can see initiatives focused on data or specifically on Big Data with different characteristics and nature. Here we review some of them (it is a reference list, not a thorough or complete list):

- **Smart Data Forum[8]:**Funded by the German Federal Ministry for Economic Affairs and Energy (BMWi),this initiative's main objectives are:: (1) the generation of an innovation ecosystem that integrates different actors, including Industry, Research Centers and politicians to promote research, innovation and education around data; (2) the alignment of different initiatives related to Big Data in the country, linking them with international activities and supporting the transfer of knowledge and good practices; and (3) to serve as a point of reference for successful cases in different sectors, as well as a point of contact for emerging companies in the data market. Smart Data Forum plays an important role in bringing together the different actors and initiatives focused on the data market in Germany with public funding programs to enhance projects of interest in the area. The following image shows some of these programs, essentially driven by the country, but also includes European funds such as the H2020 program.

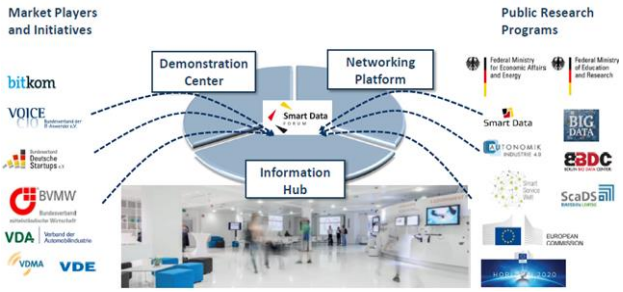



*Figure 2. Agents and roles involved in the Smart Data Forum initiative (source: SDF, Germany)*

- **The European Network of National Big Data Centers of Excellence[9]:** this initiative is promoted by Know-Center, Austria's leading research center in Big Data analytics

and data-driven businesses. Although promoted by Austria with support from the Ministry which has the data competencies, it is a joint initiative with the German government, and therefore has the political support of the Smart Data Forum and the corresponding Ministry. Its main objectives are to understand which the Big Data centers of excellence in Europe are and connect them according to different research topics, promote the transfer of knowledge from the University and Research centers to the Industry, and strengthen the European position in Big Data. It is therefore a more academic network, very orientated in promoting cooperation between leading centers in Big Data and establishing a bridge that allows for the research results of the joint teams to reach industrial applications. At the thematic level, the network includes areas such as data analytics, visualization and man-machine interaction, HW, systems and HPC, business models or education. The core team of the network is integrated by the following organizations: Know-Center, BBDC, INSIGHT, SICS, ADS and UniResearch, although it is also directly supported by the BDVe project, among others and already has more than 60 centers of excellence in 17 different countries.


*Figure 3. More than 60 Big Data Centers of Excellence from 17 countries already integrate the network promoted by Know-Center with the political support of the Austrian and German governments (source: Know-Center).*

- **Big Data Forum[10] (Finland):** this is a platform like BDVA, but its activities are limited to the country of reference. To be included in the forum, organizations must pay a fee that is different depending on the type of organization. Its main objectives are summarized as follows: generate knowledge and solutions in Big Data between Universities, Research Centers and the private industry; promote new businesses and markets where data is centerpiece, essentially at the intersection between Big Data and the so-called Industrial Internet; promote

collaboration among different actors to generate a system of innovation, and act as a vehicle towards other European initiatives in Big Data.

- **RISE Data Science:** this is the Swedish national initiative for Innovation focused on Data. Like the previous ones, it acts as a hub to integrate and combine the capacities and knowledge of the different leading entities of the country and encourage innovation projects with a greater impact on the country's different economic sectors. However, in this case we are talking about a single entity that brings together the main research and technology organizations in the country, such as SP Technical Research Institute of Sweden, Swedish ICT and Innventia. It has 2200 employees and has demonstration and experimentation centers open to other institutions and clients. In this regard, SICS ICE Data Center deserves a special mention. This is the Swedish offer in terms of Data Centers to carry out experiments and demonstrations. Due to its capabilities and excellence, it has recently received the Innovation Space label from BDVA.

*Figure 4. Coverage of the RISE initiative in Sweden*

- **Big Data Community of the Emilia Romagna Region[11] (Italy):** we conclude this preliminary list of European initiatives focused on Big Data with a regional initiative. The initiative began thanks toa survey on infrastructure providers which showed that 70% of all computing resources in Italy were found in the Emilia Romagna region (including HPC and HTC). Although initially academic, on February 7, 2016 it was also opened to the industry. The integration of the different actors is seen to be a necessary element to create an open science

platform that improves scientific knowledge and transfers it to the industry; creating a high impact on the competitiveness of companies and on the growth of the region. The initiative already involves almost 2000 researchers and includes a long list of working domains where a thorough analysis has been made to see what actors are involved, what projects have been carried out or are being implemented in the different areas, the associated infrastructure and available datasets. Among others, the list includes chapters such as: Big Data in ICT and Digital Content, Big Data in Life Sciences, Big Data in Agri-food and Bio-industry, Big Data in Transportation, Big Data in Materials, Big Data in Mechanics and industrial processing, Big Data in climate change and Big Data in environment and energy, among others. The region of Emilia Romagna plans to invest 7M Euro of regional funding precisely in Big data research infrastructures, advanced materials and genome.

As can be seen, in Europe there is a great number of different initiatives focused on Data and, specifically, in the challenges related to Big Data. In some cases, the initiatives are more academically oriented and in others they are more industrial, regional or national. But all these platformsl have one thing in common: the ambition to create an open innovation ecosystem to allow for Big Data knowledge creation and its transfer to industrial user cases in one or more sectors. Some involve educational activities, other do not. Some have public funds associated with their projects and others do not. In any case, even if this document only presents a small number of initiatives, we have confirmed the interest to establish relations with other initiatives at European level and promote collaboration between them.

## III. POSITIONING IN SPAIN

### *A. National level mapping of European strategies*

Objective 5 of the recently launched "Digital Agenda" promoted by the Ministry of Energy, Tourism and Digital Agenda, establishes the Promotion of the R & D system in Information and Communication Technologies. Investment in research, development and innovation in ICT and the application of the results achieved through it are key factors in improving the competitiveness of our companies and administrations. This is recognized by the European Union, and its Digital Agenda's goal for 2020 is to double the total annual public expenditure on research and development of ICTs, in such a way that an equivalent increase in private spending is also created.

In the case of the Digital Agenda for Spain, the main lines of action regarding ICT, are: to increase the efficiency of public investment in R &D; encourage private investment in R & D; promote R & D in small and medium sized companies; and expand the Spanish participation in R& D for ICT internationally. The initiatives proposed in this sense are aligned with the Spanish Strategy of Science and Technology and Innovation and highlights, among other elements, the challenge of moving towards a digital economy and society.

The Digital Agenda for Spain establishes the creation of a Development and Innovation Plan for the ICT sector. This plan seeks to focus resources on highly relevant digital initiatives such as cloud computing, smart cities or Big Data, and to make Spain an attractive destination for investments in future industries.

Regarding specific calls for Big Data national financing projects, these are being carried out gradually. Projects granted under the R & D line of finance from the Center for Industrial Technological Development (CDTI) are included in a a generic line of finance. CDTI does not establish specific topics such as Big Data, meaning that although there is no specific line of finance for a project of this nature, it can be awarded finance if it is positively evaluated by the Center.

