
REDUCING BARRIERS AND INCREASING UPTAKE OF TEXT AND DATA MINING FOR RESEARCH ENVIRONMENTS USING A COLLABORATIVE KNOWLEDGE AND OPEN INFORMATION APPROACH

# Deliverable 5.2

# Trend analysis, future applications and economics of TDM

# Project

| | |
|---|---|
| Acronym: | **FutureTDM** |
| Title: | Reducing Barriers and Increasing Uptake of Text and Data Mining for Research Environments using a Collaborative Knowledge and Open Information Approach |
| Coordinator: | SYNYO GmbH |
| Reference: | 665940 |
| Type: | Collaborative project |
| Programme: | HORIZON 2020 |
| Theme: | GARRI-3-2014 - Scientific Information in the Digital Age: Text and Data Mining (TDM) |
| Start: | 01. September, 2015 |
| Duration: | 24 months |
| Website: | http://www.futuretdm.eu/ |
| E-Mail: | office@futuretdm.eu |

Consortium:

**SYNYO GmbH**, Research & Development Department, Austria, (SYNYO)

**Stichting LIBER**, The Netherlands, (LIBER)

**Open Knowledge**, UK, (OK/CM)

**Radboud University**, Centre for Language Studies The Netherlands, (RU)

**The British Library**, UK, (BL)

**Universiteit van Amsterdam**, Inst. for Information Law, The Netherlands, (UVA)

**Athena Research and Innovation Centre in Information, Communication and Knowledge Technologies,** Inst. for Language and Speech Processing**,** Greece**,** (ARC)

**Ubiquity Press Limited**, UK, (UP)

**Fundacja Projekt: Polska**, Poland, (FPP)

# Deliverable

| | |
|---|---|
| Number: | **D5.2** |
| Title: | **Trend analysis, future applications and economics of TDM** |
| Lead beneficiary: | **FPP** |
| Work package: | WP5: ELABORATE: Legal framework, policy priorities, roadmaps and practitioner guidelines |
| Dissemination level: | Public (PU) |
| Nature: | Report (RE) |
| Due date: | 31.12.2016 |
| Submission date: | 22.12.2016 |


| | |
|---|---|
| Authors: | **Jan Strycharz**, FPP |
| Contributors: | **Camila R. Vieira**, FPP<br>**Brian Hole**, UP<br>**Frank Hellwig**, UP<br>**Burcu Akinci**, SYNYO |
| Review: | **Brian Hole**, UP |


**Acknowledgement:** This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 665940.

# Table of Contents

1 TDM as an Economic Asset – Definitions and Theory ..... 11
1.1 From Data to Value – Big Data and TDM Basics and Definitions ..... 11
1.2 Text and Data Mining Characteristics and its Business Connotation ..... 18
2 Company Perspective ..... 28
2.1 Introduction ..... 28
2.2 BDA process structure and its main challenges ..... 30
2.3 Adoption rates, patterns and goals ..... 38
2.4 Talents – Vital Component for TDM ..... 43
3 Market Value of Big Data and TDM ..... 55
3.1 Global and European Estimates ..... 55
3.2 Broader Perspective on Data Value ..... 62
3.3 Economic Impact ..... 63
4 Quantification of Big and Open Data Economic Impact ..... 65
4.1 Introduction ..... 65
4.2 Model Structure and Scope ..... 67
4.3 Big and Open Data Impact Assessment - Key Effects and Indicators in BOUDICA ..... 68
4.4 Research Findings ..... 73
4.5 Summary on TDM ..... 83
5 Literature ..... 85

# Table of Figures

Figure 1: Degrees of data openness ..... 13
Figure 2: Economic potential of Open Data across seven sectors globally ..... 14
Figure 3: Relations between different types of Data ..... 16
Figure 4: From Data to Action ..... 19
Figure 5: Ten Types of Innovations by Keeley et al. ..... 23
Figure 6: Fundamental Aspects of Big Data ..... 26
Figure 7: Analytical Value Chain ..... 29
Figure 8: Determining how to get value from BDA (2015) ..... 32
Figure 9: Obtaining skills becomes more challenging vis-à-vis technical issues ..... 33
Figure 10: Challenges of introducing Big Data and Analytics (2015) ..... 33
Figure 11: Discrepancy between access to and application of data ..... 34
Figure 12: Faith in Big Data Analytics ..... 35
Figure 13: Challenges in (B)D&A ..... 35
Figure 14: Levels of investment in different aspects of BDA ..... 37
Figure 15: Cultural obstacles to develop of BDA ..... 37
Figure 16: Top BDA challenges across US and EMEA (Europe, Middle East and Africa) ..... 38
Figure 17: Increased BDA uptake over time ..... 39
Figure 18: Stages of BDA Investment ..... 39

Figure 19: BDA Maturity stages ........ 41
Figure 20: BDA Maturity Stages in Detail ........ 42
Figure 21: Trends in using (B)D&A ........ 47
Figure 22: BDA Goals by Sector ........ 48
Figure 23: Recognition of BDA Value Across Industries ........ 48
Figure 24: BDA Goals ........ 49
Figure 25: What type of data do you use? ........ 50
Figure 26: Types of data used ........ 50
Figure 27: Types of data analyzed or planned ........ 51
Figure 28: Global Big Data Market Value ........ 56
Figure 29: Big Data Market by Type of Expenditure, 2012-2016 ........ 56
Figure 30: Types of Infrastructure Investments ........ 57
Figure 31: Global Spending on Big Data Infrastructure ........ 57
Figure 32: Types of Software Investments ........ 58
Figure 33: Global Spending on Big Data Related Software ........ 58
Figure 34: Subcategories of Software Expenditure ........ 59
Figure 35: Expenditure on Big Data Related Services ........ 60
Figure 36: Big Data Market Showing TDM Related Expenditure ........ 61
Figure 37: Big Data Market by World Region ........ 62
Figure 38: Share of Data Users by Country (total = 642 K) ........ 63
Figure 39: Key forces analyzed in BOUDICA ........ 66
Figure 40: ICT index values per Industry and Service (%) ........ 73
Figure 41: Additional GDP by Source of Impact - Economic potential of Big and Open Data (%) ........ 74
Figure 42: Impact of each force at the aggregate level in 2020 ........ 75
Figure 43: Additional value from each macroeconomic channel (2013-2020) – In € billion ........ 75
Figure 44: Forms of Value Extraction from Data ........ 76
Figure 45: Economic Impact of Big and Open Data by Sector (%) by 2020 ........ 78
Figure 46: Incremental GDP growth by country and sector (%) by 2020 ........ 80
Figure 47: Incremental GDP growth by country group by 2020 (in € billion) ........ 82

## Table of Tables

Table 1. Open Data Principles ........ 13
Table 2: Types of pattern detection methods ........ 20
Table 3: Process of translating data into action through TDM ........ 22
Table 4: Innovation elements ........ 26
Table 5: Challenges for Analytical Value Chain ........ 31
Table 6: TDM Spending on Software (2016) ........ 59
Table 7: Sectors Analyzed in BOUDICA ........ 67
Table 8: Country Groups in BOUDICA ........ 68
Table 9: Measures of Companies' Business Practices ........ 69
Table 10: Additional Data ........ 70
Table 11: ICT Index ........ 73

## Executive Summary

Text and Data Mining (TDM) has a vital role to play to ensure that Big Data exerts an economic impact. Gleaning value from vast informational resources has become an indispensable activity in an ever-expanding world of information. Private and public organizations are increasingly more interested in understanding how to make decisions based on sufficient and sound information to succeed in the global economy that is becoming more and more knowledge-based. By all means, the ability to glean proper and actionable intelligence from vast information sources is essential for market success, public policy impact and economic growth.

TDM is one of the most prominent techniques for extracting value from huge amounts of data. However, in order for it to work it has to be able to produce actionable intelligence on a constant basis for decision-makers to use. However, TDM in and of itself does not necessarily lead to actionable intelligence; its success greatly depends on human skills and capacities. Data miners should have an understanding not only of the data they compile, but also – even more importantly – of the context in which they operate to determine the right pieces of information to be mined and analyzed.

In business, data miners have to learn about their company and its operations – the more in-depth their knowledge is, the greater the chances for discovering useful patterns and making useful predictions. In this sense, the process of translating data into economic value needs to be strongly supported by the business side of a given organization – namely the people who are the most knowledgeable about how an organization creates, delivers and sells value. When crucial departments and decision-makers fail to engage in information interchange, this undoubtedly leads to one of the greatest organizational bottlenecks in the quest to develop a consistent analytical value chain.

Companies may innovate by tapping into Big Data to identify the many economic opportunities that lay ahead. Improvements in their response capacity and operations on the micro level can affect whole economies. According to Buchholtz et al. (2014), Big (and Open) Data will be responsible for a 1.9% increase in European GDP by 2020 and economic gains can be reaped through three types of micro level improvements, namely (1) better resource efficiency, (2) better decision-making through data-driven solutions and (3) improved products and processes as a result of R&D activities, monitoring and feedback.

Companies are struggling to derive value from TDM and Big Data. Companies' specific information shows that their success is largely driven by their ability to mobilize the right skills and talent, financial resources and leadership expertise to translate data into business and economic value. As Big Data is quickly becoming a mainstream process, the number of leaders who are struggling with the first stage of Big Data development, i.e. infrastructure building, is rapidly diminishing. As the challenge posed by building proper infrastructure becomes less pronounced, another challenge becomes more urgent, namely, talent acquisition because it is crucial for developing suitable TDM tools and methods. Without this magic ingredient, data (Big Data, too) continues to remain in a raw and unusable form.

Although data analysis is becoming more widespread, the shortage of talent is preventing companies from translating analytical insights into business actions. Even business giants have recognized that they have failed to exploit the full potential of their data due to a shortage of internal capabilities and their access to skilled analysts. This says much about the market's level of sophistication in integrating data analysis into the business decision-making process. Thus, it would be fair to claim that investments in the foreseeable future will focus on talent acquisition for TDM. Another key factor for companies to address is transforming organizational culture to make them more "data savvy" and to enable them to implement the work-products of data miners and analysts throughout their structures in a speedy way.

Increasing numbers of companies are experimenting with Big Data and TDM solutions, especially in the retail, communications and utilities sectors. Although the conversion deployment rate is equal to 60%, there is extensive room for improvement. The ability to harness infrastructure to tackle business problems and produce outcomes depends on using TDM tools appropriately so as to maximize their economic impact.

48% of all companies reported using text based data for their Big Data projects. Text mining is not as widely recognized as a source of business value compared to more structured data. However, its value grows when it is processed in (Big) Data & Analytics projects.

TDM's financial value is growing rapidly even though companies, generally speaking, continue to invest more in infrastructure and database maintenance than they do in developing analytical talents, tools and methods. However, they are starting to feel considerable demand for talent and a need to alter their organizational culture, a purpose that also demands proper training. 2016 estimates suggest that a total of $23.8 billion will be spent globally on the Big Data Market, with $6.4 billion of that figure for TDM (software purchases, support and training). In the European Big Data Market, it is estimated that TDM will be worth c.a. $2.5 billion in 2016 juxtaposed against the total expected market value of $9.37 billion for the same year. European TDM may be expected to grow rapidly to $10.3 billion in 2021.

Data value extraction using TDM may exert a significant impact on the economy as a whole, depending on the capacity of countries, sectors and more specifically, companies to deploy ICT innovations and create internal processes to facilitate wider business efficiency, decision-making improvements and entry costs reduction. According to the BOUDICA model €1 spent on Big Data products and solutions translates into €10.7 of overall economic value. This also implies that the economic impact exerted by Big Data is estimated to be over twice as powerful as the economic impact exerted by more traditional data (whose ratio ranges from 1 to 5.2). Taking all this into account one could claim that TDM's impact on the European Economy may range from $13 billion (conservative calculation) to $26.7 billion (optimum scenario) in 2016. This impact may rise to $110.1 billion in 2020.

# INTRODUCTION AND REPORT OUTLINE

This study strives to present the economic perspective of Text and Data Mining (TDM). As the study's title indicates, its purpose is to accomplish the following:

1) portray current and future trends in TDM's development and usage as an economic asset;
2) depict and clarify market dynamics surrounding TDM solutions (global and European);
3) estimate TDM's global and European economic impact.

The study focuses on Text and Data Mining treated as an economic asset and a business practice. The authors attempt to respond to questions concerning how business actors understand, approach and utilize text and data mining as a tool, its current scope and, thus, its microeconomic and macroeconomic impact.

It has to be noted up front, however, that very few existing studies treat TDM directly. As we will explain in detail, TDM is part of a bigger phenomenon initially referred to as "Big Data". In addition, it is important to note that the term Big Data is used in a business setting in two ways – (1) as a descriptor of big data sets that are unmanageable by purely human analytics and (2) as a metaphor for all business processes that support the usage of Big Data through a process of conversion into business value. In the first instance, it usually means that the data set is being developed and maintained in accordance with the three Vs (Buchholtz 2014: 10): (1) high **Volume**, (2) **Various** sources, (3) high **Velocity** data updates.[1] The development of Big Data storage technologies like, for instance, Hadoop[2], has even inspired the usage of the term data "lake" instead of "dataset" to imply the new Big Data approach to data (see box below).

