Research　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

Big Data: Witnessing the Birth of a New Discipline

Valarezo UA[1*], Pérez-Amaral T[1] and Gijón C[2]

[1]*Complutense University, Spain*

[2]*Covadonga Gijón, Universidad Carlos III de Madrid, Spain*

*****Corresponding author:** Valarezo UA, Complutense University, Spain, Tel: 913942380; E-mail: angel@abevu.com

Abstract

What if we ask ourselves about what is the operating system and the physical infrastructure behind tools and services people use almost every day, and what is making possible that well-known companies like Google, Yahoo, Bing, Amazon, Facebook, Twitter, LinkedIn, Netflix, etc. are able to deliver such high quality services like they do. What makes possible the increasing accuracy of Google Translator, the appropriate recommendations of Amazon, the contact suggestions of LinkedIn, the Netflix's hit House of Cards. Of course we have to imagine something much more bigger than the operating system we use at home or within a small company to drive our day to day work.

Though the beginnings of Big Data as hype term are not far away in time, it seems lengthy if we consider how it is evolving from just a technological phenomenon to a new discipline, which comprises many areas of knowledge, challenging not just the directly related ones as computing science, statistics, data science but others less obvious as sociology, ethics, philosophy, etc.

Three goals have driven this work: build our own definition and understanding of Big Data; get experience at using available tools based on related technologies and obtaining an approach about how interested is Spain in dealing with the new challenges Big Data represents.

Introduction

Big Data is far from being just a buzzword or being endorsed only by its large numbers. It has already obtained big relevance worldwide, given, unless in part, for the interest showed by the top-level decision makers in other countries. A good example is the 90-day study "Seizing Opportunities, Preserving Values" asked, in January-2014, directly by the president of the United States of America to his Counsellor John Podesta, [1].

Querying Google Ngram Viewer about how often "big data" appears in books since 1920 let us know that the concern about this topic is nothing new **(Figure 1)**.
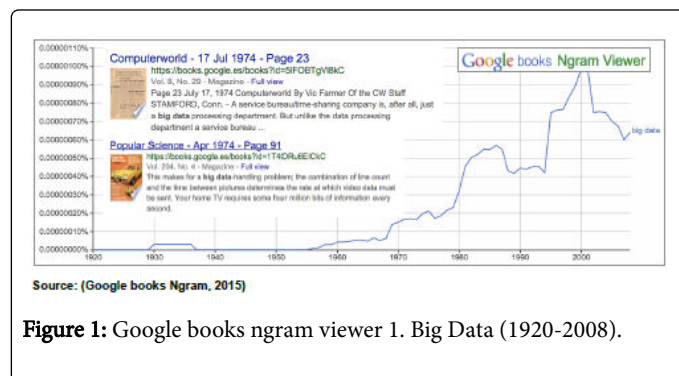

Figure 1: Google books ngram viewer 1. Big Data (1920-2008).

For many of those who have been working with data within organizations that have had kept updated their data systems to the requirements that digital age demand, Big Data could just be data.

People who are used to work extracting value from vast amounts of data are also used to face the challenges of finding alternative ways to meet the requirements set by the increasing of its availability, the speed at which they are produced and the variability of their sources and formats. So, for this reason the relatively new concept of Big Data, at least in its early days, could have represented nothing more than hype.

Already in May 2013, according to Financial Times, Big Data (BD) was one of the most hyped terms of the market. "It's one of the most popular search terms by CIOs and other IT professionals on gartner.com, and according to the Global Language Monitor (GLM) it was the most confounding term of 2012" [2].

This paper starts with the attempt of developing our own definition of Big Data, based on both a technical and a broader approach. And within the same section aiming to bringing some more clarity other subtopics have emerged: Recognising a Big Data Problem, Barriers to Big Data in Organizations, Some Challenges to Face, The Origin of Big Data, Trends that drive Data Growth and Big Data as a driver of big trends, and some examples of working with Big Data.

The second part is focused on getting an approach about what is Spain doing, but above all, how much interest is the country showing and how it is related with its level of digital capacities.

What is Big Data?

The common discussion about whether Big Data is a hype or not could be out of place. At least this impression comes to our mind if we take into consideration the 20 years old 'Gartner's Hype Cycle' in its Special Report for 2014. It tells us not only that Big Data is a hype, but also it has just already passed the top of the Hype Cycle, letting the top

for Internet of Things, and moving Trough Disillusionment, what is not a bad thing, because it means the market starts to mature, and therefore is more capable to realize how Big Data can be useful for organizations. In other words it seems to be a sign that Big Data will become business as usual **(Figure 2)** [3].
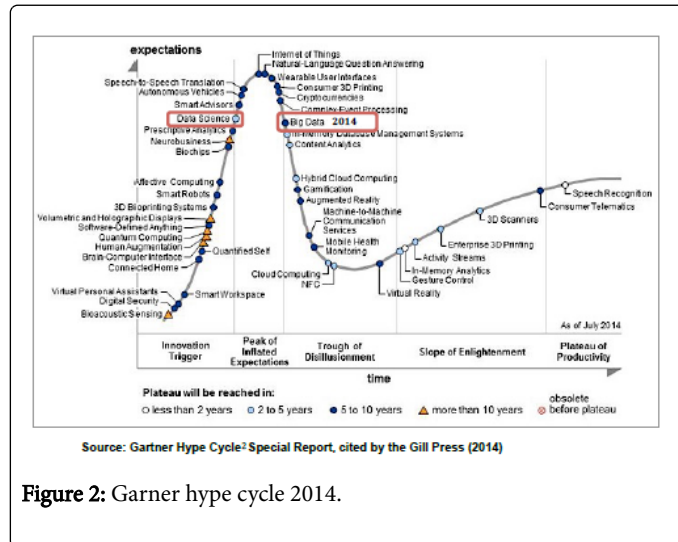

Figure 2: Garner hype cycle 2014.