Big Data has been included in other ministerial calls within the framework of the State Plan for Scientific-Technical Research and Innovation. This plan which concluded last year defined the Spanish Strategy in science, technology and innovation based on a series of key objectives. One of them, the so-called Economy and digital society challenge, established among its priorities the development, innovation and adoption of SOLUTIONS AND TECHNOLOGIES linked to: (i) cloud computing; (ii) Open / Linked / Big Data and the reuse of information from the public sector, generating value and knowledge.

During 2013-2016, MINECO's calls for the funding of "Challenges Collaboration" projects of, or the Economy and Digital Society Strategic Action (AEESD) were prioritized according to the guidelines established in the said State Plan. Below is a list of the Big Data projects that have been financed:

| BIG DATA | 2015 | 2016 |
|---|---|---|
| Requested projects | 67 | 48 |

| | | |
|---|---|---|
| Approved projects | 18 | 8 |
| Requested budget | 35,6M€ | 25M€ |
| Granted budget | 7,4M€ | 3,7M€ |

*Figure 5. Granted projects in AEESD call for 2015 and 2016. Source: Secretary of the State for Telecommunications and the Information Society and Digital Agenda.*

***B.***

To date, the agents committed with the initiative are the technological platforms that applied to join after a request for expressions of interest that was opened to all of them; Ministries that support the initiative to be aligned with its action plan (MINECO, MINETUR and MAGRAMA); and other organizations that are somehow involved in the Big Data domain, such as the Network of Excellence of Databases, the National Supercomputing Network, the CDTI or AENOR.

The initiative is open to more agents that agglutinate the interest of a sector or interest group, and that want to contribute to any of the objectives set forth in the introduction of this document.


*Figure 6. Agents and sectors in inter-platform.*

Participating technological platforms, and therefore, represented sectors are:

- Planetic (www.planetic.es)
- Food (http://foodforlife-spain.es/)
- Environmental technologies (http://www.pt-planeta.es/)
- Innovative medicines (http://www.medicamentos-innovadores.org/)
- Advanced Manufacturing (http://www.manufacturing-ket.com/)
- Logistic (http://www.logistop.org/)
- Health technologies (http://www.plataformatecnologiasanitaria.es/)
- Water (http://www.plataformaagua.org/)
- Multimedia and Digital Contents (AMETIC) (http://ametic.es/es/innovacion/plataformas-tecnologicas/enem)
- Internet of the Future (AMETIC) (http://ametic.es/es/innovacion/plataformas-tecnologicas/esinternet)
- Packaging (http://www.packnet.es/)
- Tourism (http://www.thinktur.org/)
- Roads (http://www.ptcarretera.es/)

***C. Supply and demand map of the sectors represented***

As mentioned at the beginning of this publication, different sectors are represented in the inter-platform Big Data initiative coordinated by PLANETIC. These sectors are represented bythe different platforms that are included in the initiative. Wecan propose an analysis in terms of supply and demand with Big Data technology as the denominator, both from the perspective of data and from the technology itself.

**Water:** The Water Technology Platform provides a pool of study indicators for the preparation of its supply and demand interests, both at data and technological levels.

- **Data:** it provides data related to use, management and planning of the distribution network itself. Regarding demand, it focuses on the observation of the Earth and on data produced by sensors in the distribution network.
- **Technology:** It has Information Systems for water management, flow sensors, hardware for real-time management and management of social networks. It currently offers a Big Data platform for drinking water supply networks. The Water Platform already has Big Data pilot projects and has shown its interest in large pilot projects demonstrating the integration of its own management tools with Big Data technology. It has a great potential for the deployment of Big Data experiments and the application of analytical techniques and deep learning in the governance systems of water networks.

**Health:** Represented in PLANETIC by the Platform of Innovative Medicines, it establishes what it can contribute and

its functional requirements in the future in the following terms:

- **Data:** It offers two groups of data depending on their level of privacy; a first group includes pharmacological (indications, treatments, side effects), genetic, clinical and laboratory data, with access restrictions, and a second group includes public databases usually of aggregated data and biomedical publications of free access. In terms of demand, clinical data for research, access to sources of pharmacological and toxicological information.
- **Technology:** They have filtering, data mining and machine learning tools, as well as software for the development of predictive models. These already include the management of large volumes of information. The requirement of data management doesn't mean that the computational spectrum already being used will be set aside **Industry / Manufacturing:** this sector is represented by ManuKET, an Advanced Manufacturing Platform.
- **Data:** the data used is related to the monitoring of machines and is provided by a number of different sources such as cameras, sensors, databases, OPC or file systems, in addition **to the social and financia**l logistics associated with their customers. Data on the maintenance of the production chain in real time or composition of materials would be in the group of Platform.
- **Technology:** Great set of offered systems, from those related to industrial process sensors to social network management, through data analysis and correlation tools, as well as a laboratory called Emulab. With regard to demand, we already have the right data exploitation tools.

**ICT:** Industry / Manufacturing: is represented by ManuKET, the Advanced Manufacturing Platform:

- **Data:** the data is related to the monitoring of machines, provided by multiple sources such as cameras, sensors, databases, OPC or file systems, in addition to the social and financial logistics associated with their customers. Data on the maintenance of the production chain in real time or composition of materials would be in the group of Platform.
- **Technology:** Great set of offered systems, from those related to industrial process sensors to social network management, through to data analysis and correlation tools, as well as a laboratory called Emulab. Regarding demand, we already have the right data exploitation tools.

**Tourism:** The Tourism Technology Platform THINKTUR is an organization that brings together tourism companies and technology related tourism companies, interested in improving the sector's competitiveness. Big data is a relevant element in the digital transformation of the sector, from these two points of view:

- **Data:** The generation of data is high because it is provided by different sources, such as tourists' social networks, the opinions of travelers added to different portals, the public tourist information available in destinations, the sensorics available in multiple types of devices, or the inventory, prices and availabilities of the tourist offer, among others. In terms of demand, customer knowledge, prediction of demand and prices and the carrying capacity of different destinations, are aspects where data provision is necessary for agents with valuable information to make decisions.
- **Technology:** It includes Information Systems for tourism management in destinations, tourism companies' internal management systems, for operational issues such as marketing and promotion, social network management and the management of revenue management, among others. Currently there are several initiatives at regional and local level under the umbrella of Intelligent Tourist Destinations. These initiatives are being implemented by the public authorities. The purchase of tools that help companies to know more about their clients and optimize operational management is becoming more and more frequent in the private sector. The Technological Platform of Tourism shows its interest in pilot projects demonstrating the extraction of information from various sources, as well as the treatment and visualization of said data. This makes it possible for them to obtain information with an added value, from Big Data to Smart Data. Several pilot initiatives are already being developed, many of them within the framework of smart tourism destinations and they are focused on the areas of marketing and promotion, customer knowledge, demand prediction, and waste management.

## IV. CURRENT INNOVATION POLES

Research in the field of advanced Big Data[12]analytics is essential to help organizations consolidate the technological

exploitation obtained from the finding of patterns, trends and useful information that can be used to prepare reliable diagnoses and forecasts in many different areas.

Companies that fully exploit the capacity of the data are more prepared for future planning, they strengthen their skills for the discovery of knowledge and the prediction of behavior in scenarios of uncertainty. The extraction of knowledge from databases, continues to extend to almost all areas where there is processing of large volumes of data (system logs, large databases, activity logs, etc.).