**Box 1. What is a Data Lake**

"A data lake is a storage repository that holds a vast amount of raw data in its native format until it is needed. While a hierarchical data warehouse stores [classical approach] data in files or folders, a data lake uses a flat architecture to store data [Big Data approach]. Each data element in a lake is assigned a unique identifier and tagged with a set of extended metadata tags. When a business question arises, the data lake can be queried for relevant data and that smaller set of data can then be analyzed to help answer the question.

The term data lake is often associated with Hadoop-oriented object storage. In such a scenario, an organization's data is first loaded into the Hadoop platform and then business analytics and data mining tools are applied to the data where it resides on Hadoop's cluster nodes of commodity computers." (Rouse M., *Data Lakae*, TechTarget.com, accessed on http://searchaws.techtarget.com/definition/data-lake 23rd August 2016)

Thus, to explain further, the Big Data approach to data is to maintain flat data storage in which each raw piece of data is assigned a specific set of tags. This approach allows for mining all stored data in accordance with the problem that is currently needs to be solved. How the technologies used for maintaining Big Data sets are called provides valuable insight into there being much more to Big Data

[1] The definition of Big Data is constantly evolving as we will convey below. According to the authors of this report it is currently impossible to present one coherent definition of Big Data. This is why we choose to present different aspects of Big Data in a longer narrative to reflect its complexity.
[2] "Apache Hadoop is an open source software platform for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware. Hadoop services provide for data storage, data processing, data access, data governance, security and operations." (http://hortonworks.com/apache/hadoop/)

than just storage. That is why the term Big Data is also used in a much broader way in a business setting – it is used as a metaphor for all investment and activities that need to take place in a given organization so that the development and maintenance of Big Data sets/lakes can produce value for a given company.

In our opinion, TDM is an important technique for value extraction that is part of the broader story on how companies can capitalize on Big Data. This way of understanding TDM is the underlying foundation for the economic analyses presented in this report. Therefore, most of the research we use for this analysis is related, in one way or another, to the study of Big Data – as this is how most researchers approach this topic. However, we also use this research in an attempt to grasp TDM more closely in an economic context.

The report begins with a more theoretical background – before we present our findings from business research and economic modeling we would like to clarify what we believe to be important definitional aspects by describing in detail how TDM fits into the broader discussions on Big Data. While doing this, we also explain how TDM is an important phenomenon for innovation. Thus, in the first part of this report we present theoretical argumentation supporting the thesis that the economic promise portended by the Big Data "Revolution" relies mainly on two aspects – (1) construction and development of suitable and interlinked databases and (2) mining and analytical activities (essentially TDM). These two aspects are of course interlinked. At the end of the first part, we briefly recap our quantified analyses of the economic impact exerted by Data, which will be developed into more detailed estimates of TDM's economic impact on the European Economy in chapters three and four.

After handling the definitional and theoretical foundations we proceed to present an analysis of Big Data research done on companies. We believe that the theoretical research depicted in part one presents a clear overview of current business trends in Big Data because it reveals the empirical role TDM plays within the bigger Big Data story. This section points to the main obstacles and how to overcome them. It also allows us to understand what Big Data (and, thus, TDM) is being used for specifically and by which industries for the most part; and what types of data are being stored, mined and analyzed. Furthermore, we present the patterns according to which companies develop as data-based enterprises as they undertake more extensive TDM activities as an everyday business practice.

In the third chapter we share the existing financial estimates of Big Data markets – globally and on the European level. To be able to capitalize on Big Data, organizations invest in suitable technologies, services, talents, etc. In some respects, they develop these capabilities in-house. Very often, however, they outsource them, thereby driving external demand. This has forged a Big Data market that is a Business-to-Business (B2B) market. This market has grown considerably in the last several years and is one of the most promising segments of the IT market. By analyzing its growth rate and other attributes we may estimate TDM's relative financial importance compared to Big Data.

The last part of this report discusses TDM's impact on the overall economy. Organizations invest in TDM (and other Big Data solutions) because they hope it will allow them to create distinct value. Many cases demonstrate that Big Data generates a genuine return on investment thereby being conducive to innovation, efficiency and generating greater competitive advantage. These macroeconomic effects translate into an macro impact. In Chapter 3 we attempt to account for TDM's macroeconomic impact.

To estimate its impact on the economy as a whole requires economic modeling. Importantly, we have not devised an original model to make these calculations; instead, we rely on the most up-to-date model created by the Warsaw Institute for Economic Studies to estimate the economic impact exerted by what the authors refer to as Big and Open Data on the European economy. We expand on this model through a tool known as BOUDICA (Big and Open Data Universal Impact Assessment). We find it to be the most up-to-date and most accurate model. In addition, it is the only model in existence that specifically addresses the European economy. Again, it is worthwhile to note, that this model treats Big Data as a whole (it also incorporates the impact exerted by Open Data) and we will use it to estimate the role played by TDM within the confines of the economic impact exerted by Big Data. Because of the importance we place on the BOUDCIA model in estimating TDM's economic impact, we describe this model in detail in Chapter 4. Additionally, the model constitutes an appendix to this report.

This study may lead to policy recommendations to further the goals set in Europe's 2020 Strategy. Especially, it would aid in achieving "smart growth" in reference to "developing an economy based on knowledge and innovation" (Europe 2020: 10). TDM, as we explain in detail, is often seen as part of Research and Development (R&D). Therefore, extending TDM's commercial applications will mean boosting corporate R&D spending, thereby helping Europe to meeting its "3% of GDP" goal. More importantly, it will mean that the products and services developed by European companies are more firmly rooted in methodical approaches to collecting and analyzing the world's information.

This study intends to provide as much specific information on how TDM is currently understood, implemented and developed by economic actors. This knowledge may be used to formulate policy recommendations to strengthen TDM's uptake and development and make it more economically viable. We must also reiterate that we consider TDM to be a subset of Big Data. This means that our findings may also serve to devise policies to strengthen Big Data.

# 1 TDM AS AN ECONOMIC ASSET – DEFINITIONS AND THEORY

## 1.1 From Data to Value – Big Data and TDM Basics and Definitions

Text and data mining may be understood to be an activity to derive value from vast information resources. In this sense, it may be seen as an economic resource that has the potential to boost economic performance at the micro and macro levels. However, this value materializes only if TDM leads to practical implications.

**Box 2. Practical illustration of TDM – Comparison Websites**

Price comparison websites are a perfect example of applying TDM. Consumers type in a specific product name and the site shows many different retailers' list prices. The underlying search engine is usually driven by an algorithm that uses a specific bot to mine the Web for product related data (textual, numerical and visual). This allows customers to buy less expensively. Their informed decision means that financial resources are used in a more efficient way. On a social level, it in theoretically enhances a given household's financial well-being. On a macroeconomic level, the money saved by consumer is used to boost other economic development, either through consumption or direct/indirect investment.

Text and Data Mining has become virtually indispensable in today's information-rich world. In developed countries information access appears not to pose any obstacles. However, if information access is no longer as challenging as in previous decades, then the ability to glean a particular piece of information to achieve a specific practical goal becomes problematic. Matching information and application is at stake.

*"Knowledge and information tend to be abundant; what is scarce is the capacity to use them in meaningful ways." (OECD 1996: 11)*

TDM in its generic form could mean virtually any activity aimed at finding and analyzing a specific piece of information. However, in the age of ever greater data and technological capabilities, TDM assumes a special meaning, more technological in nature. TDM becomes vital as a technique for extracting important piece(s) of information, especially when working with big data sets and ever-expanding pools of data that are often unstructured. This is why, as the box below explains Data Mining and Big Data are two separate elements of a very complex process aiming at the extraction of value from ever expanding pools of data.[3]

[3] It is worthwhile to note, though, that – as we mention in the introduction – in a business context the term "Big Data" is also informally used as a metaphor encompassing all business processes to translate the big data accessible to a given company into business value. Used in this way the term encompasses both Big Data sets and other activities such as TDM, strategy development and investment in organizational culture.

**Box 3. Big Data Versus Data Mining**

"Big data and data mining are two different things. Both of them relate to the use of large data sets to handle the collection or reporting of data that serves businesses or other recipients. However, the two terms are used for two different elements of this kind of operation.

Big data is a term for a large data set. Big data sets are those that outgrow the simple kind of database and data handling architectures that were used in earlier times, when big data was more expensive and less feasible. For example, sets of data that are too large to be easily handled in a Microsoft Excel spreadsheet could be referred to as big data sets.

Data mining refers to the activity of going through big data sets to look for relevant or pertinent information. This type of activity is really a good example of the old axiom "looking for a needle in a haystack." The idea is that businesses collect massive sets of data that may be homogeneous or automatically collected. Decision-makers need access to smaller, more specific pieces of data from those large sets. They use data mining to uncover the pieces of information that will inform leadership and help chart the course for a business.

Data mining can involve the use of different kinds of software packages such as analytics tools. It can be automated, or it can be largely labor-intensive, where individual workers send specific queries for information to an archive or database. Generally, data mining refers to operations that involve relatively sophisticated search operations that return targeted and specific results. For example, [in its simplest form] a data mining tool may look through dozens of years of accounting information to find a specific column of expenses or accounts receivable for a specific operating year.

In short, big data is the asset and data mining is the "handler" of that is used to provide beneficial results." It is also important to mention that data itself may come from large amount of sources which also complicates the particularity of access itself.

*Source: Techopedia, What is the difference between big data and data mining?, Retrieved from www on 18th July 2016,* https://www.techopedia.com/7/29678/technology-trends/what-is-the-difference-between-big-data-and-data-mining

It is, also, important, to clarify some factors related to definitions. The adjective "Big" is very frequently used colloquially in private companies. It usually means data generated by customers in their interactions with a particular company. It can, however, also mean data generated in a given company's value chain. Another important part of the chain is production: as more digital tools are used in production, more data is generated. Understanding data can lead to practical implications for optimizing production processes.

However, the term Big Data is more generic. As stated by McKinsey (2013: 1):

*"Big data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze. This definition is intentionally subjective and incorporates a moving definition of how big a dataset needs to be in order to be considered big data—i.e., we don't define big data in terms of being larger than a certain number of terabytes (thousands of gigabytes). We assume that, as technology advances over time, the size of datasets that qualify as big data will also increase. Also note that the definition can vary by sector, depending on what kinds of software tools are commonly available and what sizes of datasets are common in a particular industry. With those caveats, big data in many sectors today will range from a few dozen terabytes to multiple petabytes (thousands of terabytes)."*

Big Data is also defined by using the 3 Vs:

*"The concept of big data is usually defined by the "three Vs": volume, velocity and variety. In many areas volumes of available facts are higher than ever before, they are also expanding quicker, come*

*from many more sources and materialize in many different forms than small, well-structured data sets from the past. With the growing use of big data in business, an additional "V" became important – veracity. In contrast with the three original Vs, it refers not to the intrinsic characteristics of big data itself but rather to the quality which makes it useful for practical applications."* (Buchholtz et. al., 2014: 10)

Big Data is generated not only by private companies but also by other organizations such as public and non-governmental organizations. Therefore, Big Data has different owners, public and private alike.

Importantly, Big Data may be closed (when kept secret by its owner) or open to a varying degree. Economic openness, for instance, would suggest that data is available without imposing monetary constraints on the user (e.g. open web content). To be more open, however, data should also be free of any legal constraints: it should be published in a way enabling reuse for any purpose (for commercial or non-commercial applications) without permission from right holders. The third aspect of openness refers to the technical side of publishing data. As suggested in this paper, the full value of Big Data is only extracted by linking many datasets. It is thus of vital importance that data be published in machine-readable formats[4]. Otherwise, it cannot be processed by bots, or algorithms (Buchholtz et. al. 2014: 7). Also, it is important that data be well described by applying the correct metadata (or correct tags, as in data lakes) – otherwise, data linking cannot be performed easily (linkage is usually done by linking specific categories of data).


**Figure 1: Degrees of data openness**

One can also present slightly more nuanced conditions for data to meet the criteria of full openness. The table below summarizes the principles developed by, *inter alia*, prof. Lawrence Lessig (the initiator of the Creative Commons license). These principles address governments with the goal of making all Public Sector Information (PSI) open.

**Table 1. Open Data Principles**

| **Principle** | **What does it mean?** |
|---|---|
| *Complete* | All data is accessible |
| *Primary* | Data is collected at source with the highest possible level of granularity |
| *Timely* | Made available as quickly as possible and updated |
| *Accessible* | Data is available to the widest range of users for the widest range of purposes |
| *Machine* | Data is reasonably structured to allow automated processing |
| *Non-* | Data is available to anyone with no requirement of registration |

[4] "Machine readable data is data in a data format that can be automatically read and processed by a computer, such as CSV, JSON, XML, etc. Machine-readable data must be structured data. […]
Non-digital material (for example printed or hand-written documents) is by its non-digital nature not machine-readable. But even digital material need not be machine-readable. For example, consider a PDF document containing tables of data. These are definitely digital but are not machine-readable because a computer would struggle to access the tabular information - even though they are very human readable. The equivalent tables in a format such as a spreadsheet would be machine readable.
As another example scans (photographs) of text are not machine-readable (but are human readable!) but the equivalent text in a format such as a simple ASCII text file may be machine readable and processable." (http://opendatahandbook.org/glossary/en/terms/machine-readable/)

| | |
|---|---|
| *Non-proprietary* | Data is available in a format over which no entity has exclusive control |
| *License-free* | Data is not subject to any copyright, patent, trademark or trade secret regulation. Although reasonable privacy, security and privilege restrictions may be allowed |

Source: https://opengovdata.org/ (retrieved 26th April 2016)

There is substantial value to be produced by linking private company data with PSI – this is only possible when the latter meets the given criteria of openness (see quotation below as well as Box 4).