Big Data topics are experiencing rapid and large shifts over the peak of the hype, and this is accompanied by the rise to the peak of analytics, which suggest that the interest in information and data is shifting from supporting and managing Big Data to actually using information and data to make business decisions [4].

Once justified that no matter if Big Data is a hype term or better yet, because of it, to find a definition is a necessary starting point.

Surprisingly it is not easy to find definitions of the term beyond technical characteristics [5], which will change within very short periods of time, encompassed with the growing of the availability of data coming from different sources, sectors and almost all organizational functions. Of course technical definitions are necessary to recognise when we are facing big data challenges, in order to allow an appropriate classification and therefore an efficient resource allocation; but at the same time if we think about the current relevancy of BD and, as Victor Mayer-Schönberger and Kenneth Cukier said, "its potential to reshape how we live, work, and think", we need broader and more comprehensive approaches that help us to realize the true magnitude of what it means.

Without the goal of being comprehensive we have chosen a few of the most cited definitions, both technical as other that go beyond a technical approach.

"Hype Cycles offer a snapshot of the relative maturity of technologies, IT methodologies and management disciplines. They highlight overhyped areas, estimate how long technologies and trends will take to reach maturity, and help organizations decide when to adopt" [6].

Technical definitions

Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making [7]. The three key dimensions of big data were first coined by Laney in 2001, when he warned that the business conditions and mediums are pushing traditional data management principles to their limits [8].

It is the data that exceeds the processing capacity of conventional database systems. To fit into the concept the data should be too big, moves too fast, or doesn't fit the structures of regular database architectures [9]. In this concept we can find the three V's of Big Data.

Against the misconception of that BD is just about size, EMC defines it as any attribute of data that challenges constraints of a business capability or business needs [10].

Big data refers to dataset whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse [11]. As the authors mentioned in the report where this definition was published, it is intentionally subjective and can be seen as a moving definition which adapts instantaneously to how big a dataset needs to be in order to be catalogued as big data.

Here is a very comprehensive approach that refines the Gartner definition [12]: "Big Data (Data Intensive) Technologies are targeting to process (1) high-volume, high-velocity, high-variety data (sets/assets) to extract intended data value and ensure high-veracity of original data and obtained information that demand (2) cost-effective, innovative forms of data and information processing (analytics) for enhanced insight, decision making, and processes control; all of those demand (should be supported by) (3) new data models (supporting all data states and stages during the whole data lifecycle) and (4) new infrastructure services and tools that allows also obtaining (and processing data) from (5) a variety of sources (including sensor networks) and delivering data in a variety of forms to different data and information consumers and devices.

Big Data Properties: 5V (volume, velocity, variety, value and veracity).

New Data Models: Data linking, provenance and referral integrity, data lifecycle and variability/Evolution.

New Analytics: Real-time/streaming analytics, interactive and machine learning analytics.

New Infrastructure and Tools: High performance computing, storage, network, heterogeneous multi-provider services integration, etc.

Source and Target: High velocity/speed data capture from variety of sources, data delivery to different visualization and actionable systems and consumers, full digitalised input and output, etc."

After this, it is easy to find place for a sixth characteristic:

(6) Highly comparable well trained data analysts.

Beyond of just technical definitions

Wrobel [5] argues that besides that BD refers to the trend towards availability of ever more detail than ever, closer to real time data, and its economic potential, the really important thing about BD is a kind of philosophical issue, which consists in the switch from a model-driven approach to a model and data-driven approach to science, to society to business.

BD is not just volume, variety, velocity, in memory computing, real time analysis or its effects of scale, but, although BD is all this things, the long running trend is that we don't model, we don't look at society and business the way we thought it should be, but we are looking at

these things the way they are. This new approach brings severe philosophical implications in terms of the relative reliability of conclusions, statistical phenomena that are not reliable because of the nature of the data, etc.

Boyd et al. [13] define Big Data "as a cultural, technological, and scholarly phenomenon that rests in the interplay of:

Technology: Maximizing computation power and algorithmic accuracy to gather, analyse, link, and compare large data sets.

Analysis: Drawing on large data sets to identify patterns in order to make economic, social, technical and legal claims.

Mythology: The widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy." Newly trained and competent analysts.

Our definition

Big Data could be defined as new field of knowledge as well as a new socio-economic and technical phenomenon founded in the high availability of data, which, given a set of characteristics that demand new approaches of recollection, storage, processing, analysis, protection, ways of thinking, etc., brings unusual challenges, opportunities and threats to society.

Recognising a big data problem

Besides a definition it is equally useful trying to define the main features to answer the question: How do we know when we have a big data problem? In 'Tackling the Challenge of Big Data', an online course taught by experts from MIT Computer Science and Artificial Intelligence Laboratory (Csail), Samuel Madden (2014) adds a fourth characteristic (Non -Scalable Analysis) to the well-known 3 V's, first mentioned in 2001 in the Garner's report "3D Data Management: Controlling Data Volume, Velocity and Variety" [8].