In a MCKinsey[13]market study, the potential of Big Data was shown in figures. For example, the retail industry, using all the capacity offered by Big Data, could increase its profits by 60%. Also, European governments could save more than 100 billion Euros in improvements According to the Chief Marketing Officer of EMC in a study carried out on the digital universe of 2014 supports the conversion of organizations into software-defined companies: “IT must press the reset button and look for new ways to optimize storage and take advantage of these masses of data. "

Although some Big Data technologies are old, such as artificial intelligence or expert systems, their development as an integral discipline is relatively recent. This means that there are limited explicit references in sectoral strategies and Platforms, except those that have been published in recent months. But, the movement that is taking place around these technologies in Europe and around the world is impressive, given the power they have to transform the conception and exploitation of existing businesses in practically all sectors. The digitization of business and the instrumentation of the world (sensors, actuators, Internet of Things, cyber physical systems), the increasing availability of all kinds of data and the ability to intelligently analyze these amounts of data is a decisive vector to improve processes, products and solutions of all kinds as well as a source of new business. The facilitators are, therefore:

- **Scientific and technological maturity**. The volumes, speed and variety of data grow exponentially and, in parallel, there are new systems and technology to process them. Technological evolution seeks to treat and follow the exponential rhythm of this huge amount of data, more quickly, from more diverse sources and more efficiently.
- **Maturity of the systems.** The development of data storage, processing and communication systems continues to expand exponentially from the point of view of power and cost reduction; the new XaaS model also appears and provides almost any organization with the benefits of the new systems. And this also happens for compact and embedded systems.
- **Maturity of the market.** Although incipient, the digital mentality is already being established in many different segments; there is a shortage of talent and from the market point of view, the organizational priority is mainly focused on efficiency and cost reduction. An although always with the need to find new channels of income.
- **Social maturity Growing** concern about the use of data and privacy and property issues; demand for open data; partial legislative vacuum. These are issues that should be considered from these multiple perspectives.

### *A. Sectoral analysis*

To analyze the sensitivity to the incorporation of Big Data by sectors, four dimensions are considered:

- The existence and availability of data
- The existence and availability of talent and qualified people who work with the data
- The existence of infrastructures and technology to work with data, even in compact and embedded systems
- The existence of a specific organization within the organization that takes advantage of the Big Data asset

A sectoral analysis of these four dimensions can provide an idea of the sectoral sensitivity of Big Data.Although, obviously, case by case and company by company there may be singularities and important differences:

| SEGMENTS | Data availability | Access and talent retention | ICT infrastructure implantation | Organizational maturity |
|---|---|---|---|---|
| INDUSTRY | «« | «« | « | « |
| HEALTH | «« | « | « | « |
| ENERGY (UTILITIES) | «« | «« | «« | « |
| WATER (PUBLIC BODIES and UTILITIES) | «« | « | « | « |
| ADMINISTRATION | « | « | « | « |
| MOBILITY | « | « | « | « |
| FINANCIAL AND INSUREMENT. | «« | « | «« | «« |
| COMMERCIAL | « | « | « | « |
| MEDIA | « | « | « | « |
| TELECOMM. | « | « | «« | «« |
| TOURISM | «« | « | « | « |

In terms of potential, although the application of Big Data is enormously transversal, some institutions and analysts consider that there are sectors that are better positioned to

obtain short-term profits through Big Data. In any case, a detailed analysis must combine the potential with the feasibility that is determined by the sensitivity of the business to take advantage of the potential of Big Data.

The following tables are part of an analysis of Big Data opportunity areas for certain productive sectors. The examples below allow us to see the economic impact potential of the Big Data disciplines in almost all sectors of economic activity.

**–The Machine-Tool Sector.** The Machine-Tool sector is in early stages of adopting Big Data in fields such as:

| OPPORTUNITIES | USAGE |
|---|---|
| -Manufactured in series, flexibility in manufacturing<br>-Energy efficiency<br>- Digital Cybersecurity Protection | - Optimization of the storage processes.<br>- Predictive maintenance of machines.<br>- Optimal allocation of production resources.<br>- Prediction of failures in production lines.<br>- Logistic optimization of supply and distribution.<br>- Preview of the demand.<br>- Energy efficiency in production.<br>- Explicit deployment of goods and production chains (layouts).<br>- Security in production lines. |

**The automotive sector.** The applications within the vehicle indsutry are beginning to appear (consumption optimization, sensory of all types, adaptation to change, protection of cyber-attacks, etc.).It is an area that will continue to grow as vehicles are beginning to relate to each other and with infrastructure, or progress in autonomous driving. In the aeronautical sector, world-leading companies such as GENERAL ELECTRIC are starting to exploit Big Data as a source of new business growth. For companies in the sector, Big Data could begin to be adopted in areas such as:

| OPPORTUNITIES | USAGE |
|---|---|
| -Manufactured in series, flexibility in manufacturing<br>-Energy efficiency<br>- Digital Cybersecurity Protection | Optimization logistics and storage processes.<br>Predictive maintenance of machinery.<br>Planning and optimal allocation of production resources.<br>Prediction of failures in production lines.<br>Logistic optimization of supply and distribution.<br>Demand review.<br>Energy efficiency in production.<br>Explicit deployment of goods and production chains (layouts).<br>Security in the plant<br>Analysis of risks and prediction of failures |

The Aeronautical sector presents a certain similar behavior in terms of short-term demands.

**The Energy Sector.** The energy sector is one of the largest number of users of Data Analytics application in fields such as:

| OPPORTUNITIES | USAGE |
|---|---|
| **Renewable energies** | |
| -Reduction of greenhouse gases<br>-Decentralization and fragmentation of the generation<br>-Advanced management systems<br>-Reduction of the cost of energy<br>-Preparation of production | Prediction of renewable energy production: Wind, hydraulic and solar.<br>Explicit deployment of infrastructures for the collection of renewable energy.<br>Predictive maintenance of infrastructures.<br>Planning of dispatch and storage of renewable energy produced, based on economic criteria. (Micro generation by domestic users, for example solar panels in private homes). |

| | |
|---|---|
| | Structural design of the elements that make up the infrastructures for collecting renewable energy (Example: design of the blades of a wind turbine).<br>Advanced monitoring of infrastructures |
| **Energetic efficiency in buildings** | |
| -Rules and regulations in energy efficiency<br>-Efficient and responsible consumption<br>- A more delegated management towards users<br>-Request of demand | Characterization of energy consumption of architecturally complex buildings.<br>Minimization of the building's energy consumption, comfort criteria.<br>Optimization of production plans with energy efficiency criteria for industrial plants<br>Efficient use and dispatch of energy in energy-producing buildings<br>Predictive maintenance of air conditioning systems<br>Prediction of failures in air conditioning systems |
| **Smart Grids** | |
| -Predictive Maintenance<br>-Custom attention<br>-New storage capacities<br>-Improvement of the operation of the network | Exploitation of information from Smart Meters for:<br>Detection of non-technical loss<br>Predictive characterization of subscribers' energy consumption<br>Prediction of network failures<br>Balance in demand<br>Transfer of micromarketing strategies for users<br>Prediction of the demand and therefore possible saturation of the network.<br>Intelligent treatment of infrastructure operation data for the detection of faults.<br>Exploitation of the data provided bythe distribution network equipment |
| **Gas** | |
| -Optimization of the Infrastructure Network | Prediction of incidents in the network, leaks in the distribution circuit.<br>Fraud detection.<br>Explicit deployment of distribution infrastructures.<br>Prospection of natural gas deposits. |

**The Water Sector.** The domain of the water cycle in this sector has common elements with some segments of the energy domain, as opposed to other complementary ones related to the natural environment or to the singularity of the facilities for the supply of drinking water, sanitation or irrigation.