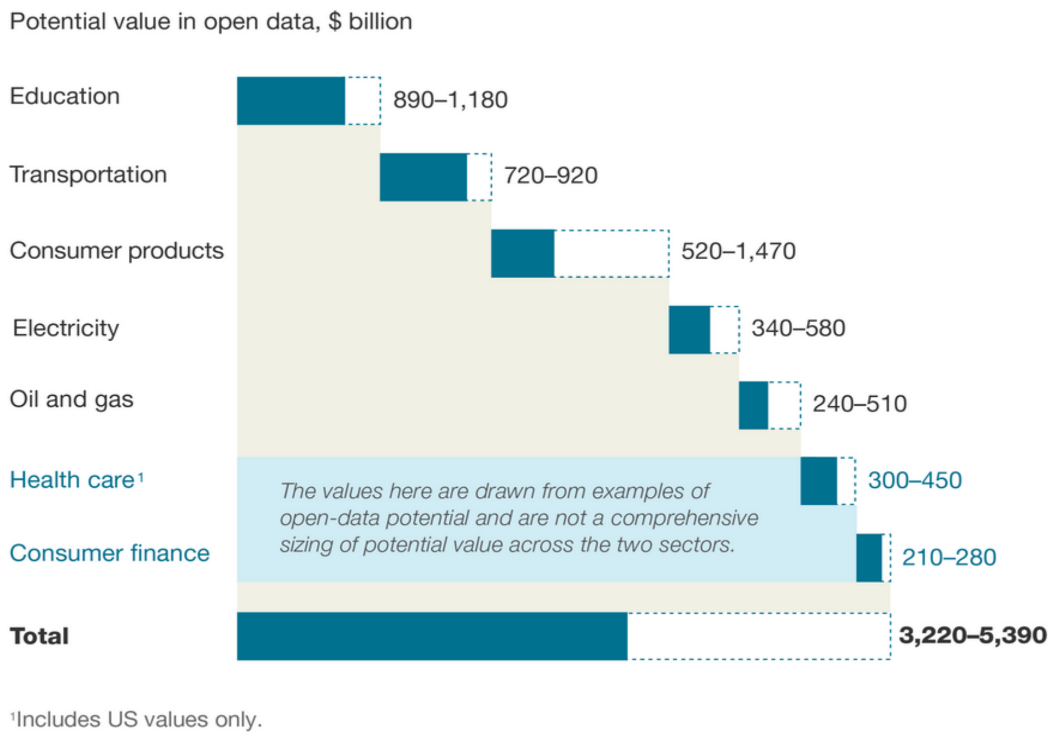


**Figure 2: Economic potential of Open Data across seven sectors globally[5]**

*"Although the open-data phenomenon is in its early days, we see a clear potential to unlock significant economic value by applying advanced analytics to both open and proprietary knowledge. Open data can become an instrument for breaking down information gaps across industries, allowing companies to share benchmarks and spread best practices that raise productivity. Blended with proprietary data sets, it can propel innovation and help organizations replace traditional and intuitive decision-making approaches with data-driven ones. Open-data analytics can also help uncover consumer preferences, allowing companies to improve new products and to uncover anomalies and needless variations. That can lead to leaner, more reliable processes."* (McKinsey 2014)

[5] McKinsey, Open data: Unlocking innovation and performance with liquid information, October 2013, McKinsey Global Institute. Retrieved 9th May 2016

**Box 4 . Importance of external data for innovation - case of Sytnhos**

For the purpose of picturing how companies need a sound Research Infrastructure (also involving open data infrastructure), we present a brief case study of Synthos - one of the largest manufacturers of chemical raw materials in Poland being, also, the first European manufacturer of emulsion rubbers and a leading manufacturer of polystyrene for foaming applications.

Synthos is a chemical company with a well established global footprint, especially in rubber (it works with global brands such as Goodyear and Pirelli). Synthos operates in two main sites in southern Poland and the Czech Republic. What is important is that this company focuses its growth strategy on innovation.

"The strategy of Synthos S.A aims at establishing the company's value by upgrading innovation and marketing modern and technologically sophisticated products. R&D focuses on three strategic sectors: synthetic rubbers, expanded polystyrene and dispersion, adhesives and latex." (Synthos website)

For this reason, Synthos established its own R&D Center at the end of 2009. It took 4 years for it to achieve its first outcomes underlining how long-term and complex innovation projects are.

"We worked on multiple projects in 2014. Five of them resulted in marketing new products. Others will be gradually marketed in coming years." (Ibidem)

Synthos R&D Center employs c.a. 50 employees with 25% of them holding a doctoral degree in Chemistry. The Center has a proprietary market intelligence unit that tracks new trends and developments in the markets of interest to Synthos. This unit develops leads and ideas for research projects that are then screened by the Center's director and talked over with the Management Board. After management screening, decisions about kicking off a specific research project are taken.

This company does in-house research and collaborates with universities. It aims to collaborate with university-based researchers with greater knowledge. R&D Management states it is hard to identify such researchers. Is is done mainly through traditional human intelligence gathering – browsing the web and gathering leads when participating in conferences and other less formal meetings. This intelligence unit also actively follows the patents obtained by chemical researchers and Synthos's competition. All this work is done "manually" without harnessing any algorithmic power, or with only limited amounts thereof.

We cite this case because tapping into the power of open data mining and analytics could boost the work of this intelligence unit. Free and open access to many scientific and patent databases would allow it to streamline its work. Especially Text Mining could prove conducive to Synthos's R&D – effective mining of research articles, for example, could help identify suitable researchers whose knowledge or IP could boost the company's research without carrying out expensive experimental R&D phases. Open Science Data could save corporate resources.

It is also clear, however, that harnessing the power of third party open data does not equate to providing suitable datasets for easy access by the company's intelligence workers. More importantly, it entails a change in skills and learning and decision-making culture in the intelligence unit and management.

* This brief study was developed based on two In-depth interviews carried out by the authors with individuals familiar with Synthos's R&D Center and its Intelligence Unit. The interviews were carried out in 2014 as a part of the consultancy project for the local government of Malopolska (a region in southern Poland). The project was concerned with the Knowledge Transfer between Cracow based Jagiellonian University's Chemistry Department and the region's chemical companies. This qualitative information was reused for the purpose of producing this brief case study.

Although the term Open Data is usually – on a colloquial level of understanding – associated more with public sector information (whether Big or not) and the academic sector that publishes research data in a more open manner, private companies also make their data open to various degrees (see Box below).

**Box 5. Opening Data by Private Companies**

Facebook provides developers with APIs (Application Programming Interfaces) enabling them to access, use and reuse some data collected by Facebook - https://developers.facebook.com/. By doing this Facebook facilitates the interaction between other applications that are being developed through interaction with the Facebook environment. This means that Facebook inspires the development of a plethora of applications and, thus, businesses that exist only because of Facebook's open data . In this sense Facebook is a stronger catalyst of economic development than it would have been without allowing others to make use of their data infrastructure.

It has to be noted, however, that Facebook does not provide its data without any legal constraints. Strict rules apply as to how the data can be used: https://developers.facebook.com/policy/. Thus, this data is only open to a certain degree and this is the approach championed by most commercial actors.

Thus, publicly accessible data published by private and public entities create an environment in which data is ubiquitous. The vastness of existing information leads many researchers and analysts to proclaim that the age in which we live is the information age. Below we present a diagram developed by McKinsey to illustrate different types of data and the relationships between them as well as the complexity of the information-rich age.

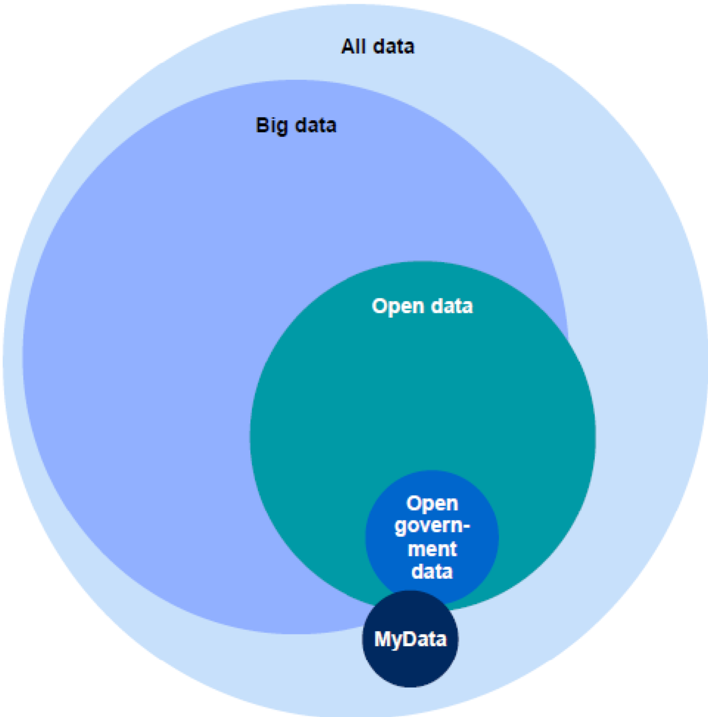


**Figure 3: Relations between different types of Data**[6]

Big Data is understood here in its narrow sense: it refers rather to a characteristics of the sets that are huge in **Volume**, include an opulent **Variety** of sources and are updated at a fast pace of **Velocity** (the

[6] McKinsey (2013: 4)

three Vs). Open Data is information that is accessible publicly. It can have various levels of openness as described above. Public Sector information published openly and meeting openness criteria may be labeled Open Government Data. It is also worthwhile to remember that private companies and individuals also publish data in an open manner. McKinsey also identifies the concept of "MyData" "which involve sharing information collected about an individual (or organization) with that individual"[7] (McKinsey 2013:4).

In the information age economies tend to become more knowledge-based – where "production and services [are] based on knowledge-intensive activities" (Powell and Snellman 2014: 201). What it means is that "[t]he key component of a knowledge economy is a greater reliance on intellectual capabilities than on physical inputs or natural resources" (*Ibidem*). Hence, to succeed in the knowledge economy one has to make decisions basing them on sound information. Even more – the ability to derive adequate information determines whether one succeeds in the economy. In other words, the ability to crystallize proper and actionable intelligence (see Box below) from abundant information sources is of vital importance to market success, public policy impact and economic growth.

**Box 6. Actionable intelligence explained**

Understanding the difference between "information" and „actionable intelligence" is instrumental in understanding what role TDM plays in the wider Big Data picture and how it is vital in safeguarding the economic impact of data. "Information" is a more generic term that refers to virtually any fact or opinion that can be expressed either textually or numerically.

"Actionable Intelligence", on the other hand, refers to the world of actions that relate one to goals, deliverables, achievements and so on. The concept of "actionable intelligence" points to the fact that not all information can be used at any given time by anyone. For a piece of information to have qualities of an actionable intelligence it has to be delivered in an adequate form to a person in the specific context whose work is to advance some kind of process in order to achieve a specific goal. Farmer and Carter (2014) define the concept by giving the following example:

"For a concrete example of what actionable intelligence can do, consider this quote from a product director in a large fast-moving consumer goods company:

*We've been carefully tracking an important product launch, worth $55.6 million in the next 12 months. The actionable intelligence tool showed me the global forecast increased by 160,000 pieces globally for the next six months in the last four weeks alone, a $10 million increase. This information enables us to stay ahead of the demand increases and proactively coordinate a response, reducing costs and ensuring we don't go out of stock and avoiding airfreight.*" (Farmer and Carter 2014: xv)

The challenge in translating information into actionable intelligence lies mostly in the fact that the latter is context specific and, thus, it has no specific characterization. For, one never knows which kind of insight will give one concrete grounds for moving forward with a specific decision.

To sum up – to make information useful, each time it has to be processed and transformed into actionable intelligence. Otherwise, its power to move the world of organizations and processes is only a potential one.

[7] "For example, some hospitals now provide individual patients with access to their own medical records. Providing aggregate statistics (a form of open data) alongside MyData allows useful comparisons; some utilities show consumers how their energy use compares with that of neighbors to encourage conservation. In some cases, individuals are allowed to modify or correct the data provided to them, which improves the quality of the data." (*Ibidem*)

What complicates the picture is that, as it turns out, “"mining" text or data is not a very good metaphor for what people in the field actually do. Mining implies extracting precious nuggets of ore from otherwise worthless rock. If data mining really followed this metaphor, it would mean that people were discovering new factoids within their inventory databases. However, in practice this is not really the case. Instead, data mining applications tend to be a (semi)-automated discovery of trends and patterns across very large datasets, usually for the purposes of decision making” (Fayyad and Uthurusamy, 1999; Fayyad, 1997 as quoted in Hearst (1999)).