Volume = Too many bytes: This is sort of literal definition: when you have too much data.

Velocity = Too high a rate: If you have a lot of data coming at you really fast.

Variety = Too many sources: When the data comes from lot of different sources that need to be integrated together.

Non-Scalable Analysis: If the data is just hard to process form some reason. Usually it might be that it takes a lot of manual labour to find what is needed. And, also it might be some algorithm that takes a long time to run on the data, and, in order to extract the insight pursued it needs to be improved o processed on a smaller subset.

It is common to find Veracity, Value and other features, whether or not they start with V. But even though other features, different than those already mentioned, help to define the ideal of Big Data and allows to realize the correct path, not necessary helps to determine if we are in front of a Big Data problem. In other words, some features are useful to identify a Big Data problem, and others are useful to identify challenges of Big Data.

Barriers to Big Data in organizations

The main barriers to face the challenge of Big Data come from: novelty of the technologies, easy of confusing causation with correlation, easy of finding fallacious patterns in the data, need for cultural change, and the privacy and security concerns [14]. Other somehow related with the already mentioned barriers have been identified by a survey of 507 German companies in 2014: Too few big data specialists (70%), Technical IT Security Requirements (61%), Insufficient Budget (59%), Privacy Regulations (48%), Big Data Tools/Solutions are not sufficient yet (42%), I know of too few suppliers of big data solutions (35%), We don't have enough data (22%), I don't know of sufficiently many usage areas (22%), and, Our data are of insufficient quality (9%) [5].

Some challenges of Big Data

Whether the barriers mentioned above themselves reflects some challenges and are among the most important, it is possible to get a more comprehensive approach from the rated responses about 'Key Challenges of Big Data Across Regions of World' of the TCS 2013 Global Trend Study [15]. The respondents rated 16 challenges that the authors of the study found most frequently in press articles, public speaking presentations, and their client work. The result is the following ranking:

- Getting business units to share information across organizational silos.
- Being able to handle the large volume, velocity and variety of Big Data.
- Determining what data (both structured and unstructured, and internal and external) to use for different business decisions.
- Building high levels of trust between the data scientist who present insights on Big Data and the functional managers.
- Finding and hiring data scientist who can manage large amount of structured and unstructured data and create insights.
- Getting top management in the company to approve investments in Big Data and its related investments (e.g., Training, etc.).
- Putting our analysis of Big Data in a presentable form for making decisions (e.g, visualization/visual models).
- Finding the optimal way to organize Big Data activities in our company.
- Understanding where in the company we should focus our Big Data investments.
- Determining what to do with the insights that are created from Big Data.
- Reskilling the IT function to be able to use the new tool and technologies of Big Data.
- Getting functional managers to make decisions based on Big Data, rather than on intuition.
- Determining which Big Data technologies to use.
- Keeping the data in Big Initiatives secure from external parties.
- Getting the IT function to recognize that Big Data requires new technologies and new skills.

Keeping the data in Big Data initiatives secure from internal parties.

For ranking the list the scale was set between 1 (not at all challenge) and 5 (very high challenge). And given that the 16 challenges received mean ratings between 3.0 and 3.4, none of them really stands far above the others.

A clearer picture about the challenges of Big Data was showed by the Profesor Etefan Wrobel. In the European Data Forum 2014 [5]. They summarized en only 7 points:

Big Data is not an isolated IT topic, but must address business value end-to-end in company/sector specific ways.

- Technical solutions must be designed-to-fit.
- Further solutions needs beyond off-the-shelf software.
- Data Linking and brokering need open standards.
- Security and Privacy are demanded by business and society alike- "by design".
- Enormous educational and training needs.
- SME and start-ups face special challenges and need a supportive ecosystem.

Other challenges which are more related with technical aspects are focused in overcome the obstacles in the development of Big Data applications: Data Representation, Redundancy Reduction and Data Compression, Data Life Cycle Management, Analytical Mechanism, Data Confidentiality, Energy Management, Expendability and Scalability, and Cooperation [16].

Threats of Big Data

The vast and detailed amount of data that can be gathered from almost every area of our lives, activities, cities, companies, institutions, governments, scientific and academic production, etc., coupled with the ability of analysing and combining with other pertinent data, could represent an enormous threat if the outcome is used without attend to legal restrictions or in its absence without attend to generally accepted principles of ethics.

It is not an easy task to determine all the risks Big Data brings, above all because this exercise require putting yourself in the role of many different actors and consider many different scenarios, interests and variables. Nevertheless it is noteworthy to point some of the treats mention by Viktor Mayer-Schönberger, Kenneth Cukier, who spend an entire chapter of their book to this topic:

Given the expected value that any personal data can reach in the future many actors have incentives and means to gather, store, use and reuse it. The problem is that the second or subsequent uses could not been known neither authorized by those affected.

Protecting privacy is now much harder.

If we take into consideration the possibility of making big-data predictions about people it is closer the possibility or the tempting of establish penalties based in propensities, what undermines the basic principles of justice and free will.

We could face the risk of letting data and the outcome of its analysis dominates decisions that should have in consideration other factors. And also if it falls into wrong hands it may be used as an instrument of repression. The origin of Big Data.

When we talk about the origin of Big Data, we can think in data as a raw material, data generation and data acquisition are exploitation process, data storage is a storage process, and data analysis is a production process that utilizes the raw material to create new value [16]. So, when we ask 'where the data comes from?', we want to focus our attention in the first step of big data.