| OPPORTUNITIES | USAGE |
|---|---|
| -Predictive maintenance<br>-Customer Support<br>-New storage capacities<br>-Improvement of the operation of the network<br>-Optimization of the Infrastructure Network<br>- Improved management against extreme weather events<br>- Improvement of pollution control in the medium | Exploitation of information from Smart Meters for:<br>Detection of nontechnical losses<br>Predictive characterization of the consumption at the subscriber level and prediction of the demand and therefore possible saturations in the network.<br>Balance in demand<br>Transfer of micromarketing strategies at user level<br>Prediction of incidents in the network, leaks in the distribution network<br>Intelligent treatment of infrastructure operation data for the detection of faults<br>Exploitation of the data coming from the equipment of the distribution network<br>Management and maintenance of assets |

| | |
|---|---|
| | Advanced monitoring of infrastructures and aquifer reserves<br><br>Advanced calculations for water quality simulation in watersheds and reservoirs<br><br>Simulation of diffusion of pollutants in basins and reservoirs<br><br>Simulation of the behavior of the receiving environment before actions for the mitigation of risks derived from climate change. Water quality control. |

**The Health Sector.** The areas related to health, both in the field of public health and the prevention of diseases and in the clinic, and in Biomedicine, including genomic information or the development of new drugs, are generating an increasing volume of data and present extraordinary growth and enormous potential for the application of Big Data technologies:

| OPPORTUNITIES | USAGE |
|---|---|
| - Comparative analysis of treatments and therapies<br><br>-Precise and personalized medicine<br><br>-Support systems for clinical decisions<br><br>- Automation of repetitive processes<br><br>- Development of new medicines<br><br>- Improvement in the quality of health data<br><br>- Integration of information from wearables and health apps for the monitoring of biological data and diseases<br><br>- Enable the integration and analysis of very diverse and / or isolated health information sources | Better understanding and monitoring of chronic diseases<br><br>Reuse of electronic medical record information for research<br><br>Support for the improvement of lifestyles and promotion of healthy life. Analysis of images for the discovery of similar pathologies.<br><br>Optimal deployment of health centers<br><br>Improve user satisfaction through social media analysis<br><br>Personalized medical care based on the user's medical history, with criteria of efficiency in care and economic costs.<br><br>Multimodal prediction of medical problems<br><br>Segmentation and stratification of patient profiles for a more personalized care<br><br>Optimal planning of medical personnel in health centers (both in hospitals and primary care units)<br><br>Optimal and predictive management of priorities in clinical emergencies<br><br>Personalization of preventive activities according to lifestyles, clinical data and genetic information available<br><br>Management and planning of preventive activities and health crises in the field of Public Health<br><br>Reduction in the cost and research of new drugs<br><br>Reuse of existing drugs for new indications<br><br>Detection of unsuspected adverse effects with the analysis of large volumes of pharmacological data |

**The Digital Sector:** As well as holding a “producer” role related to Big Data systems and technologies, the Digital Sector was one of the first sectors to demand been data solutions.

| OPPORTUNITIES | USAGE |
|---|---|
| - Marketing micro personalized and contextual<br><br>- Customization of customer use<br><br>-Integration of services | Reduction of customer leakage (logistics)<br><br>Advanced management of call centers: identification of problems in critical time, maximization of return of the clients and efficiency in the conservation of clients.<br><br>Analytics for networks: identification of bottlenecks based on logs, prediction of capacity and demand for optimal sizing of networks, downloading of cellular traffic to opportunistic networks.<br><br>Explicit deployment of network infrastructure with economic criteria, capacity, visual impact ... |

| | |
|---|---|
| | Predictive analysis of levels of occupation of network resources (spectral).<br><br>Advanced value services based on location information estimated by telecommunication networks (useful in the transport and commerce domain).<br><br>Analysis of network content and user profiles (Deep packed analyses) |

**The Mobility and Logistics Sector.** In addition to private use, the mobility and logistics sector has many applications in the public domain. The development of IoT, ITS and other intelligent technologies is exponentially increasing the volume of data used in the logistics and mobility sector. Their application in the operation and in the creation of new business models is of extreme interest:

| OPPORTUNITIES | USAGE |
|---|---|
| -Digital transformations<br><br>-Optimization of origin/destination matrix<br><br>-Combination of data in real time, dates, accidents<br><br>-Electric cars<br><br>-Efficient deployment of electric mobility<br><br>- Implementation of the Physical Internet concept (Physical Internet)<br><br>- Creation of new interconnected logistics operators<br><br>- Increase the visibility of the supply chain<br><br>- Increase the load factors<br><br>- Reduction of the carbon footprint of logistics operators<br><br>- Increase transport safety<br><br>- Development of automatic logistics operations<br><br>- Boost autonomous transport / self-driving (trucks, boats and planes)<br><br>- Integration of cargo and logistics units in automatic data exchanges | Prediction and mapping of optimal routes based on information about incidents, weather and environmental pollution criteria.<br><br>Multimodal transportation.<br><br>Aggregation of urban logistics flows for aggregate analysis in order to plan the logistics of the city.<br><br>Explicit deployment of electric car infrastructures.<br><br>Dissuasive parking.<br><br>Design of emergency plans before eventualities on the road network.<br><br>Detection of points of interest according to the routes of different users.<br><br>Predictive maintenance of road infrastructures.<br><br>Management of electric vehicle charging and roaming cargo managers.<br><br>Shared transport applications.<br><br>Optimization of logistics distribution plans.<br><br>Traced geo-located and dynamic isochrones.<br><br>Electric Urban Transportation.<br><br>Change of speakers: logistics operators and zones.<br><br>Management of intelligent and automatic warehouses.<br><br>Intelligent load management, especially perishable products.<br><br>Creation of open digital platforms that optimally integrate the data of interconnected logistics in collaborative governance.<br><br>Ability to readjust logistic processes in real time. |

**The Packaging Sector.** The practical application of Big Data criteria in the packaging sector by means of connectivity through intelligent sensors undoubtedly provides higher quotas of quality, efficiency, security, in addition to contributing to the cost optimization of the entire process.

| OPPORTUNITIES | USAGE |
|---|---|
| - “Smart Packaging”<br><br>- IoT: Internet of Things | The interaction of technological devices with active and intelligent packaging provides a great deal of information about the product: freshness, nutritional value, etc.<br><br>In pharmaceutical packaging, it can also provide information on composition, dosage, contraindications, mode of preservation, etc.<br><br>Through various technologies (RFID, NFC, Bluetooth, smart labeling, etc.) included in the packaging you can obtain more refined information regarding the traceability, which allows companies to make better decisions regarding the "time to |

| | |
|---|---|
| | value "<br><br>In the same way, the insertion of intelligent sensors either in packaging machines or in complete packaging production lines will help in the proves perfecting the automation of the process. |

**The Tourism Sector.** The interest in the tourism sector is based on an intelligent management of the data that makes it possible to obtain relevant information for the launching of strategic data. For this we have relied on the "Big data: challenges and opportunities for tourism" study by INVATTUR and Creative Territory:

| OPPORTUNITIES | USAGE |
|---|---|
| - Smart Data<br><br>- IoT: Internet of Things<br><br>- Optimization of processes<br><br>- Customer knowledge<br><br>- Demand and market analysis<br><br>- Predictive Pricing<br><br>- Customer modeling<br><br>- Assessment and creation of new products<br><br>- Energy efficiency<br><br>- Sustainability<br><br>- Water management<br><br>- Waste management<br><br>- Intermodality<br><br>- Cybersecurity | Find patterns of behavior and consumption trends.<br><br>Anticipate the behavior of the market and customers and respond proactively.<br><br>Improvement in the management of the information generated by tourists in the destination with the intention of creating an offer of comprehensive leisure services<br><br>Elaborate market segments and individualized studies to determine actions<br><br>Greater knowledge of the tastes and preferences of different segments and customer profiles<br><br>Maximization in the personalization of experiences and tourist offers adapted to the needs of the traveler<br><br>Discover and measure the influence of customers and non-customers<br><br>Identification of the variation of supply and demand, establishing a warning system<br><br>Through different technologies (RFID, NFC, Bluetooth, WIFI, etc. included in the packaging you can obtain more refined information regarding traceability, which allows companies to make better decisions regarding the "time to value".<br><br>Optimization of operational management processes<br><br>Attraction and customer loyalty through data mining techniques, machine learning and natural language processing, to satisfy, attract and retain customers<br><br>Optimization of distribution and transport routes<br><br>Monitoring and management of energy consumption<br><br>Efficient and coordinated management of public and private resources |