We develop the characterization of mining in more depth further in this chapter. What we believe to be crucial to appreciate at this stage is that capitalizing on Big Data is not about finding the proper piece of information, but finding many pieces that may be synthesized into something meaningful for improving a practical solution. Perhaps, reflecting this inadequacy, instead of the term data “mining” often the term data “analytics” or “analysis” is used. In fact, many practitioners claim that data mining and analysis (when applied to Big Data) mean the same thing and are done by the same people – or, the practical distinction between mining and analyzing cannot be established.

## 1.2 Text and Data Mining Characteristics and its Business Connotation

All the above corresponds with the idea that information is not “easily transformed into the object of standard economic transactions” (OECD). In other words, to translate it into an economic asset, various processes need to take place so that the information is translated into actionable intelligence (or, at least, makes this translation possible). The activity of Text and Data Mining, in this context, is vital and instrumental in situations where the source of information is too big to handle in any foreseeable time prospect by purely human effort.

*“The sheer scale of […] data has far exceeded human sense-making capabilities. At these scales patterns are often too subtle and relationships* ***too complex or multi-dimensional to observe by simply looking at the data****. Data mining is a means of automating part this process to detect interpretable patterns; it helps us see the forest without getting lost in the trees.”* (Furnas, 2012)

The importance of TDM is seen even more clearly when one takes into account how Big Data practices and the ideas behind them begin to evolve. It seems, that “with the growing use of big data in business, an additional “V” has become important – veracity. In contrast with the three original Vs [Volume, Variety, Velocity], it refers not to the intrinsic characteristics of big data itself but rather to the quality which makes it useful for practical applications.” Hence, the “veracity” criterion alerts attention to the fact that when treating Big Data as a resource, one has to take into account not only its quality but also its adequacy, which in turn may influence its potential to be translated into actionable intelligence. In turn, the data miner must determine whether he or she is mining and analyzing the right pieces of information. The importance of miners is starting to overshadow the importance of infrastructure as “Big Data” as a business discipline matures.

*“In 2016, companies will move away from irrelevant data noise, acknowledge that the variety and speed of data can be daunting and will take a more thoughtful approach to analyzing “useful” data to reach fast, meaningful, holistic insights. Rather than investing time and money in IT infrastructure to manage high volumes of data, the trick will be managing the data diversity and speed at which data streams to glean valuable insights and to do something worthwhile with them.” (Gutierrez 2015)*

TDM facilitates the extraction of useful and instrumental pieces of information from typically large corpora of essentially unstructured text and other types of data; it also allows for the translation of this information into actionable intelligence for advancing a specific process – be it public policy intervention, market actions or actions performed by other entities for various reasons.

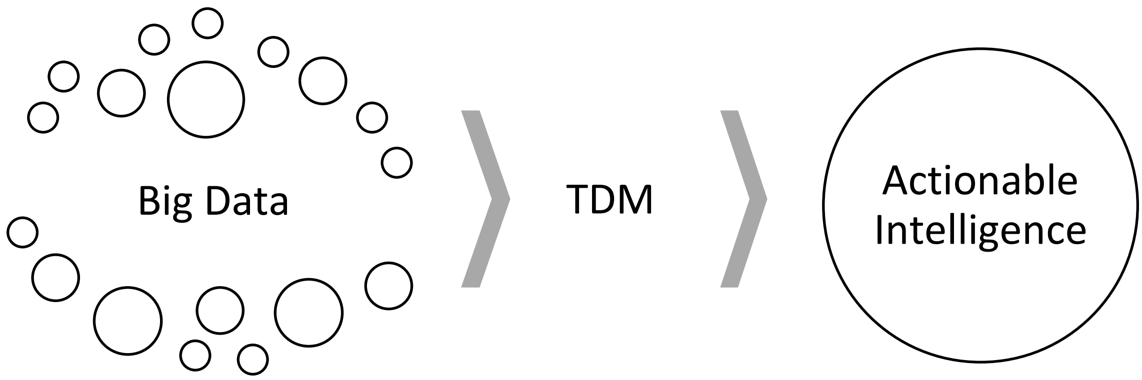


**Figure 4: From Data to Action**

The ways in which data miners work are as in any creative line of work hard to standardize. However, “The two "high-level" primary goals of data mining, in practice, are prediction and description.” (Fayyad *et al.* 1996)

*“Discovering information from data takes two major forms: description and prediction. At the scale, we are talking about, it is hard to know what the data shows. Data mining is used to simplify and summarize the data in a manner that we can understand and then allow us to infer things about specific cases based on the patterns we have observed.”* (Furnas *op cit.*)

This means applying a plethora of statistical methods – we present several main types of pattern detection in the table below linking them to real world business applications.

| Type of pattern detection | Description |
|---|---|
| Anomaly detection | In a large data set it is possible to get a picture of what the data tends to look like in a typical case. Statistics can be used to determine if something is notably different from this pattern. For instance, the tax administration could model typical tax returns and use anomaly detection to identify specific returns that differ from this for review and audit. |
| Association learning | This is the type of data mining that drives the Amazon recommendation system. For instance, this might reveal that customers who bought a cocktail shaker and a cocktail recipe book also often buy martini glasses. These types of findings are often used for targeting coupons/deals or advertising. Similarly, this form of data mining (albeit a quite complex version) is behind Netflix movie recommendations for specific users based on their earlier history of viewings. |
| Cluster detection | One type of pattern recognition that is particularly useful is recognizing distinct clusters or sub-categories within the data. Without data mining, an analyst would have to look at the data and decide on a set of categories which they believe captures the relevant distinctions |

| | |
|---|---|
| | between apparent groups in the data. This would risk missing important categories. With data mining it is possible to let the data itself determine the groups. This is one of the black-box type of algorithms that are hard to understand. But in a simple example - again with purchasing behavior - we can imagine that the purchasing habits of different hobbyists would look quite different from each other: gardeners, fishermen and model airplane enthusiasts would all be quite distinct. Machine learning algorithms can detect all of the different subgroups within a dataset that differ significantly from each other. |
| Classification | If an existing structure is already known, data mining can be used to classify new cases into these pre-determined categories. Learning from a large set of pre-classified examples, algorithms can detect persistent systemic differences between items in each group and apply these rules to new classification problems. Spam filters are a great example of this - large sets of emails that have been identified as spam have enabled filters to notice differences in word usage between legitimate and spam messages and classify incoming messages according to these rules with a high degree of accuracy. |
| Regression | Data mining can be used to construct predictive models based on many variables. Facebook, for example, might be interested in predicting future engagement for a user based on past behavior. Factors like the amount of personal information shared, number of photos tagged, friend requests initiated or accepted, comments, likes etc. could all be included in such a model. Over time, this model could be honed to include or weight things differently as Facebook compares how the predictions differ from observed behavior. Ultimately these findings could be used to guide design in order to encourage more of the behaviors that seem to lead to increased engagement over time. |
| The patterns detected and structures revealed by descriptive data mining are then often applied to predict other aspects of data. Amazon offers a useful example of how descriptive findings are used for prediction. The (hypothetical) association between cocktail shaker and martini glass purchases, for instance, could be used, along with many other similar associations, as part of a model predicting the likelihood that a particular user will make a particular purchase. This model could match all such associations with a user's purchasing history and predict which products they are most likely to purchase. Amazon can then serve ads based on what that user is most likely to buy. | |

**Table 2: Types of pattern detection methods[8]**

However, by definition, TDM does not necessarily lead to Actionable Intelligence, a condition for making data usable in a business context. This issue was indirectly addressed by Fyyad *et al.* in 1996 (:39) when they stated that the "[b]lind application of data-mining methods (rightly criticized as data dredging in statistical literature) can be a dangerous activity, easily leading to the discovery of meaningless and invalid patterns." The authors coined the term "Knowledge Discovery in Databases" which they explain in the following way:

[8] Furnas, 2012.

*"In our view, KDD refers to the overall process of discovering useful knowledge from data and data mining refers to a particular step in this process. Data mining is the application of specific algorithms for extracting patterns from data. The distinction between the KDD process and the data-mining step (within the process) is a central point of this article. The additional steps in the KDD process, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge and proper interpretation of the results of mining, are essential to ensure that useful knowledge is derived from the data." (Ibidem)*

In the business (or organizational in general) context we equate "useful knowledge" with "actionable intelligence". In other words, useful knowledge can be used for action. Therefore, just like specific conditions have to be met in order to be able to secure the finding of useful knowledge, the same applies to TDM understood as tool for developing actionable intelligence. First, miners must understand not only the data they use but, even more importantly, the context in which they operate. In business this usually means familiarity with the company and its operations: the more in-depth this knowledge is, the greater the chances of discovering useful patterns and making useful predictions. The lack of mutual understanding between data miners and a given company's strategic decision-makers is one of the most prominent obstacle to developing Big Data projects (we develop this in more detail in Chapter 2).

After the context is understood miners can start making decisions about the data they should take into focus and start cleaning and preprocessing it. Then, they may proceed to shed unimportant data, choose the appropriate data mining method, start exploratory analysis, search for patterns and start interpreting them. The final interpretation should be actionable to make a decision to solve a given problem. What is important is that this process (summarized in the table below) may have many iterations and loops between any two steps as the process itself is highly creative and requires the miner to possess a high level of complex and interdisciplinary sensitivity (*Ibidem*: 42).

| Step tag | Explanation of the process |
|---|---|
| Understanding | First is developing an understanding of the application domain and the relevant prior knowledge and identifying the goal of the KDD process from the practical problem viewpoint. |
| Selecting | Second is creating a target data set: selecting a data set, or focusing on a subset of variables or data samples, on which discovery is to be performed. |
| Preprocessing | Third is data cleaning and preprocessing. Basic operations include removing noise if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields and accounting for time-sequence information and known changes. |
| Transformation | Fourth is data reduction and projection: finding useful features to represent the data depending on the goal of the task. With dimensionality reduction or transformation methods, the effective number of variables under consideration can be reduced, or invariant representations for the data can be found. |
| Choosing the method | Fifth is matching the goals of the KDD process (step 1) to a particular data-mining method. For example, summarization, classification, regression, clustering and so on (see Table 1). |

| | |
|---|---|
| Exploring | Sixth is exploratory analysis and model and hypothesis selection: choosing the datamining algorithm(s) and selecting method(s) to be used for searching for data patterns. This process includes deciding which models and parameters might be appropriate (for example, models of categorical data are different than models of vectors over the reals) and matching a particular data-mining method with the overall criteria of the KDD process (for example, the end user might be more interested in understanding the model than its predictive capabilities). |
| Discovering patterns | Seventh is data mining: searching for patterns of interest in a particular representational form or a set of such representations, including classification rules or trees, regression and clustering. The user can significantly aid the data-mining method by correctly performing the preceding steps. |
| Interpreting | Eighth is interpreting mined patterns, possibly returning to any of steps 1 through 7 for further iteration. This step can also involve visualization of the extracted patterns and models or visualization of the data given the extracted models. |
| Acting | Ninth is acting on the discovered knowledge: using the knowledge directly, incorporating the knowledge into another system for further action, or simply documenting it and reporting it to interested parties. This process also includes checking for and resolving potential conflicts with previously believed (or extracted) knowledge. |

**Table 3: Process of translating data into action through TDM**[9]

All these steps in the process point to necessary conditions for TDM to offer value in business projects. Pattern recognition and prediction have to take place within a deep understanding of the complexity of a specific business model. This necessarily has to mean that the business knowledge of data miners has to be updated and aligned to strategic decisions constantly. When they are properly linked, solutions may be mined for thousands of specific business problems to hone greater competitive advantage.

However, the application of TDM may go far beyond what we are observing currently. From a theoretical and a more abstract, perspective, TDM may lead to improvements (or innovations) in (1) business configuration, (2) offering and/or (3) customer experiences (Keely *et al.* 2013). The figure presented below helps grasp what type of improvements companies are pursuing.

[9] Fyyad et al., 1996: 42.