It is well known that the sources are highly diverse and that the Variety is one of the defining elements, so it could be useful trying to establish at least some clues for sorting data by its origin. Taking into account the importance of the traceability of data and its relationship, both direct and indirect with the key challenges of Big Data, we propose three ways of classification based in the following questions: from who?, from what?, and where does it come from?

Whom it comes from?

Identifying who has generated the data matters as much as the data itself. This is important because how the data is going to be stored, treated, analysed, presented and undisclosed, if it is the case, depends on whom the data came from. Individuals, public and development sector, private sector, all of them have their own types of data, sharing incentives and requirements. The following box is just an example of the kind of classification that can be built when we think on who originated the data **(Table 1)**.

Individuals	Data type	Crowd sourced information
		Data exhaust
		Personal information
	Sharing incentives	Pricing offers
		Improved services
	Requirements	Privacy standards, 'opt out' ability
Public and Development Sector	Data type	Census data
		Health indicators
		Tax and expenditure information
		Facility data
	Sharing incentives	Improved service provision
		Increased efficiency in expenditures
	Requirements	Privacy standards, 'opt out' ability
Private Sector: Companies, institutions, science, education	Data type	Transaction data
		Spending and use information
	Sharing incentives	Improved consumer knowledge and ability to predict trends
	Requirements	Business models
		Ownership of sensitive data
Source: Adapted from (World Economic Forum, 2012)		

Table 1: Whom data comes from?

From what?

Data may come from a variety of sources and it could be classified in different ways. But if we just think about it, the kind of things that we meet often and are easy to find in any organization, we can easily build a very large list, but hardly a complete one: Smartphones, telecommunication network, telecommunications infrastructure, Sensors, personal digital devices, wearable devices, websites, business operations, big data systems, digitalized data from analogic formats,

historical data already stored, health records, mobile applications, web applications, payment transactions, geographic location, academic publications, scientific experiment, emails, weather, audio, video, social networks, etc.

What is relevant here is not just identifying the sort of things but identifying the sort of trends that this collection of sources can represent. And it is also important because of its relationship with key data management concepts as they are: Data model and schema. As Anthony D. Joseph says within an edX3 course about Big Data [17]: Data model refers to a collection of concepts for describing data; and Schema means the description of a particular collection of data, using a given data model. Based on this, data can be structured, semi-structured and unstructured, ranging from schema first (structured) to schema never (unstructured). Structured data is schema-first and comprise relational databases, formatted messages, etc. Semi-structured or schema-later data includes documents, Extensible Markup Language content, tagged text or media, etc. And, Unstructured data or also called schema-never, which can be plain text, media, etc.

Where it comes from?

This question refers not to geographical place but to various industries and sectors within any economy **(Table 2)**.

Nowadays data comes from everywhere and if we wanted to be exhaustive we would have to number all industrial, economic and social sectors. But the objective is just to show how any sector can generate big data and therefore be interested in exploit it. For this purpose we mention just 15 of them:

Banking/financial services
High Tech
Retailing
Consumer products manufacturing
Travel, hospitality and airlines
Insurance
Heavy manufacturing
Telecommunications services
Pharmaceuticals/life sciences
Media and entertainment
Utilities
Science and Academia
Government
Development
Source: Adapted from Tata Consultancy Services [15]

Table 2: Some of the sectors that produce Big Data.

The Global Trend Study 'The emerging Big Returns on Big Data' [15] contribute with the following highlights:

- Telecom, travel, high tech and banking firms spend the most on Big Data.
- However, utilities and energy and resources companies expected the biggest ROI.
- Media and entertainment companies use the most unstructured data; high tech and telecom companies use the most external data, which means data that comes from external sources.
- Utilities and telecom companies are most likely to sell their digital data, but insurance companies make the most from that data.

Trends that drive data growth and Big Data as a driver of big trends

From the review of the literature it is easy to find that as well as Big Data feeds modern trends, it is feeding by different sources and new technologies. But what call our attention is how the Big Data phenomenon is fed by the outcomes from trends that itself has fostered before. Let's take Google search as an example. Given that the big amount of data that Google already has from crawling and indexing the web of trillion of documents [18], as well as all the searches people do, fuels its system allowing Google to improve the accuracy of its searching tool. This kind of virtuous circle allows the continuous improvement and enrichment of its system and can be seen in many different examples **(Figure 3)**.

Source: (Google Trends, 2015b)
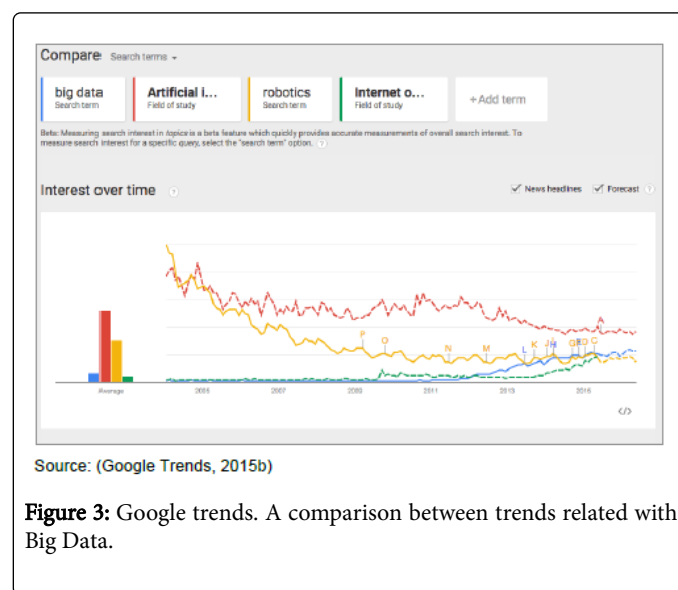
Figure 3: Google trends. A comparison between trends related with Big Data.