***B. Analysis of the evolution of Big Data***

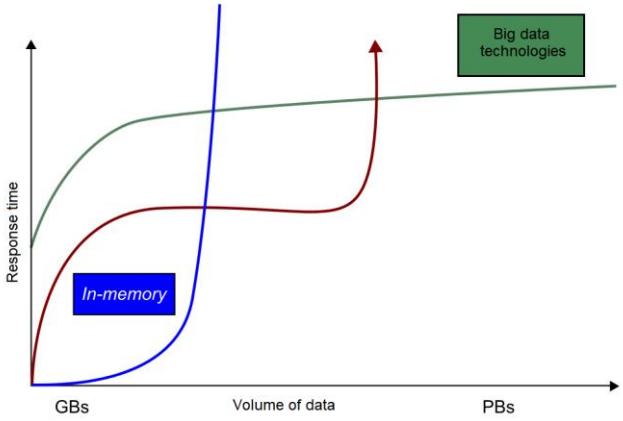
Size, formats complexity and delivery speed are overcoming capacities of traditional data management technologies. So, the new use of ICT to manage huge volumes of data is mandatory. New technologies that make a great impact will emerge, such as in-memory database managers to attend new requirements of data.

In any case, the future will surely be hybrid; it will not only include a storage technology, but it will be a mixture of all of them. It is important not to forget that not all applications have their bottle neck alone in the data. On the other hand, even though storage costs in memory are cheaper, it will not always be profitable to store certain data at higher costs. Therefore, one of the important research topics is to consider the global management of storage resources inside the CPD, considering the requirements of the applications and at the same time deciding where they are executed and where their data can be found. In the following image, the behavior of the new proposals is represented schematically compared to the traditional information storage systems that we have presented before.

A very important fact is that these new technologies are available to everyone and not just corporations that can make large investments. In the same way that webs evolved and Web 2.0's were created, thanks to open source software packages such as LAMP (Linux, Apache, MySQL and PHP),

now we are entering a new era for data: Data 3.0 and this is thanks to a new generation of technologies and architectures, designed to extract value from great amounts of different tyrpes of data and in some cases, event in real time.

The following image shows some of the possible technologies that are available to everyone to enter this world of Big Data.




When it made its first appearance, Big Data emerged in its incarnation 1.0 as a technology-oriented event. Big Data research focused mainly on the technological requirements that companies needed if they wished to correctly process the large amounts of data available. Some of the main problems addressed were (i) **the management of the volume and scalability** of Big Data sources with the paradigms of new databases such as Hadoop, NoSQL or the Map / Reduce the environment through the use of scalable[15] [16], **(ii)** architectures (ii) **the aggregation** of Big Data sources focusing on the analysis of the transactions generated by these large data sources. Due to its difficulty, efforts are currently still being made focused on improving the data patterns that currently exist and are being achieved in different domains and (iii) **analyze the data** and learn from the data with a strong belief in the resolution of problems driven by the data. The analysis carried out in this first stage focused on the multidimensional analysis of the sources -type Data Warehouse-, the visualization tools and, statistical and data mining techniques applied to large volumes of data.

In a second phase (which could be called Big Data 2.0) and due to the difficulty of extracting real knowledge from the different domains, many investigations were conducted focusing on (i) vertical applications that would develop techniques specially adapted to the problems under the data of the domain. Some of the main domains addressed were advertising, social media (with strong integration of feelings analysis in social networks or blogs), retailers, financial services, telecommunications or health, among others; (ii) aggregation of data for vertical applications previously introduced; and (iii) the interoperability of new Big Data solutions with the already existing Business Intelligence applications (Business Intelligence), since most of the Big Data solutions were not integrated with these BI applications. At this stage many issues are still left open for the next stage. For example, even though considerable progress has been made in the analysis of feelings when reading tweets, the results are still far from automating this analysis since the 2.0 paradigm only takes into account the textual interaction of the user, and ignores variables such as intonation, cadence or vocal tone, the facial expressions, etc. This leaves the field open to the next generation Big Data 3.0.

Nowadays, in the third and what would be the last stage for now (Big Data 3.0), a lot of effort has gone into the scalability of the solutions in the sense that, the design of Big Data solutions follows the same traditional paradigm of BI, being inter-sectorial applications. Therefore, the main research objectives of this third era are: (i) to develop a top-down analysis guided by the objectives and real information needs, (ii) the integration of data from a large number of sources of information. Large heterogeneous data (for example, audio, text, video, sensors, mobile devices), ETL (Extraction-Transform-Load) processes already developed for BI applications and adapted to the different needs of the domains, and finally (iii) Research focused on Smart Data instead of Big Data, paying close attention to the quality, origin, explanation, trust and semantics of Big Data.

Reaching this last paradigm is not an easy task.It requires a platform that allows from the collection of multidimensional information established in Data 1.0, and an accurate interpretation of human feelings contained in the various forms of interaction that this may have. This platform will give the sellers the possibility of getting feedback from users almost in real time thanks to all the interaction media added in Data 3.0. All of this keeping it in an exploitable database of considerable dimensions at the same time. This is the Big Data 3.0 promise.

Note that in this stage of Big Data 3.0, instead of collecting all the sources of Big Data and processing them, special emphasis is going into identifying the value (focusing on this era in the fifth V of Big Data) of the sources being they are processed. So, there is a strong commitment to specify the Big Data sources' Key Performance Indicators (KPIs) and to identify how these can complement the current internal KPIs to improve strategic decision making. When this is achieved we will be able to say that Big Data will be providing value and improving strategic decisions.

### *C. Prospects*

We are currently in a time of growth in the use and implementation of sophisticated software/hardware tools. These are used to analyse, process and extract large volumes of data generated on a daily basis. We must incorporate new concepts such as Zettabytes or Yottabytes. Nowadays, the digital data stored all around the world would is equivalent to the memory size of 10.000 million personal computer and this amount that will double in the next 18 months, and so on every year and a half.

It is important to emphasize the current capacity to detect the added value of Big Data sources and therefore, the capacity for the early detection of that value and its incorporation into the KPIs of current organizations for the strategic decision-making process. This is one of the main challenges for researchers and organizations with regard to Big Data.

The main problems related to Big Data according to a study by the consulting firm Gartner[17]are: (i) **How to extract the value of the data**, (ii) **The definition of a Big Data strategy,** (iii**) Obtaining the necessary skills for its processing and analysis,** (iv) **The integration of multiple data sources** and (v) **The poor quality of the data**. Curiously, the data volumes do not appear to be a serious problem because a series of solutions (discussed previously in the Big Data 1.0 Era that deals with this matter) already exists. Instead, the main concern is how to successfully combine the tools available to integrate and extract value from Big Data.

Another aspect summarized in this report refers to how the growth of organizations that implement Big Data projects has been a lot less in the 2013-2016 period and that the number of new projects seems to be stagnate.