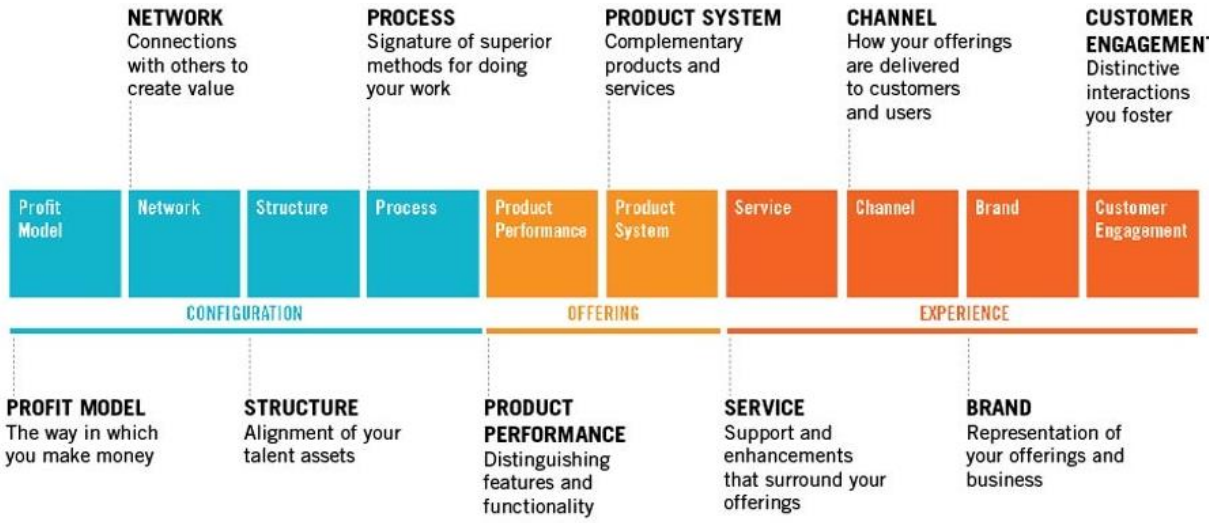


**Figure 5: Ten Types of Innovations by Keeley et al.[10]**

We find this theoretical approach to innovation very illuminating as many analyses of Big Data and Data Mining point to its influence on decision making or developing new products. However, decision-making is a ubiquitous activity in an organizational practice. Hence, breaking it down according to how companies actually operate (and more importantly – what those decisions are aimed at) helps in organizing one's thoughts around the subject. And so, "configuration" type of innovations "are focused on the innermost workings of an enterprise and its business system", "offering" refers directly to enterprise's core product or service or the whole collection of such. "Experience" type of innovations focus on "more customer-facing elements of an enterprise and its business system" (Keely *et al.* 2013). Below we present a more detailed explanation of each element in the given "innovation" category.

| Innovation category | Innovation element | Description |
|---|---|---|
| Configuration | Profit model | Innovative profit models find a fresh way to convert a firm's offerings and other sources of value into cash. Great ones reflect a deep understanding of what customers and users actually cherish and where new revenue or pricing opportunities might lie. Innovative profit models often challenge an industry's tired old assumptions about what to offer, what to charge, or how to collect revenues. This is a big part of their power: in most industries, the dominant profit model often goes unquestioned for decades. |

[10] Keeley L., Pikkel R., Quinn B., Walters H., *Ten Types of Innovation: The Discipline of Building Breakthroughs,* Hoboken, New Jersey: John Wiley & Sons, 2013; infographic provided by: https://doblin.com/ten-types/

| | | |
|---|---|---|
| | Network | In today's hyper-connected world, no company can or should do everything alone. Network innovations provide a way for firms to take advantage of other companies' processes, technologies, offerings, channels and brands—pretty much any and every component of a business. These innovations mean a firm can capitalize on its own strengths while harnessing the capabilities and assets of others. Network innovations also help executives to share risk in developing new offers and ventures. These collaborations can be brief or enduring and they can be formed between close allies or even staunch competitors. |
| | Structure | Structure innovations are focused on organizing company assets—hard, human, or intangible—in unique ways that create value. They can include everything from superior talent management systems to ingenious configurations of heavy capital equipment. An enterprise's fixed costs and corporate functions can also be improved through Structure innovations, including departments such as Human Resources, R&D and IT. Ideally, such innovations also help attract talent to the organization by creating supremely productive working environments or fostering a level of performance that competitors can't match. |
| | Process | Process innovations involve the activities and operations that produce an enterprise's primary offerings. Innovating here requires a dramatic change from "business as usual" that enables the company to use unique capabilities, function efficiently, adapt quickly and build market–leading margins. Process innovations often form the core competency of an enterprise and may include patented or proprietary approaches that yield advantage for years or even decades. Ideally, they are the "special sauce" you use that competitors simply can't replicate. |
| Offering | Product performance | Product Performance innovations address the value, features and quality of a company's offering. This type of innovation involves both entirely new products as well as updates and line extensions that add substantial value. Too often, people mistake Product Performance for the sum of innovation. It's certainly important, but it's always worth remembering that it is only one of the Ten Types of Innovation and it's often the easiest for competitors to copy. Think about any product or feature war you've witnessed—whether torque and toughness in trucks, toothbrushes that are easier to hold and use, even with baby strollers. Too quickly, it all devolves into an expensive mad dash to parity. Product Performance innovations that deliver long-term competitive advantage are the exception rather than the rule. |
| | Product system | Product System innovations are rooted in how individual products and services connect or bundle together to create a robust and scalable system. This is fostered through interoperability, modularity, integration and other ways of creating valuable connections between otherwise distinct and disparate offerings. |

| | | |
|---|---|---|
| | | Product System innovations help you build ecosystems that captivate and delight customers and defend against competitors. |
| Experience | Service | Service innovations ensure and enhance the utility, performance and apparent value of an offering. They make a product easier to try, use and enjoy; they reveal features and functionality customers might otherwise overlook; and they fix problems and smooth rough patches in the customer journey. Done well, they elevate even bland and average products into compelling experiences that customers come back for again and again. |
| | Chanel | Channel innovations encompass all the ways that you connect your company's offerings with your customers and users. While e-commerce has emerged as a dominant force in recent years, traditional channels such as physical stores are still important — particularly when it comes to creating immersive experiences. Skilled innovators in this type often find multiple but complementary ways to bring their products and services to customers. Their goal is to ensure that users can buy what they want, when and how they want it, with minimal friction and cost and maximum delight. |
| | Brand | Brand innovations help to ensure that customers and users recognize, remember and prefer your offerings to those of competitors or substitutes. Great ones distill a "promise" that attracts buyers and conveys a distinct identity. They are typically the result of carefully crafted strategies that are implemented across many touchpoints between your company and your customers, including communications, advertising, service interactions, channel environments and employee and business partner conduct. Brand innovations can transform commodities into prized products and confer meaning, intent and value to your offerings and your enterprise. |
| | Customer engagement | Customer Engagement innovations are all about understanding the deep-seated aspirations of customers and users and using those insights to develop meaningful connections between them and your company. Great Customer Engagement innovations provide broad avenues for exploration and help people find ways to make parts of their lives more memorable, fulfilling, delightful – even magical |

**Table 4: Innovation elements**[11]

Companies struggle to find improvements in each of the abovementioned fields of their business. By tapping into Big Data, they may improve their business performance. By improving actions and operations at the micro level, the overall economy will derive benefit.

It has been estimated that Big (and Open)[12] Data will give an incremental boost of 1.9% to European economic growth by 2020 (Buchholtz et al. 2014). This growth will transpire mainly through three types of economic gains to organizations. However, one has to remember that these gains follow improvements at the micro level:

1. Resource efficiency improvements through better use of information concerning resource waste in production, distribution and marketing activities.
2. Product and process improvements through innovation based on R&D activities, day-to-day process monitoring and consumer feedback.
3. Management improvements through evidence-based, data-driven decision making (Buchholtz et al. 2010: 11).

We develop an in-depth analysis of the economic model that foresees a 1.9% uplift impact on Europe's GDP in the last chapter of this report. In chapter three we also try to estimate how big the role that TDM plays is. At this stage, it is important to note that Big Data's economic potential is based on two main basic conditions, namely: (1) data quality and timely accessibility and (2) mining and analysis capacity. This relationship is inter-conditional – without good (quality) and accessible data no mining can be performed; and without TDM capabilities, little value can be extracted from the data. More importantly, the quality of TDM may often depend on the quality and accessibility of data itself. It is clear that without TDM Big Data is much less capable (if not utterly incapable) of delivering actionable intelligence.

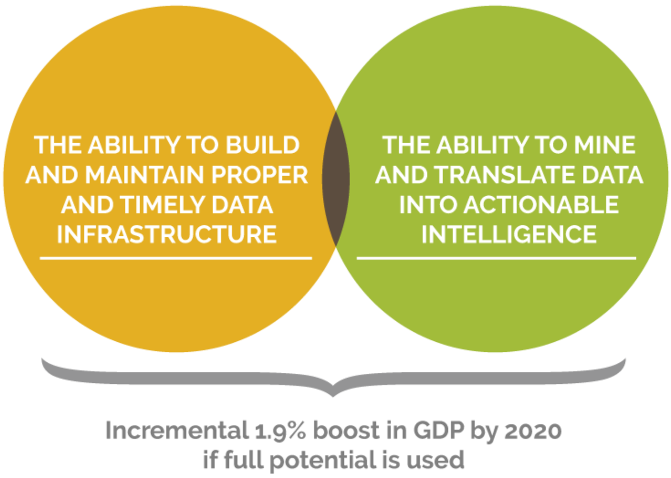


**Figure 6: Fundamental Aspects of Big Data**[13]

[11] *Keely et al. 2013*

[12] The authors of this study include the economic impact of Big Data (which they treat as proprietary company data) and Open Data (PSI meeting openness criteria). However, according to these calculations, Open Data will only account for for 4% of this GDP boost.

[13] Created by the authors using the estimates in Buchholtz et al (2014).

To recapitulate, the argument we put forward in this report (based on the premises in the first chapter) is that TDM accessibility and quality will, to a large extent, determine its uptake and therefore also the economic impact exerted by Big Data. While Big Data discussions have so far pointed more to aspects concerned with data infrastructure development and maintenance, TDM – as shown in chapter two – presents an opportunity to address the issue of human skills and capacities to extract value from data leveraging data infrastructure. We believe that the most essential elements of the Big Data promise can be broken down into two building blocks: (1) deploying storage infrastructure and (2) extracting value through TDM (see Figure above). Although the former topic has been addressed to a great extent, the latter one lacks the attention of business leaders and policymakers alike.

# 2 COMPANY PERSPECTIVE

## 2.1 Introduction

In this chapter, we will present market data that shows how companies approach Big Data and Analytics (BDA). For promoting the uptake of TDM requires us to take an organizational perspective. We will mainly rely on data provided by leaders of organizations across many sectors to depict their perception, understanding, challenges and trends in the quest toward translating ever larger data sets into business and economic value.

Before we go into detail, it is worthwhile to recapture some of the key findings that are important to keep in mind when purveying company specific data. Chapter one presented some lexical discussions and arguments concerning TDM's vital role in poising Big Data to make an economic impact. We concur with the assertion that Big Data as a "brand" is less than fortuitous when it comes to fathoming the subject. For the challenge organizations face is not to construe big data sets (or lakes) or develop TDM capabilities for their own sake, but, to put it bluntly, *translate data into value*. It is, thus, not about gaining access to big pools of data, but the ability to reuse them in a way that proves valuable to the operations at hand.

For these reasons, the creators of models of Big Data's economic impact quote three main positive effects it may exert on organizations: (1) resource efficiency improvements, (2) product and process improvements through innovation, (3) management improvements (Buchholtz et al. 2014: 11). To produce effects, data needs to be transformed into specific pieces of information to make the right decisions at a given time and venue. To explain the complexity of this challenge we also rely on the term "actionable intelligence" as explained in the previous chapter. This approach is consistent with what the top consulting firms contend when dealing with Big Data:

*"The time is ripe for big data and analytics initiatives to pull leaders ahead of the pack. New and varied sources of data, including weblogs, sensors and social media streams, are weaving their way into the analytics mix, **making it easier than ever to create connections that convert into actionable insights** [underlined by authors]."* (Forbes Inisghts 2015: 2)

"*The ultimate success of Big Data projects lies in realizing business value."* (Capgemini & Informatica 2016)

It is important to note, however, that the characteristic of actionable intelligence depend very much on the problem that is being debated at a precise moment – there is no one golden rule of what is an actionable insight and what is not. That is why the process of translating data into economic value needs to be strongly supported by the business side of a given organization, namely, the people who are the most knowledgeable about how an organization creates, delivers and sells value. Also, this is why the whole analytical value chain in a given company needs to be supported by information flows between many crucial departments and decision makers.

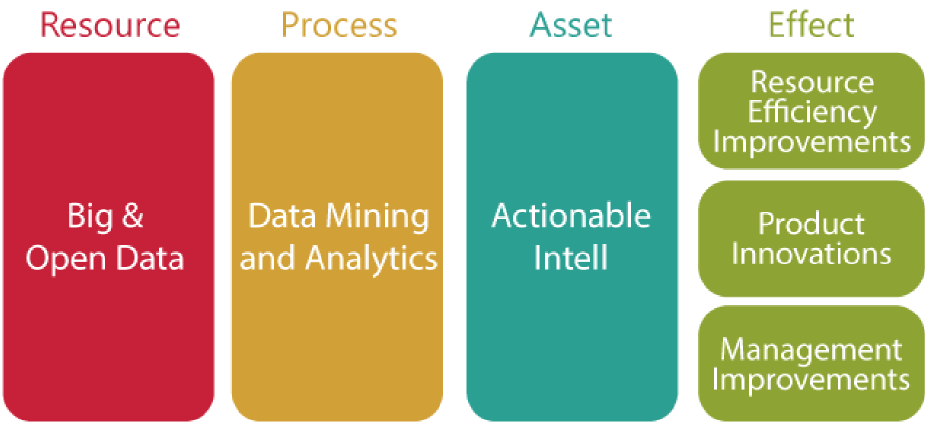


**Figure 7: Analytical Value Chain**

From this perspective, Big Data is perceived as a resource to be transformed into an asset through mining and analytical activities. It is important to remember that organizations do not strive to contrive Big Data strategies for their own sake; rather, they must see value in setting up the whole value chain and make it operational in a way to produce actionable intelligence.