Some of the main trends related with Big Data are:

- Internet of Things
- Machine Learning
- Deep Learning
- Artificial intelligence
- Programmatic Advertising
- Data Science

Examples of working with Big Data

There are many examples of applications, products, services, or, in general, projects and uses of Big Data. To justify why there is so much excitement around the topic, lets look at several examples:

Analytics at Spotify:

Analyse a massive and growing data set collected by Spotify from its users and operations make possible to recognize trends, discover bugs, and analyse the effect of an event on a user and the entire ecosystem [19]. At the end most of the efforts are user-centric and allow Spotify to provide music recommendations, choose the next song to hear on radio, encompass rhythms and kinds of music with the intensity of a sport activities like running.

Real Time Bidding (RTB):

RTB and Programmatic Buying and Selling of Online Advertising are address to solve a central issue in performance display advertising, which is matching campaigns to add impressions that can be formulated as a constrained optimization problem. The aim is to maximize revenue subject to constraints such as budget limits and inventory availability [20]. "Not like the conventional digital advertising, in the process of RTB, the impressions of a mobile application or a website are mapped to a particular advertiser through a bidding process which triggers and held for a few milliseconds after an application is launched" [21]. To deal with this complexity needs Bid Data and Machine Learning expertise to sift through the data noise and get the meaningful signals that match with the campaign goals [22].

Google:

Companies that were born digital such as Google, Amazon, Facebook, Twitter, etc., are already masters of Big Data. So, when we talk about Google it is difficult to choose only one example, therefore it is necessary to take at least some of them: Google Search, Google Flu, Google Trends, Google Correlate, Google Translate, Google Ngram, Google Self Driving Car, etc. Especially interesting is Google Photos4, not only because it has been recently re-launched, but because it is an example of what can be done with Big Data and other related technologies and above all it raises a big question, what are they going to do with all the images they are collecting from people all over the world?

The GDELT project:

Based on a research work about Global Data on Events, Location and Tone [23], and supported by Google Ideas, this project consists, as is mentioned in the its homepage, in an online application that monitors the world's broadcast, print, and web news from nearly every country in over 100 languages and identifies the people, locations, organizations, counts, themes, sources, emotions, counts, quotes and events driving the global society every second of every day, creating a free open platform for computing on the entire world [24].

Madiva soluciones:

A Spanish start-up, acquired by BBVA in 2014, specialised in offering services based on Big Data and cloud computing technologies. Only some of their success stories are [25]:

Being able to anticipate the potential demand of telecommunications of a family, based on public information available on Internet.

Automatic studies of housing market.

Predict who are going to change home appliances and when.

Just trash:

This comes from a personal experience of one of the authors of this paper, who was part of a team, which participated in a worldwide Hackathon about smart cities. Although JusTrash is not finished yet, it was the first prize winner project in Datafest Madrid edition. The idea is to use sensors to see how full a trash can is; whenever someone throws something into the trash can, an update is sent to the servers; the client can then access the data form a web application and obtain the most efficient pick up route, district by district. The real strength of the project is not that can help gain efficiency but that it can be used to get information about population and human activity within cities [26].

Big Data in Spain

How much interested is Spain in big data?

In Spain there has not yet been such initiatives like the United States Government led in 2014, when President Obama called on the administration to conduct a broad 90-day review of Big Data and Privacy, which had as a result an insightful report titled "Big Data: Seizing Opportunities, Preserving Values" [1]. Although is fair to say that the Congress of Deputies of Spain has already approve a proposal to forest: Big Data through the collaboration between private and public organizations, applications for smart cities and other topics like tourism and health, and taking into account privacy protection and the necessary measures to guaranty anonymization. Additionally according to the proposal the Government should align Big Data with Open Data Strategy and deliver an awareness campaign addressed to citizens with the aim of informing about Big Data, its benefits and the risks it could represent.

Turning the view towards the participation and interest of private sector in Spain it is easier to find several cases of business, users, events, communities, experts, Big Data evangelists and other actors. Just to mention some of them:

Business and Start-ups: Stratio, Xeridia, Pivotal, T-Systems, PiperLab, Bidoop, Daedalus, DataSalt, Cubenube, The Data Republic, Vizzuality, U-tad, Treelogic, Amadeus, etc.

Users/Clients: Telefónica, Easyap, Puleva, JustEat, BBVA, SAP, SONY, Santander Bank, etc.

Events: Big Data Spain 2013–2014-2015, Big Bang Data 2015, Global Urban Datafest 2015, Big Data Value Association Summit in Madrid, Big Data Week 2014, Strata + Hadoop World–Barcelona 2014, Innova Challenge, Ojo al Data (Medialab Prado), Living in a Sea of Data 2015 (Fundación Telefónica), Dare2Data (BBVA), etc.

Some sponsors and event organizers: BBVA, Fundación Telefónica, UCM, IBM, Amazon Web Services, Stratio, Fundación Bankinter, Esri España, Madrid Emprende, KPMG, Pivotal, Mongo DB, Microsoft, Silver Datastax, Synergic Partners, BEEVA, Paradigma, JavaHispano, Ticbeat, etc.