In this regard and as described in the sectoral analysis mentioned above, each sector and domain has its own idiosyncrasy and, therefore, we are in the era of providing vertical solutions dedicated to each one of them. Another important problem in the current market of Big Data applications is that even today there are no specific guides and methodologies that facilitate the application and cohesion in different sectors. This fact represents an impediment for organizations that wish to undertake a Big Data project and do not have the knowledge or training to do so.

Sergio Martínez-Cava, CEO of Innovation Bankinter Foundation, states that the emergence of tools for processing, interpreting and integrating this huge amount of data will revolutionize our day to day lives. The report by Innovation Bankinter Foundation "Big Data: The power of Data", contains the keys and challenges of a phenomenon that is set to transform our daily life and to revolutionize numerous sectors of society in the years to come, from the economic sector and the public administration to health or education.

Companies and governments have just begun to understand the possibilities of Big Data technology. According to the experts at the Innovation Bankinter Foundation, this technology offers a world of opportunities to increasing the profitability and operating efficiency of organizations. It will increase revenues through the personalized services, adapted to the characteristics and conditions of each consumer, the retention of the best customers, the definition of new products, the identification of new opportunities in the market, and even the conversion of customers into proactive sales agents. It also allows for the reduction of costs, through the optimization of supply channels, the limitation of really relevant communications, fraud detection and dimensioning of trading platforms.

The effective analysis of the data will also allow for the identification of new market opportunities, accelerating the launch of new products, modifying prices, accurately assessing the return on investments, anticipating customer willingness to purchase another product or choosing the most appropriate time to make an offer. "It will allow for the generation of patterns of behavior and for the possibility of predicting habits. This will be done not only for commercial use, but also to combat crime, increase the intelligence of cities to the point of being able to self-manage themselves through algorithms or improve aspects of human life such as health, personal performance, sports, work, etc. This is just the beginning. Big Data has opened the closure to a more informed society and to a society that is more efficient and capable of performing feats that up until now were reserved to the field of science fiction.

Among others, the report highlights the future impact of Big Data on the functioning of cities. Technology accelerates the ability of governments to listen and resolve its citizens' problems a lot quicker.

This report also highlights the importance of Big Data on Health. The Healthcare System is another great actor that is feeling the effects of Big data. Relationships between patients, caregivers and health professionals, inside and outside the health system is changing and new rules and methods are being adopted. Self-care emerges as a phenomenon, with new applications and devices that provide valuable information about our health and day to day habits.

In addition, when looking at the corporate world, Big Data has filled a gap and has become the incentive to move forward and renovate old hierarchical structures.

New forms of organizations and relationships between managers and employees will emerge, collaborating to improve labor and business efficiency.

Cybersecurity will lead to new ways of fighting crime and ensuring systems. As in sports, Big Data is changing the role of trainers and is contributing to the improvement of the physical performance of athletes.

Experts from the Foundation warn that Big Data still faces a great number of challenges before it is fully implemented and before its "golden age" is here. There are also many obstacles to overcome including legal issues, human, market, format, access, technical, human skills and resources that facilitate the development of this technology. The authors of the report warn that these challenges, , will make it necessary to rethink protection and privacy formulas to protect identities and to avoid identity theft; create new educational disciplines for the massive treatment of data; develop new operational levels for the self-management of smart cities; adopsensors' technology that will monitor all our movements and create new quantum computers that process and store the data, etc.

In March 2016, FORBES tech publication released an article[19] with a pool of predictions about Big Data future:

- Data is growing more and more all the time. There is absolutely no question that we will continue generating larger and larger volumes of data, especially considering that the number of handheld devices and Internet-connected devices is expected to grow exponentially.
- While SQL continues being the standard language, tools to analyse data such as Spark are emerging as complementary systems for analysis and will continue to grow, according to Ovum[20.]
- Tools for analysis (without analysts) will emerge. Microsoft MSFT and Salesforce both recently announced features to let non-coders create apps to view business data.
- *IDC*[21]predicts that half of all business analytics software will include the intelligence where it is needed by 2020. In this way, prescriptive analytics will be built into business analytics software.
- Real-time streaming insights into data will be the hallmarks of data winners going forward, according to Forrester.
- [22]. Users will want to be able to use data to make decisions in real time with programs like Kafka and Spark.
- Gartner[23]considers that machine learning is a top strategic trend. And Ovum predicts that machine learning will be a necessary element for data preparation and predictive analysis in businesses moving forward.
- Big data will face huge challenges relating to privacy, especially in terms of the EU's new privacy regulation. Companies will be forced to address the 'elephant in the room' around their privacy management and procedures. Gartner[24] predicts that by 2018, 50% of business ethics' violations will be related to data.
- Forrester predicts that the number of CDO's (Chief Data Officer) will see a rise in the near future and that this position will be created in more and more companies. But certain types of businesses will not see the need to cover this position and generational differences may also be a deciding factor for the creation of this position in certain companies.
- According to Gartner, autonomous agents and things will continue to be a huge trend, including robots, autonomous vehicles, virtual assistants and smart advisers.
- According to IDC the shortage of big data profiles will be managed by analysts and scientist, but it will also be necessary to hire architects and experts in data management.
- The fact is that the big data talent crunch may s as companies employ new tactics. The International Institute for Analytics[25]predicts that companies will use recruiting and internal training to resolve problems concerning staff.
- DaaS or the data-as-a-service business model is on the horizon. Forrester suggests that after IBM, IBM +0.59%'s acquisition of The Weather Channel, more businesses will attempt to monetize their data.
- Another emerging market will be that of algorithms. Forrester surmises that businesses will quickly learn that they can purchase algorithms rather than program them and add their own data. Existing services like Algorithmia, Data Xu, and Kaggle can be expected to grow and expand.
- Cognitive technology will be the new buzzword. For many businesses, the link between cognitive computing and analytics will become synonymous in much the same way that businesses now see similarities between analytics and big data.

- "All companies are now data businesses," according to Forrester. More companies will attempt to drive value and revenue from their data.
- By 2020, organizations using data will see $430 billion in productivity benefits over their competition that is not using data, according to International Institute for Analytics.
- The Big data concept will soon be replaced by "fast" or "actionable" data, according to some experts. The argument is that big isn't necessarily better when it comes to data, and that businesses don't use a fraction of the data they have access too. Instead, the idea suggests that companies should focus on asking the right questions and making use of the data they have — big or otherwise.

Thus, we can conclude that only time will tell which of these predictions will come to pass and which will merely pass into obscurity. But the important thing to takeaway is that Big data is only going to get bigger, and companies that ignore it will fall further and further behind.

## V. BIG DATA PILOT PROJECTS

### *A. Projects financed at European level*

Including 2014, 2015 and 2016 calls. On the PLANETIC site there is an updated list of Big Data projects: http://planetic.es/content/proyectos-big-data

### *B. Nationally funded projects*

In the 2011-2016 period, CDTI approved 94 Big Data related projects., Nineteen of these projects were implemented in conjunction with CIEN or INNTERCONECTA calls:

| Call | Projects | Total budget |
|---|---|---|
| **CIEN (Cooperation)** | 4 | 34 685 367 € |
| **ID (Individual)** | 67 | 44 492 734 € |
| **ITC ("Innterconecta", coop)** | 15 | 31 092 566 € |
| **LIC (Innovation)** | 2 | 3 428 785 € |
| **NEO (Technological Base)** | 6 | 2 140 406 € |
| **Total** | **94** | **115 839 858 €** |

| Year | Projects | Total budget |
|---|---|---|
| **2011** | 3 | 5 420 483 € |
| **2013** | 10 | 8 483 353 € |
| **2014** | 11 | 17 556 116 € |
| **2015** | 39 | 51 998 693 € |
| **2016** | 31 | 32 381 213 € |
| **Total** | **94** | **115 839 858 €** |

Titles of the most relevant projects

- **'E-TUR2020, TURISMO & RETAIL'**. The Big Data Platform offers precision marketing solutions for the intelligent processing of information, in the Tourism sector particularly. The goal is to reach people at the right moment offering nearby services and product recommendations depending on their preferences, profiles and daily activities.
- **'LPS-BIGGER:** Lines of Software Product for Big Data through innovative applications applying to the real environment'.
- **'PRODUCTIO**: Productivity industrial enhancement through enabling technologies.