Because of the nature of this challenge, the term used by International Data Corporation (IDC – a major ICT consulting firm) Big Data and Analytics (BDA) is much more general, inclusive and explanatory. Big Data and Text and Data Mining are crucial elements of BDA. It also reflects how thinking about Big Data has transformed recently as it matures and joins the mainstream. As industry and business practices develop, "Big" has lost significance in favor of accuracy encompassing adequate variety (proper sources) and velocity (speed measured in time). This development also stresses that much more analytical precision in data mining is required to ramp up the value of mining activities *vis-à-vis* Big Data sets.

*"Big data is no longer a buzzword. It's mainstream, says Gartner. The definition has changed where big is more influenced by data variety and velocity versus volume. When it comes to volume of data, big is relative. In 2016, companies will move away from irrelevant data noise, acknowledge that the variety and speed of data can be daunting and will take a more thoughtful approach to analyzing "useful" data to reach fast, meaningful, holistic insights. Rather than investing time and money in IT infrastructure to manage high volumes of data, the trick will be managing the data diversity and speed at which data streams to glean valuable insights and to do something worthwhile with them." (Gutierrez 2015)*

As Big Data hype matures, companies develop different ways of engaging with BDA. Buytendijk (2014: 9) proposes a framework of understanding what companies actually do in BDA claiming that:

- "Big data analytics for customer service, sales analytics, intent-driven customer systems and big data analytics for e-commerce have all become part of **Big Data Analytics for Customer Intelligence**.
- Speech recognition, text analytics and video search have become part of **Content Analytics**.
- Social media monitoring has become part of **Social Analytics**.
- Cloud-based grid computing and cloud parallel processing have become part of **Cloud Computing.**
- Context-enriched services have become part of **Context Brokers**.

- Supply chain analytics have been superseded by new and more specific **Supply Chain Big Data Analytics**."

This classification points to the most promising fields of BDA. It is, however, important to remember, though, that the list is rather a group of trending tags and is also: (1) far from complete, (2) constantly changing and (3) evolving together with innovative practices. To illustrate complexity one may easily notice that, for instance, *Content Analytics* is used in *Social Analytics,* but also can support the process of gathering *Customer Intelligence*. *Cloud Computing* shows in turn the importance of cloud based storage for BDA.

Thus, understanding BDA movement is not an easy task as many developments are very recent and remain in constant flux. However, the general understanding is that in all these practices, the challenge lies in the capacity of transforming data into actionable intelligence for solving a practical issue, and this remains intact in our firm belief.

## 2.2 BDA process structure and its main challenges

When attempting to understand, what organizational processes stand behind BDA, one may rely on information provided by market players. In 2015 Gartner (one of the most prominent technology focused consultancies) commissioned a survey in which 437 leaders across eleven sectors were asked to rank the challenges related to Big Data[14]. In other words, the respondents identified what stands in the way of extracting value from data in their organizations. Those challenges present – in our opinion – insight into how companies approach BDA and furnish a more nuanced understanding of where TDM plays a role.

| Name | Description |
|---|---|
| Determining how to get value from Big Data | Requires knowledge what data is being gathered by a company, also requires deep understanding of business and the company's business model. The challenge here is to link data to the current business model, either improving it, or changing it utterly. It requires a theory on how a company may gain by developing and acting upon a Big Data strategy. Only when the initial idea of value determination is developed can it be further translated into a specific strategy. |
| Obtaining skills and capabilities needed | Obtaining the right skills and talents goes across all the aspects of having a Big Data program in a given company. However, it mostly refers to three crucial capabilities: (1) IT engineering, (2) Data analytics, (3) Leadership. |
| Risk and governance issues | Refers to security, privacy and legal issues but also delves into data quality. For if decisions are to be influenced by data analytics, the data must be robust. |

[14] Gartner performed this research for the third time. Comparing results over time aids in understanding important BDA trends. We develop these comparisons further in the text.

| | |
|---|---|
| Funding for big-data related initiatives | Running a Big Data program within a company needs dedicated resources. It may be seen as an investment; however, it appears as a cost in the company's P&L. Thus, starting a Big Data undertaking will require using either internal or external financing – this is why in general big companies are seen as the drivers of the Big Data revolution as they have – on average – greater access to the necessary resources. On the other hand, in government agencies a political decision is usually needed. |
| Defining strategy | Developing a proper strategy to deploy a Big Data program, its goals and time-related framework. |
| Integrating multiple data sources | Decisions need to be made about which sources of data represent valuable information outlets. |
| Integrating Big Data technology with existing infrastructure | Companies have legacy IT structures. Very often they are fragmented across a company's various departments. Running a Big Data program means that a spate data infrastructure dedicated to Big Data has to be constructed. This infrastructure needs to use legacy IT solutions. Thus, integration is as much an engineering issue as an analytical one that involves data mining decisions. |
| Infrastructure and/or architecture | Developing proper Big Data IT architecture. Picking the technology to be used. |
| Leadership and organizational issues | As with every business process, sound leadership and organizational alignment is needed. Processes need champions to lead organizational change. Especially with Big Data projects new management styles or practices that may face organizational obstacles may be introduced. |
| Understanding what Big Data is | Some companies still have trouble understanding this concept with all its complexity. |

**Table 5: Challenges for Analytical Value Chain[15]**

From the above elements one can paint a picture of what has to happen in a given company to collect and mine the proper data and translate it into actionable intelligence to take a business decision for future evaluation. Such an organization must pursue a profound process of creative thinking leading the management to see the opportunities inherent in linking data to the current business model.[16] Only from there may a strategy be developed. As a consequence, a given organization will be looking for talents and/or external support to create a proper data infrastructure, integrate it with the legacy structure and manage execution with all the related risk and governance factors in mind.

55% of the respondents in Gartner's survey claim they struggle with determining how to get value from data – within this group, 33% rank it as their top challenge. It is the most single challenging factor that

[15] Created by authors based on Gartner (2015), Big Data Insights, p. 8

[16] We understand the term business model within the framework proposed by Osterwalder & Pigneur (2014: 14). They propose that: "A business model describes the rationale of how an organization creates, delivers and captures value". We endorse this definition as it puts stress on value, not on "how an organization makes money". We believe that this approach also makes it possible to include other types of organizations that do not make money per se, but create scientific, social, cultural or public value through their actions. For those organizations may also make use of BDA. For a brief discussion on business model definitions see: Ovans (2015), *What is a business model?*, Harvard Business Review, retrieved from the web on 3rd August 2016, https://hbr.org/2015/01/what-is-a-business-model

holds companies back from investing in more complex data solutions. This percentage rises to 62% when only companies with no plans for this investment are taken into consideration. These findings show that companies may have the technical skills to develop big data infrastructure (only 22% indicate this as being problematic); however, they lack visionary and creative competencies allowing them to see business opportunities in the uptake of big data and data mining.

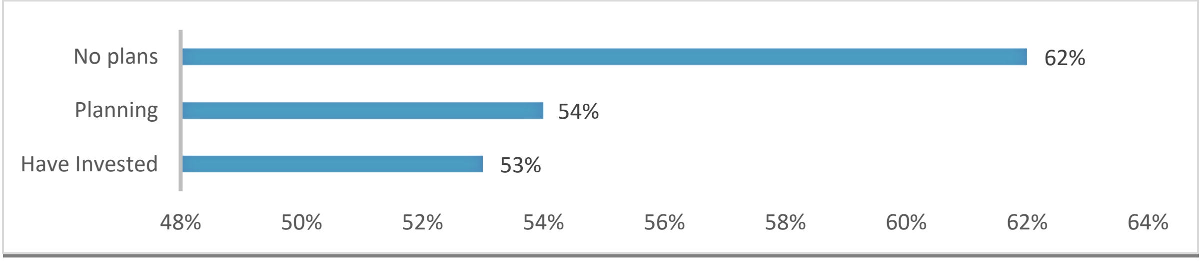


**Figure 8: Determining how to get value from BDA (2015)**[17]

What is optimistic, though, is that the number of leaders who are struggling to determine how to get value from Big Data and Analytics is constantly dropping – compared to 2014 it fell 10 percentage points. This trend is visible in Gartner's research. This may point to the conclusion that Big Data is rapidly becoming a mainstream process (in 2013 fewer than 8% of the 720 companies surveyed by Gartner stated that they were entertaining the adoption of Big Data and Analytics solutions). This means that more resources are being devoted every year to mining and analytical activities (which is also empirically proven by the growth of the Big Data Market as presented in detail in chapter 3). It is fair to assume that mining and analytics-related investments are followed by big data infrastructure development.

This assumption seems to be confirmed when seeing how the challenge of sourcing suitable talents has gained importance – from 30% in 2014 to 36% in 2015; in 2013 it was 33%. On the other hand, developing proper infrastructure and solving technical issues related to integration are becoming less of an issue. Hence, it seems to be true that as companies become more skillful in maintaining and managing large data sets they also start to see more need for the appropriate analytical competencies.

[17] Gartner (2015), Big Data Insights, p. 8

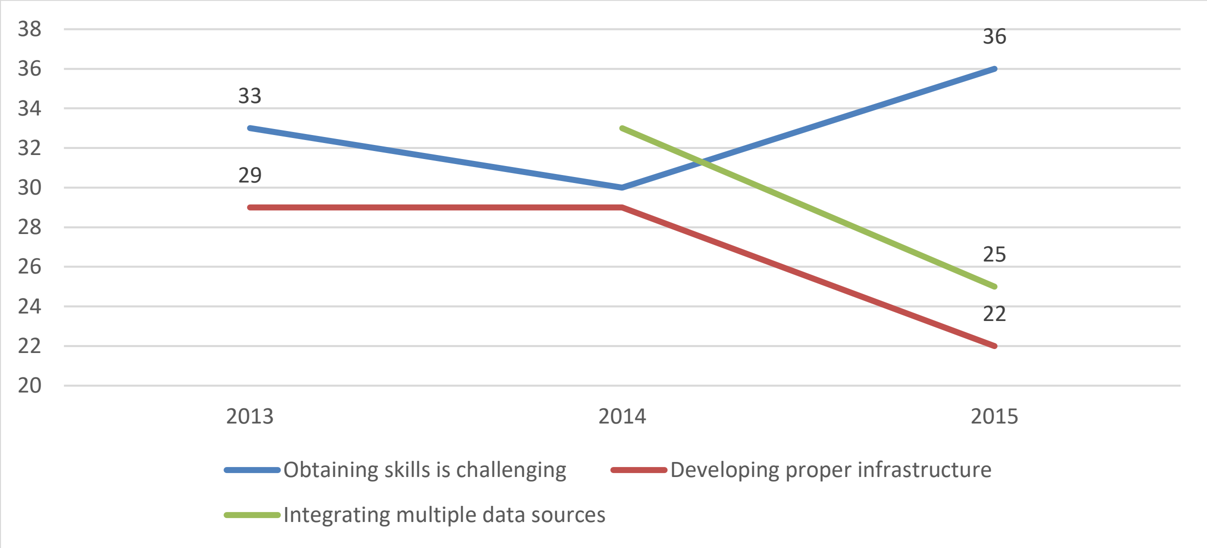


**Figure 9: Obtaining skills becomes more challenging vis-à-vis technical issues**[18]

Obtaining proper skills is followed by managing risks, funding and strategy definition issues. So, it is by no means a given that even the organizations that introduce BDA are able to capitalize on them easily. It seems companies do not struggle that much with introducing proper IT infrastructure; however, analytical skills (including mining) to translate data into business value is much more problematic.