Communities: Data Beers, Hackathon Lovers, Data Science Spain, Big Data 4 Success, Several Meetup groups, etc.

Speakers, Leaders, Experts and Big Data evangelists: Óscar Méndez (CEO and Foundig partner Stratio), Elena Alfaro (CEO at BBVA Data & Analytics), Esteban Moro (Associate Professor at Universidad Carlos III de Madrid), Victoria López (Head of G-TEC Researching Group at Universidad Complutense de Madrid), Javier Ramírez (CEO, Teowaki),

Marcelo Soria (VP Data Services, BBVA Data & Analytics), Jesús Escudero (Data Journalist, El Confidencial), Chema Alonso (CEO, Elevan Paths–Telefónica), Pau García (Founder and Researcher, Domestic Data Streamers), Pedro Pablo Pérez (Head of CyberSecurity, Telefónica), Soraya Paniagua (Data Journalist), etc.

In order to get closer to realize how interested is Spain in Big Data, and how its interest has evolved over time this research relies in a series of tools, powered and provided for free by Google and based on Big Data technology. The main tools are: Google Trends, Key Planner Tool of Google Adwords, Google Ngram and Google Correlate **(Figures 4 and 5)**.

From the analysis of the results obtained for Spain and the United States it is worth to note the following:

As **Figure 6** shows, Big Data trend starts to be relevant within Google Searches in Spain around November 2011, what contrasts with earlier adopters countries such as United States or United Kingdom.

It's not surprising how well ranked are the interest showed by Catalonia, Community of Madrid or Basque Country, given that these are three of the most important engines of the national economy. The same happens in United States **(Figure 7)** where Massachusetts and California are on the top positions.

In Spain, the top searched terms related with Big Data show kind of interest of first discovering. In other words it seems that Spanish people are already still interested in knowing what is Big Data? Especially relevant is how well positioned is the query 'master big data' and the opportunity that it could represent for training and educational sector. On the other hand, the interest of US people about the topic seems to be related with tools, applications, vendors and perils.

Figure 4: Google trends 5. Searched term: Big Data, Spain (2004-2015).

Figure 5: Google trends. Searched term: Big Data, USA (2004-2015).

Looking at **Figure 6** it is not hard to realize how the trend behaves worldwide. But although the figure shows a static map, within the tool we can see how the trend started in United States and passed to other countries. Anyway, most of best ranked countries do not represent any surprise given its development and economic growth; however Kenya case could be related with Big Data applications developed and used to alert about Ebola epidemic [27].
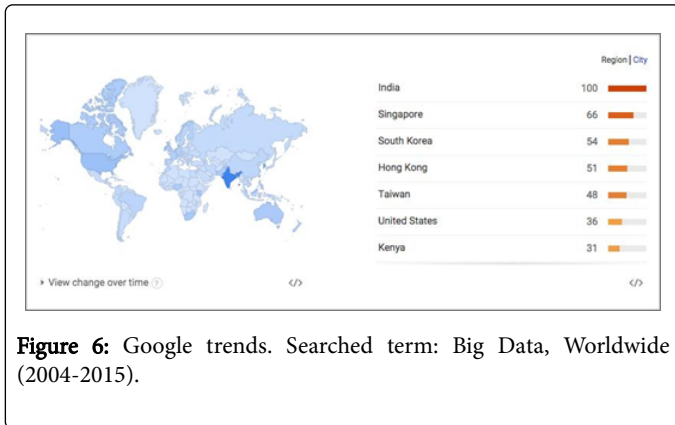

Figure 6: Google trends. Searched term: Big Data, Worldwide (2004-2015).

The next figures allow us to perform an analysis about the relationship between the interests showed by the countries and how well or bad they are placed within an international comparison about their digital capacity.

Taking into consideration **Figures 7, 8 and 9**, here are the key highlights:

From comparing the chosen countries in Figure 8 with their behaviour at searching for "big data", rises that match the digital capacity **(Figure 7)** of each country with how early it show a relevant interest in Big Data and how high is its compared average of searching for the term [28-35].

Although Spain started later showing a relevant interest, in June 2014 the country stands out from the other countries with which is compared in **Figure 8**.

Singapore deserves a mention apart. The great distance of its digital capacity from the other countries corresponds with the very high interest that can be perceived if we look at **Figure 9**.
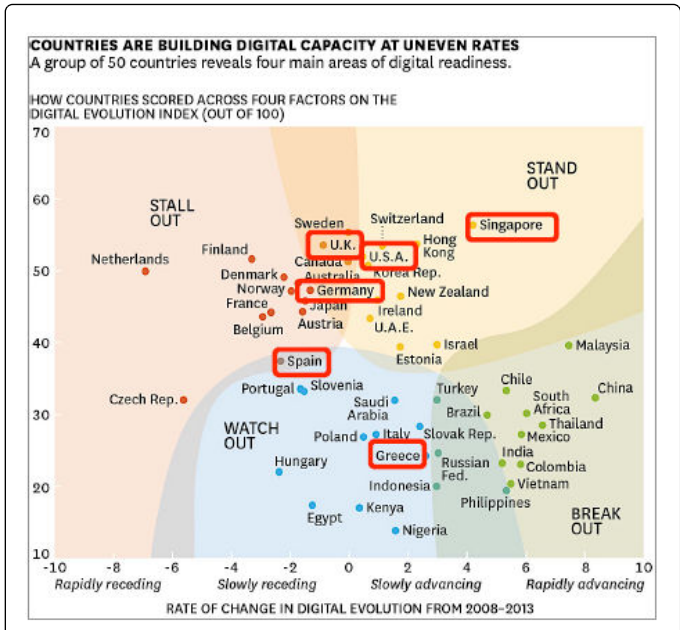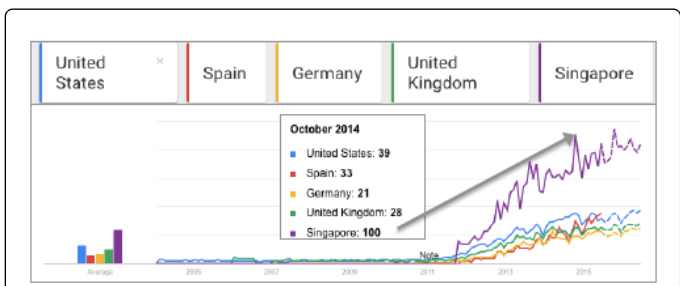