Other projects and initiatives:

- **Alianza BID3A - Big Data para RIS3. The** BID3A initiative has several projects co-financed by the ELKARTEK Program offered by the Basque Regional Government. This programis focused on the generation and validation of Big Data technologies that can be applied to the priorities of the Basque Country's RIS3 Strategy: Advanced Manufacturing, Energy and Health/Biosciences. The BID3 alliance is composed of the Technology Centre TECNALIA, several technological centers form IK4 (a private and independent alliance of R&D centers) like VICOMTECH-IK4, TEKNIKER-IK4, IKERLAN-IK4, the Ibermatica Innovation Institute I3B, the Basque Center of Applied Mathematics, the

University of the Basque Country, Mondragon University and Deusto University.

- **WATERSIG. This is an** application of Information Technologies to integrate the management of water and sanitation through Big Data and parallelizing. The general project objective is the design, development and experimental validation of a precommercial information system to facilitate the management of water supplies and sanitation network infrastructures. On the one hand technological opportunities related to Big Data technologies and to software acceleration techniques are taken into consideration and on the other hand, all kinds of restrictions related to the commercial exploitation of results are also taken into consideration. The project is aligned with the Regional Government of Galicia's RIS3 Strategy and is funded through the CONECTAPEME regional program. The project is being implemented by CYE-PROYFE, Torusware and ITG Technological Centre and Ingegnerie Toscane and Klink S.R (Italian companies) are also collaborating in the project.
- **WaMaS – IoT**. Water Infrastructures Maintenance Information Support Decision System based on the data added value in a well-balanced SQL No SQL Architecture. The project wasled by the SME Computer3 S.L.and was supported scientifically by the Instituto Teccnologico de Galicia. WaMaS –IoT was funded by the Centre for the Development of Industrial Technology (CDTI).
- **IRIDA**. Innovative Remote and Ground Sensors, Data and Tools into a Decision Support System for Agriculture Water Management. IRIDA will use communication based on M2M protocols for the transmission of information. This includes the possibilities that 5G brings to the Internet of the Future and Things combined with wireless sensor networks for distributed information communication and for data processing, considering Big Data Technologies to support the four V's: Variety, Volume, Velocity and Value. Project of the WATER JPI - WaterWorks 2014 led by the CEBAS-CSIC.
- **Producto.** Information system for water and sanitation infrastructures management over Big Data architecture.

Some of the initiatives developed by the private and public tourism sector are listed in an online publication "Open data y Smart data", published by the Tourism Technological Centers of several regions. This publication mentions the great value of the developed activities.

## VI. REFERENCES


1. M. Chen, S. Mao, and Y. Liu, "Big data: A survey," Mobile Networks and Applications, vol. 19, no. 2, pp. 171–209, 2014.
2. C. Philip Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," Information Sciences, vol. 275, no. 0, pp. 314–347, 2014.
3. P. Zikopoulos, C. Eaton et al., Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hill, 2011.
4. P. Russom, Managing Big Data. TDWI best practices report. The Data Warehousing Institute, 2013.
5. S. Kaisler, F. Armour, W. Money, and J. A. Espinosa, "Big data: Issues and challenges moving forward," in System Sciences (HICSS), 2013 46th Hawaii International Conference on, 2013, pp. 995–1004.
6. 2015/2612 (RSP) - 10/03/2016 Texto aprobado por el Parlamento.
7. http://www.bdva.eu
8. http://smartdataforum.de/
9. http://magazin.know-center.tugraz.at/program/big-data-labs-europe/
10. http://www.tieke.fi/pages/viewpage.action?pageId=48366073
11. http://www.regione.emilia-romagna.it/fesr/notizie/2016/febbraio/bigdata-nasce-la-community-emilia-romagna
12. L. Columbus, "84% Of Enterprises See Big Data Analytics Changing Their Industries' Competitive Landscapes in The Next Year", 2014. http://www.forbes.com/sites/louiscolumbus/2014/10/19/84-of-enterprises-see-big-data-analytics-changing-their-industries-competitive-landscapes-in-the-next-year/#1cb8b3863250

13. MckInsey Global Institute," Big Data: The next frontier for innovation, competition, and productivity", 2011. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.
14. J. Campo-Ávila, R. Conejo, F. Triguero, R. Morales-Bueno, "Mining Web-based Educational Systems to Predict Student Learning Achievements", International Journal of Artificial Intelligence and Interactive Multimedia, Vol III, no. 2, España, 2015, pp. 49
15. Patel, A. B., Birla, M., & Nair, U. (2012, December). Addressing big data problem using Hadoop and Map Reduce. In 2012 Nirma University International Conference on Engineering (NUiCONE) (pp. 1-5). IEEE.
16. Hamlen, K. W., & Thuraisingham, B. (2013). Data security services, solutions and standards for outsourcing. Computer Standards & Interfaces, 35(1), 1-5.
17. Survey Analysis: Big Data Investments Begin Tapering in 2016". https://www.gartner.com/lite/document/3446724
18. https://www.fundacionbankinter.org/documents/20183/42758/Publicaci%C3%B3n+Big+data/cc4bd4e9-8c9b-4052-8814-ccbd48324147 Fundación Bankinter 2015.
19. http://www.forbes.com/sites/bernardmarr/2016/03/15/17-predictions-about-the-future-of-big-data-everyone-should-read/#411550c2157c Bernard Marr, March de 2016
20. https://www.ovum.com/big-fast-data/
21. https://www.idc.com/getdoc.jsp?containerId=prUS41826116
22. http://blogs.forrester.com/brian_hopkins/15-11-09-forresters_2016_predictions_turn_data_into_insight_and_action
23. http://www.gartner.com/newsroom/id/3143521
24. http://www.gartner.com/newsroom/id/3143521
25. http://iianalytics.com/analytics-resources/2016-predictions


## AUTHORS

**Felipe Miron** is the European Projects Coordinator of KTO at FIDESOL:He manages the organisation's foresight and technological surveillance actions, both at state and European level, as well as project proposals. He is a Computer Engineer with a degree from the University of Granada. Before joining FIDESOL, he worked for the Regional Government of Murcia leading a development team for the transformation of the Labor Counseling and Social Policy's File Processing system. He has been part of development teams for national and international projects (Banesto, Telefonica, Onda Cero) (Credit Suisse, IBM) andhe has collaborated in several publications related to disciplines such as Artificial Intelligence, Mobility and Big Data.

**Mabel López** has a degree in Computer Science Engineering. She is Knowledge Manager at Fideso and articipates in the research and development strategy of this entit y, technology transfer and analysis of technological trends, such as big data, internet of things, virtual reality, cognitive engines, machine learning, etc. Currently, she is involved in several R & D projects related to the above mentioned technologies and she is the current coordinator of the Big Data Inter-Platform Initiavive in Spain. She has been part of development teams for different national projects (Vodafone, Generalitat Valenciana, IBM, …).