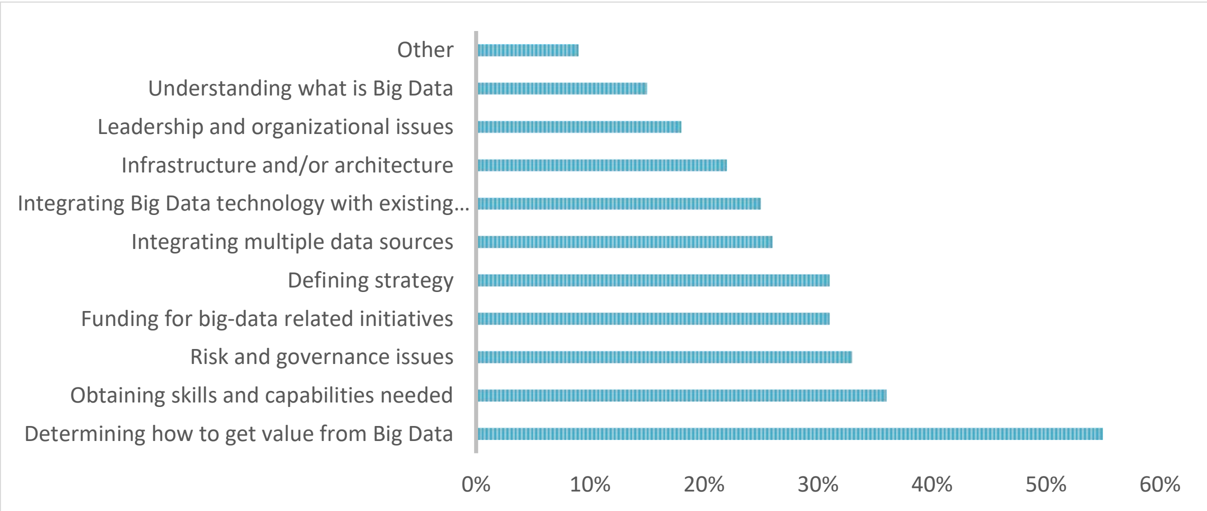


**Figure 10: Challenges of introducing Big Data and Analytics (2015)**[19]

One could, thus, conclude that much faith is placed in including more data analytics into business processes; however, there is much less satisfaction when it comes to getting actual outcomes. For instance, Mu Sigma (a data analytics firm that is valued at over $1 billion according to Forbes) presented its "State of Analytics and Decision Science" Report that "shows that many businesses are still misguidedly prioritizing data and technology needs over the need for better decision making"

[18] Gartner (2013, 2014, 2015), Big Data Insights
[19] Gartner (2015), Big Data Insights, p. 8

(2016). So it seems, that, although the lack of analytical talent is felt, the main driving force behind Big Data initiatives has so far not been strongly related to specific business problems.

One illustration of this is that 74% of the respondents indicated their confidence that "they lead with data when it comes to problem solving", while only 26% are satisfied with the outcomes[20]. This finding is seconded in a research conducted by Ransbotham, Kiron and Prentice (2015: 3)[21] showing that as access to useful data has increased over the last years, it has become much more difficult to apply it. **This means that there is a substantial gap between exceling in maintaining data and translating it into valuable insights and decision-making**.

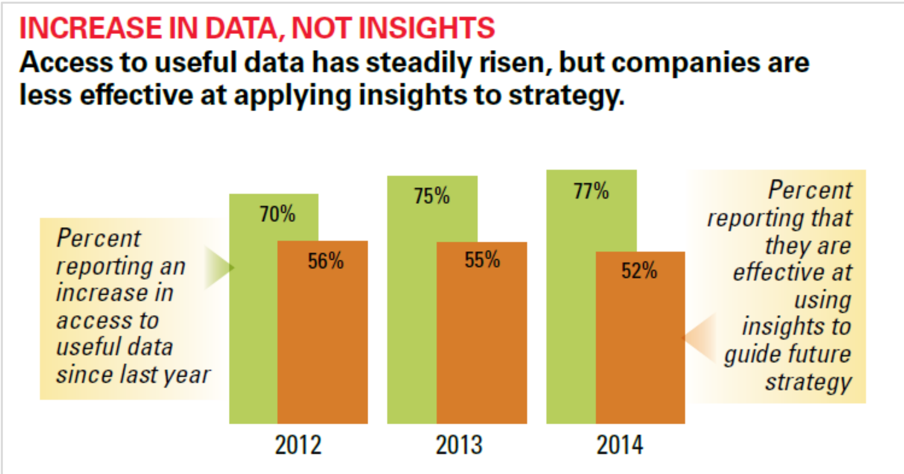


**Figure 11: Discrepancy between access to and application of data[22]**

It can also serve as explanation of why 41% of the leaders surveyed by Gartner (2015: 24) say "they don't know if Big Data's Return on Investment will be positive".

Data shows that as Big Data movement matures, companies become better at building data infrastructures – only 22% of the respondents in 2015 indicated this as a challenging issue (compared to 29% in 2014) – however mining, translational and other analytical skills will be needed more.

[20] Although this particular research project was done only with the participation of US business leaders.
[21] Based on a survey of 2,719 respondents from across the globe and interviews with 28 executives and thought leaders.
[22] Ransbotham, Kiron and Prentice, 2015: 3.

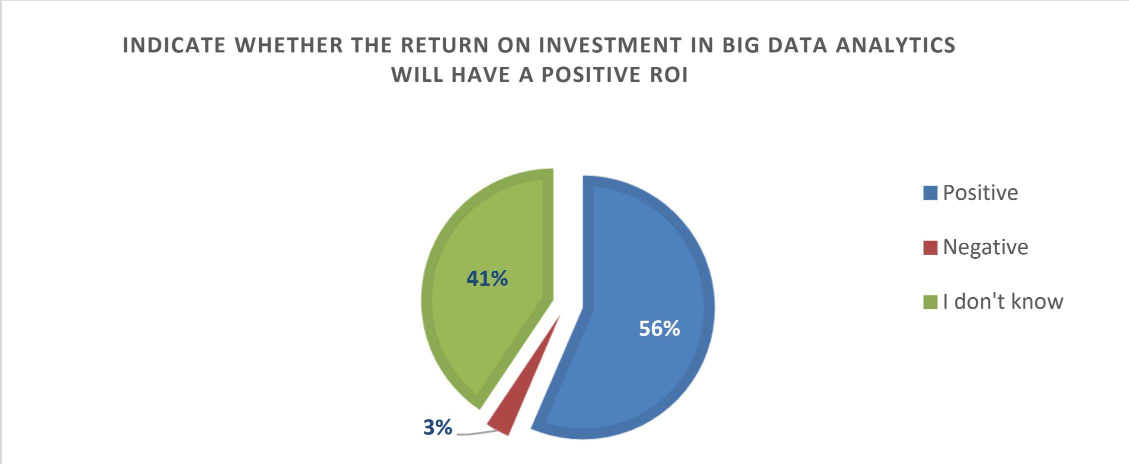


**Figure 12: Faith in Big Data Analytics[23]**

Gartner's (2015: 9) research also shows that organizational challenges become more practical as they start executing their Big Data strategies. For instance, only 21% of the organizations with no plans of investing in Big Data cite "obtaining skills" as a challenge. In turn, organizations that have concrete plans, or have invested in Big Data recently, report numbers of 43% and 39%, respectively. This trend further proves that mining and analytical skills will become more important as BDA as a field of practice matures. Although as Wikibon (Kelley 2015: 2) foretells, at subsequent stages of BDA development, mining and analytics will also become standardized and embodied in a specific software to facilitate greater automatization.

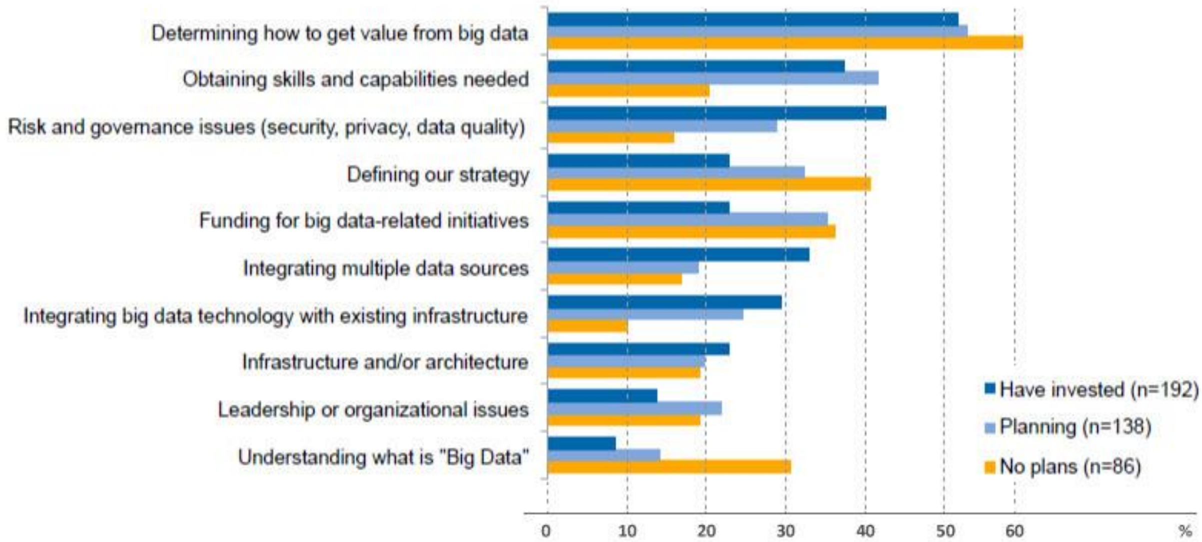


**Figure 13: Challenges in (B)D&A[24]**

Nevertheless, this data shows that organizations are aware that having the correct Big Data infrastructure is only a prerequisite and this awareness is also reflected by their leadership approach. Mu Sigma stated that "23% [of the surveyed companies] noted that CIOs [Central Information Officers] are in charge of data and analytics while 17% named the CFO [Central Financial Officer] and 13%

[23] Gartner (2015), Big Data Insights.
[24] Gartner, 2015.

referred to the relatively new position of CAO", i.e. the Central Analytics Officer. Hence, it seems that as the discipline matures CIOs will be more responsible for maintaining the information infrastructure while CAOs will be the direct link to business goals and problems.

It seems that understanding that much investment is needed in talent acquisition and development is on the rise. It would be fair to claim that this will be the investment focus in the near future. However, one could say, that this awareness is just budding and does not signify maturity. The empirical evidence on where companies actually invest their resources reveals that mining and analytical talent is still overshadowed by investment in infrastructure:

*"Not quite as encouraging, however, is where these investments lie. The largest percentage—37%—of respondents are making very significant investments in storage, while 35% are focused on data acquisition. The least cited investment is in analytics (30%), while only 34% of respondents allocate funds for talent. This prioritization of investments shows how big data and analytics are still very much in their infancy.*

*In fact, heavy investments in data storage indicate that many organizations are still in the preliminary stages of managing, storing and sorting huge volumes of data."* (Forbes Insight 2015: 14)

On the other hand, when one wants to reach the stage at which business value is constantly driven by data[25], talent and culture need to become as much of a priority as storage, hardware and software.

*"Organizations ahead of the maturity curve, on the other hand, are focused on deriving value from their analytics efforts by investing in both organizational change and in-house talent. Ask Lochner where RBI invests most heavily and he'll say, "Definitely not in storage, definitely not in hardware or software." Rather, Lochner says funding must support "how you are changing the company" and teaching employees "to move into this new digital world. Change management is really the top topic of investment here." For this reason, RBI relies on everything from large-scale training sessions to company newsletters to evangelize its big data initiatives across the organization."*

[25] We present a more detailed account of the stages of maturity when it comes to (Big) Data and Analytics deployment below. Here, we just want to underline that the challenges to successful implementation point directly to mining and analytical talents as well as organizational culture. These soft issues are very often overlooked by companies that believe that introducing new technology will work like a "magic wand" and unlock hidden potential on its own.

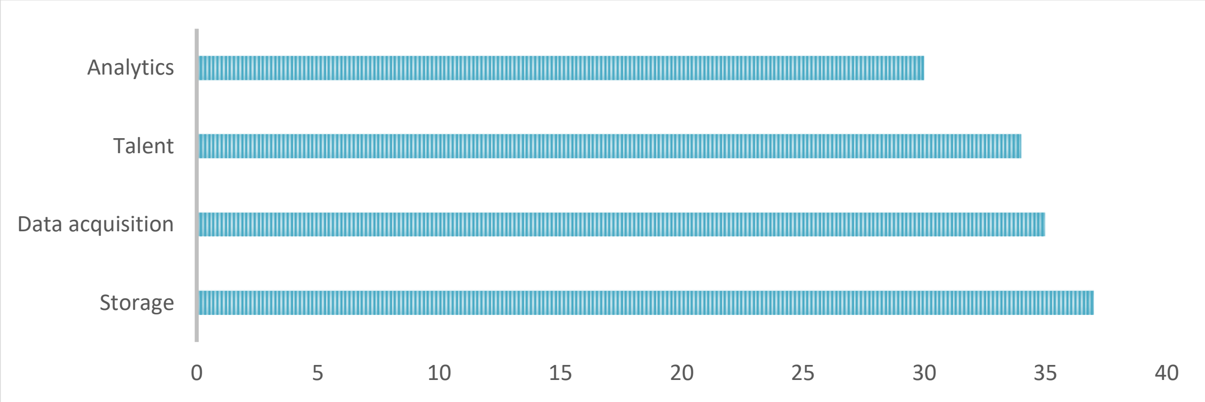


**Figure 14: Levels of investment in different aspects of BDA**[26]

Forbes' approach clearly shows that the challenges to deploy and further develop BDA solutions (necessarily including TDM) also can be looked at from an organizational culture perspective.