Figure 7: Countries building digital capacity at uneven rates.
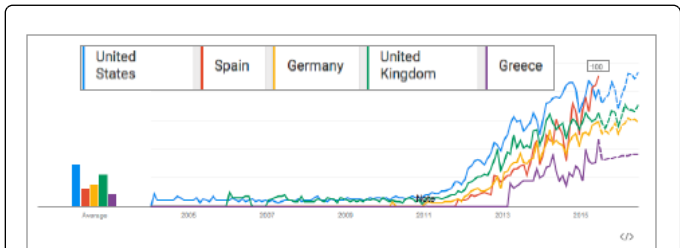
Figure 8: Google trends. Searched term: "Big Data" a comparison between different countries (2004-2015).

Figure 9: Google trends. Searched term: "Big Data". A comparison between different countries (2004-2015).

Figures 10 and 11 bring us another perspective. The gathered information from the Key Planner of Google Adwords gives information about the average volume of searches.
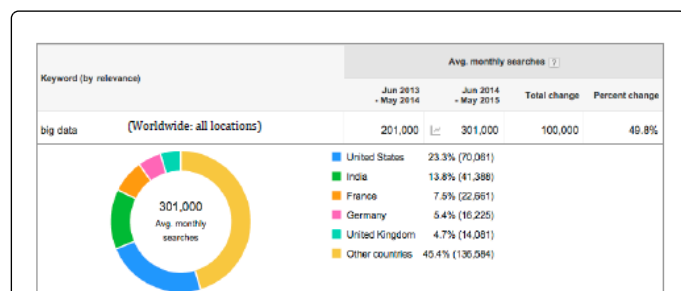
The most important points extracted are:

- The average monthly searches of Spain are around 9,900 and it represents 3.3% of the total worldwide.
- Although the average monthly searches of Spain are quite far to the volume of searches of United States or India, there is no much distance with neighbouring countries as France, Germany or United Kingdom.
- Again the results show what can be expected if we consider the digital capacity of its country.
- The importance of this information resides in its power to describe the level interest that Big Data attracts, both general public and business.

Figure 10: Average monthly search of "Big Data". All locations Jun 2014-May 2015.

Figure 11: Average monthly search of "Big Data". Spain Jun 2014-May 2015.

This last **Table 3**, beside the information about the volume of searches, gives information of the degree of competition and the suggested bid in order to place an online advertising campaign with Google Adwords [36-44].

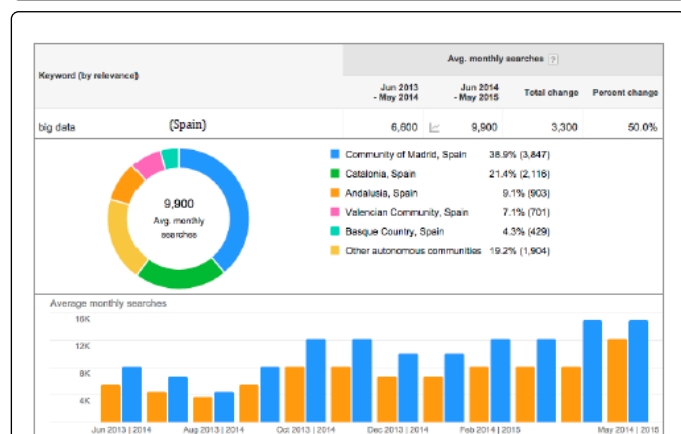| Ad group (by relevance)	Ku/words	Avg. monthly searches	Com potto n	Suggest ed bid
Keywords like: Sig Data (t)	Big data (Local: United Stat...	60:500	High	1.29
Keywords Ike: Big Data CI)	Big data (Location: UK)	12100	High	€ 16.60
Keywords Ike: Big Deta (.%)	Big data (Location: Germany)	14,800	High	€ 16.41
Keywords Ike: Sig Data (I)	Big data (Location: Spain)	9,900	Mediu m	€ 3.19
Keywords Ike: Dig Data (ii)	Big data (Location: Greece)	590	Low	I2.91
Source: Keyword planner of Google Adwords (2015)				

Table 3: Average monthly search and sugested big for "Big Data". Diferente locations.

Conclusions

Even though Big Data can be considered as a hype term and it has already passed the top of the hype cycle the relevance of the phenomenon deserves resources, analysis and a deep discussion, which should incorporate the participation of all the actors and stakeholders.

It is needed a definition that goes beyond such technical aspects. It should be comprehensive and above all should allow the understanding of the depth implications it can represent for society.