**Clara Pezuela** has a degree in Computer Science from the Technical University of Madrid "Universidad Politécnica de Madrid". She has eighteen years of experience in the development and management of R&D projects and she is currently the person responsible for the Information Technologies market within Atos's R&D group. Her main responsibilities are the management of the research groups focused on cloud, parallel computer science and software engineering, managing the projects and the teams, preparing new proposals and transferring the research results to ATOS's business units. She is the President of the PLANETIC technological platform for the adoption and dissemination of ICTs in Spain.

**Nuria de Lama** studied Telecommunications engineering at the Polytechnical University of Madrid. Nuria is Atos's formal representative in the Software and Services ETP (NESSI) and in the CELTIC Plus Eureka programme. She is also a member of OISPG - Policy and Open Innovation Strategy Group. For the past years she has been involved a great deal in the implementation of Future Internet PPPs (FIWARE), where she has been a member of this initiative's Steering Committee. Some examples of projects are: Synchronicity, BDVe (Big Data Value ecosystem), FIWARE, FI-CORE, BIG, SOA4All, NEXOF, LUISA, TAO, INFRAWEBS, AIM, C@R, CORELABS, BROADWAN, HEROE, ISTforCE, MESSENGER, WALKONWEB, RuralWins and MOSAIC. She has worked as an independent expert for the EC on a number of occasions and she has participated as speaker in many different international events and conferences. She is currently the Deputy General Secretary for the Big Data European Association, where, amongst other tasks, she is responsible for the coordination of the Big Data national initiatives.

**Juan C. Trujillo** is a Full Professor at the University of Alicante (Spain) and the leader of the Lucentia Research Group. His main research topics include Business Intelligence (BI) applications and Big Data. He has also participated in the official registration of different tools related to Data Warehouse modelling and advised 12 PhD students. He has published more than 150 papers in different highly impact conferences, more than 60 papers in highly ranked international journals (JCR) and has been co-editor of 9 special issues. Juan C. Trujillo has been the Principal Investigator of many different Regional and National Research Projects; and Knowledge-Transfer Projects related to BI and PI of several International Projects of the H2020 program. He is also the CEO and Co-Founder of the Lucentia Lab Spin-off. This Spin-off's main activity is focused on Business Intelligence and Big Data Applications. Finally, Juan C. Trujillo holds the international credential of the Project Management Professional (PMP®) awarded by the prestigious Project Management Institute (PMI).

**Juan Luis Sobreira,** holds a PhD in Economics from the University of A Coruña and an Industrial Engineering degree from the University of Vigo. Currently he is a member of the Board of Directors of the Spanish Water Technology Platform, PTEA. He has extensive experience not only in the definition and management of R&D projects, but also in the design of business models and exploitation of results, both at national and international level. In the field of water, he has led projects that address the development of decision support systems in the area of urban planning integration with water management and sanitation or the applicability of big-data as a data valorization and information tool in drinking water and sanitation networks. Projects in which he has participated, such as WIZ (www.wiz-life.eu) and WETNET (www.wetnet.it) have obtained different international recognition for their excellence from programs such as LIFE (European Commission), international events such as WEXGLOBAL or international organizations such as OECD.

**Miguel Angel Mayer** is a researcher at the Research Programme on Biomedical Informatics (GRIB) of the Universitat Pompeu Fabra and the Hospital del Mar Medical Research Institute (IMIM). Miguel A. Mayer is a medical doctor, specialised in Family and Community Medicine. He holds a Master's Degree in Public Health and a PhD in Biomedical Informatics from the Universitat Pompeu Fabra. With more than 25 years of experience as a clinician and researcher, he is a member of the Big Data Value Association (BDVA) of the European Commission and of the Bioinformatics Barcelona Association (BIB). He is also an associate professor and the coordinator of the Biomedical Informatics course belinging to the Bachelor's degree in Medicine at UPF,. He has written a number ofnational and international scientific publications and is involved in several EU projects related to the use of ICTs in healthcare, to the reuse of clinical information for research and the application of Big Data in Biomedicine.

**Amelia Martin** holds a PhD in Law and for the past 12 years has been responsible for the Spanish Innovative Medicines Platform, an initiative whereby programmes to promote Public-Private Partnerships (PPP) within the area of biomedical research are developed. Amelia is a member of EFPIA's (European Federation of Pharmaceutical Industries and Associations) Data Protection Working Group and participates in a PPP IMI consortium, the so-called DO-IT Project. This project aims to harmonise certain issues regarding Data Protection and biomedical Big Data. Amelia is also a member of CEOE's R&D Commission and works actively in promoting and disseminating activities related to

R&D innovative medicines. She is a regular speaker at a number of R&D forums and postgraduate courses and is a professor and author of several publications regarding Bioethics and Biorights.

**Fernando Martín** is the ICT National Contact Point (NCP) at CDTI for the H2020 program. He has a Telecommunications Engineering degree from the University of Valladolid and has worked as an Engineer and Project Manager since 1999. Fernando Martín has been an expert for the H2020 FET Committee, representing Spain and has taken part in the management of a number of different European programs, such as the one he is currently coordinating and acting as national reference point.

**Patricia Miralles** isInnovation manager at ITH. She has a bachelor's degree in Economics, specializing in rural and urban economic development, from UAM (Universidad Autónoma de Madrid), and an Executive Master's degree in Tourism Business Management from EOI (Escuela de Organizacion Industrial). She is responsible for the analysis of trends, technology foresight, promotion of entrepreneurship and management of R&D projects. Previously, she was responsible for industry projects at SEGITTUR and Head of the SME in the State Society changeover to the Euro. She is also coordinator of the Spanish Tourism Technology Platform-ThinkTur. She participates in various committees and working groups related to technology, innovation and entrepreneurship. She is also coordinator and professor of tourism innovation in different schools and universities, as well as a regular lecturer at tourism technology and innovation conferences such as IMAT, Ocitur, Inavar, Turistec, etc.

**Belén García** holds a PhD in Law and is the Director of the Spanish Technological Platform for Containers and Packaging (PACKNET). She holds an Executive MBA from IE Business School and in her professional career of over twenty years she has held different managerial positions both in Spain and abroad.

**Jesús Poveda** is Technical Manager of Logistop. He is an Industrial Engineer with experience as coordinator of the Technology Platform of integrated logistics, intermodality and mobility, LOGISTOP. Jesús Poveda supports the development and implementation of R&D projects related to logistics and transport and is a point of union between the industry and public financial agents: European Commission, CDTi, Clusters and other national and international platforms. Previously he was manager for R&D proposals and projects at the National Aerospace Technology Institute (INTA), managing the participation of INTA in European and national R+D programs.

TAYRNE BUTLER holds a degree in Tourism Activities and Companies from the University of the Balearic Islands and a Specialized University Course Certificate in Translation from English to Spanish as well as a Specialized Postgraduate certificate in English for Specific Purposes. In 2014, she cofounded AnySolution S.L., a company that develops strategic methodologies and projects in the fields of R&D, IoT, Big Data, Tourism, SmartCities and Destinations, SmartMobility and Emergency mitigation amongst others and also offers specialised translations in these fields. Before this she worked for a company specialized in the MICE tourism sector organizing a number of different networking events and international workshops and while working there she also helped to set-up and launch the Trip&Treat interactive publicity campaign platform, working on the company's business model, market studies and back-office development. Previously she worked for the Regional Government of the Balearic Islands where she co-managed the European Enterprise Network and worked on the coordination and financial management of European Projects and Community Initiatives in a number of different fields.