There is a kind of virtuous cycle in which Big Data phenomenon is fed by the outcomes from trends that itself has fostered before.

There are free and powerful tools, as the ones we used in this paper, which allow researchers access to insightful information.

Spain has started late to show a relevant degree of interest on Big Data. Nevertheless, if we look at the range of activities, initiatives, vendors, users, and trends, it seems that the country as an economy is trying to get ready to tackle the challenges of Big Data.

The fact that the digital capacity of Spain is not as good as would be desirable is reflected in the degree of interest showed over time and could represent a limitation in order to leverage the potential of Big Data.

Next steps

Explore the range of technologies needed to develop Big Data projects.

Define best practices to lead Big Data projects.

Experiment with Big Data available solutions in order to find alternative and cost efficient methods to analyse socioeconomic information.

References

1. Executive Office of the President (2014) The White House. BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES.
2. Laney D, LeHong H, Lapkin A (2013) What Big Data means for business. Financial Times.
3. Buytendijk F (2015) Hype Cycle for Big Data.
4. Burton B, Willis DA (2015) Gartner's Hype Cycle Special Report for 2014.
5. Wrobel S (2014) The Value of Big Data - From Data-Driven Enterprises to a Data-driven Economy.
6. Fenn J, Raskino M (2015) Understanding Gartner's Hype Cycles.
7. Beyer MA, Laney D (2012) The Importance of "Big Data": A Definition.

8.   Laney D (2001) 3D Data Management: Controlling Data Volumen, Velocity and Variety.

9.   Dumbill E (2012) What is big data? - O'Reilly Radar.

10.   EMC (2012) How Big is Big Data?.

11.   Manyika J, Chui M, Brown B, Bughin J, Dobbs R, et al. (2011) Big data: The next frontier for innovation, competition, and productivity.

12.   Demechenko Y (2014) Big Data and Data Intensive (Science) Technologies.

13.   Boyd D, Crawford K (2012) CRITICAL QUESTIONS FOR BIG DATA. Information, Communication & Society 15: 662-679.

14.   McAfeeErik A, Brynjolfsson E (2012) Big Data: The Management Revolution.

15.   Tata Consultancy Services (2013) The Emerging Big Returns on Big Data.

16.   Chen M, Mao S, Zhan Y, Leung VCM (2014) Big Data. Related Technologies, Challenges and Future Prospects.

17.   Joseph AD (2015) Introduction to Big Data with Apache Spark.

18.   Google (2010) Crawling & Indexing.

19.   Palmer J (2013) Analytics at Spotify.

20.   Chen Y, Berkhin P, Anderson B, Devanur NR (2011) Real-time Bidding Algorithms for Performance-based Display Ad Allocation. In Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp: 1307-1315.

21.   Adikari S, Dutta K (2015) Real Time Bidding in Online Digital Advertisement. In: Donnellan B, Helfert H, Kenneally J, Vander Meer D, Rothenberger M, & Winter R (eds.). New Horizons in Design Science: Broadening the Research Agenda, pp: 19-38.

22.   Sense Networks (2014) Real Time Bidding Platform.

23.   Leetaru K, Schrodt PA (2013) GDELT: Global Data on Events, Location and Tone.

24.   The GDELT Project (2014) The GDELT Project.

25.   Madiva (2009) Madiva.

26.   Altamirano F, González Y, Piña C, Sánchez J, Valarezo-Unda A (2015) Justrash - Global Urban Datafest.

27.   Basir A (2015) Africa: Is Big Data the Solution to Africa's Big Issues?

28.   Carr D (2013) For "House of Cards," Using Big Data to Guarantee Its Popularity. The New York Times.

29.   Chakravorti B, Tunnard C, Chaturvedi S (2014) Digital Planet: Readying for the Rise of the e-Consumer.

30.   Congreso de los Diputados (2014) Boletín Oficial de las Cortes Generales.

31.   Google (2015) Keyword Planner Google Adwords.

32.   Google books Ngram (2015) Google books Ngram Viewer, searched query: "big data", between 1920 and 2008.

33.   Google Trends (2015a) Google Trends - Web Search interest - Worldwide, 2004 - present.

34.   Google Trends (2015b) Google Trends - Big Data, Artificial Intelligence, Robotics, Internet of Things - Web Search interest - Worldwide, 2004 - present.

35.   Google Trends (2015c) Google Trends - Web Search interest - Spain, 2004 - present.

36.   Kimball R, Ross M (2013) The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, 3rd Edition. John Wiley & Sons.

37.   Nimpact K (2014) Bigprivacy,data security and consumer welfare. Telecommunications Policy, 38: 1134–1145.

38.   van der LJ (2013) Using Big Data to Build Value to Operators.

39.   Lokanathan S, Gunaratne RL (2015) Mobile Network Big Data for Development: Demystifying the Uses and Challenges.

40.   Madden S, Stonebraker M, Zaharia M, Zeldovich N, Karger D, et al. (2014) Tackling the Challenges of Big Data.

41.   Mayer-Schönberger V, Cukier K (2013) Big : AdataRevolution That Will Transform How We Live, Work, and Think. Madrid: Turner.

42.   O'Reilly T (2010) The State of the Internet Operating System - O'Reilly Radar.

43.   Press G (2014) It's Official: The Internet Of Things Takes Over Big Data As The Most Hyped Technology.

44.   World Economic Forum (2012) Big Data, Big Impact: New Possibilities for International Development. World Economic Forum.