*"Part of the problem is that data-driven business models represent a break from the past that can call for a huge cultural upheaval. Almost overnight, employees must make data a priority, be willing to share data sets across departments and assume shared responsibility for data collection, quality and analysis. It's a tall order that can cause workers to recoil."* (Forbes Insight, 2015: 11-12)

Taking this perspective shows that 43% of the companies in the *Forbes Insight* (*Ibidem*) claim difficulties in fostering a culture "that rewards the use of data", or "values creativity and experimentation with data". Also, 41% leaders state that their companies encounter problems when they aspire for their staff to "view data as a valuable asset".

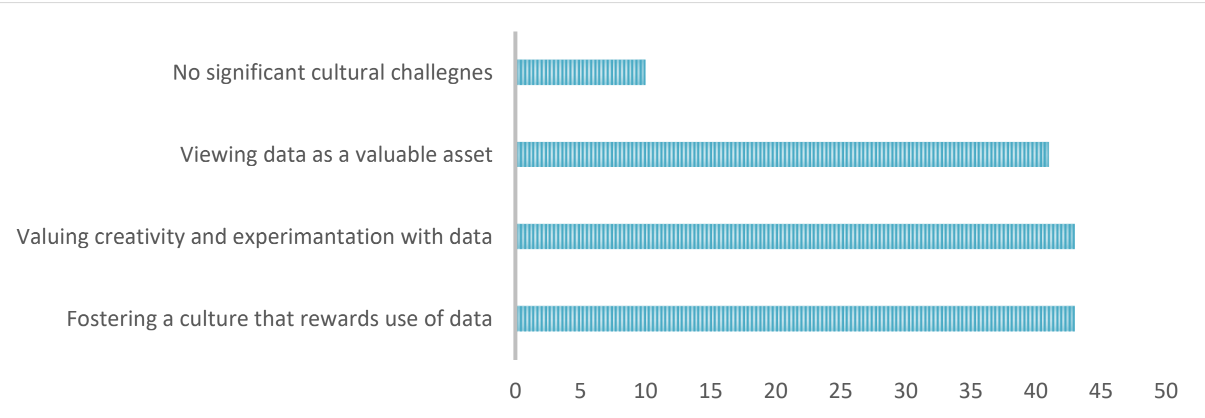


**Figure 15: Cultural obstacles to develop of BDA**[27]

The challenges presented in this section may be compared in two different regions using the study done jointly by Capgemini and Informatica (2016). Their research was carried out in 2016 with the participation of 100 US and 110 European business leaders knowledgeable about the Big Data initiatives in their companies. It shows that EU businesses feel much less constrained by integration

[26] Forbes Insight, 2015: 15.
[27] Forbes Insight, 2015: 11-12.

challenges; however, data quality seems to be much more of an issue in Europe. Both sides of the Atlantic report similarly strong concerns about the lack of proper skills and corporate culture.

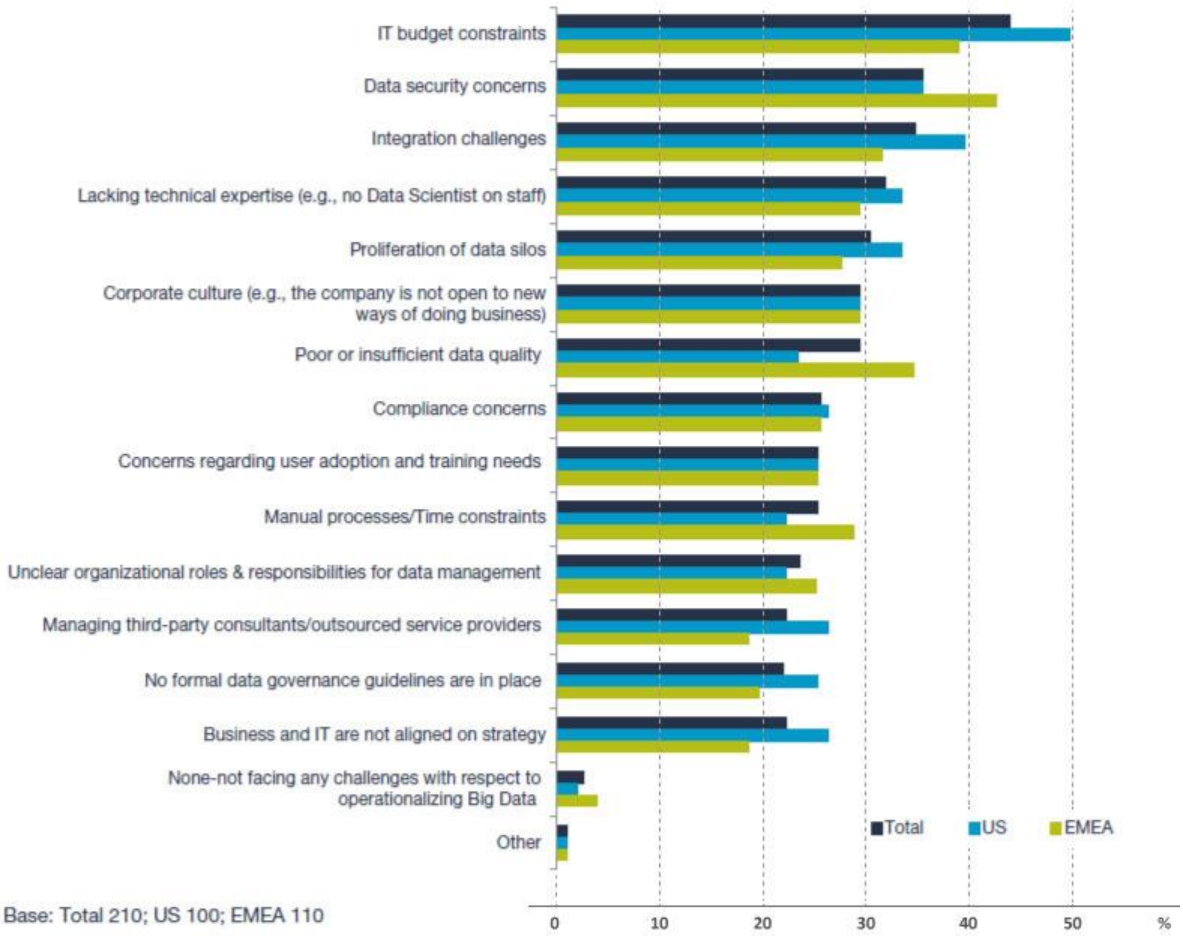


**Figure 16: Top BDA challenges across US and EMEA (Europe, Middle East and Africa)[28]**

## 2.3 Adoption rates, patterns and goals

When studying Gartner's data one also sees that the adoption of Big Data and Analytics has vastly improved in recent years. Since 2012 – the first year of Gartner's research – the percentage of companies reporting investment as well as firmly planning to invest has climbed 18 percentage points to 76% in 2015. Each year fewer organizations have "no plans" and are undecided when it comes to BDA investment. This clearly means that integrating vast amounts of data with business decisions is becoming more mainstream each year. Gartner's analysts also claim that "Based on the tone of inquiries from Gartner clients, we are seeing the second wave of big data adoption, but it is slower and more deliberate than the first. This second wave is less enamored by technology, instead being driven by business value" (Heudecker and Kart 2015: 3).

[28] Capgemini & Informatica, 2016: 3.

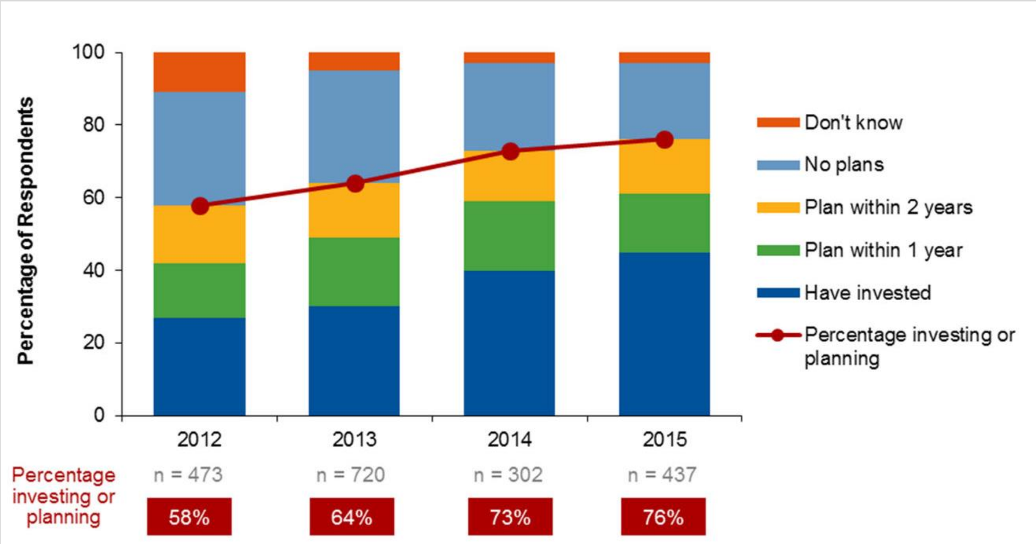


**Figure 17: Increased BDA uptake over time**[29]

Gartner also takes a different perspective on BDA adoption by companies. It breaks down the process of investment into stages: (1) knowledge gathering, (2) developing strategy, (3) piloting and experimenting and (4) proper deployment. This breakdown is very insightful as it shows that there is much more to BDA than merely IT technology. We will return to this observation in the chapter on BDA market valuation as the data on economic transactions reveals only a fraction of the real financial value created by BDA.

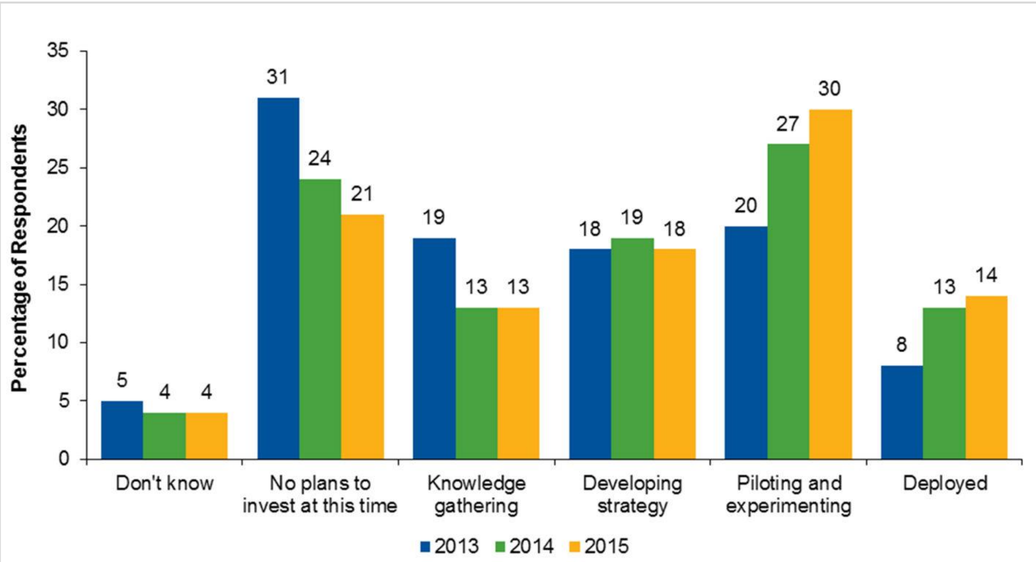


**Figure 18: Stages of BDA Investment**[30]

The data presented above shows that each year more companies are experimenting with BDA solutions. This also translates into greater deployment; however, the conversion rate (the percentage of companies experimenting with big data that go proceed to the deployment stage) is only 60%. In other words, the 10% increase in piloting and experimenting that took place between 2013 and 2015

[29] Gartner, 2015.
[30] Gartner, 2015.

only corresponds to a 6% uptick in deployment. This means that there is substantial room for improvement. This rate is suboptimal because many companies are acting on the Big Data hype and starting to develop proper IT infrastructures. However, they lack the ability to connect these infrastructures to business problems and – hence – business outcomes. We believe that the missing link here is the proper application of TDM and other analytical tools to facilitate the translation of data into value.

To deepen the understanding of how burdensome actual value extraction from data is (and why TDM is a crucial factor), IDC's maturity classification is poignant (Buytendijk, 2014). Its analysts have observed that in the early stages of experimentation Big Data and Analytics have an ***ad hoc*** characteristic – the management usually greenlights pilot projects that are most probably data driven and not connected to specific business challenges. At this stage, it is still difficult for a company to derive a return on its investment in a BDA solution and, thus, the value usually comes down to gaining new knowledge and learning. IDC experts claim that it is not until the second stage of maturity referred to as **Opportunistic** that a company starts seeing specific business opportunities. If BDA programs gain more stable funding, the business opportunity is realized and individual BDA champions gain buy-in from other stakeholders in a company. Even so, business value is still restricted to specific business units at this stage. If an organization is able to build a strong business case, management will make the decision to invest more resources. In effect, Big Data and Analytics projects will become more **Repeatable**. In addition, these projects will get more management attention and that usually means that proper strategies and evaluation standards are implemented. If BDA still proves useful at this stage, projects will be **Managed** properly and in a more sustainable way. On the other hand, the outcomes will be translated into business plans to a greater extent. From there a company may want to start to **Optimize** – this becomes possible because the link between BDA projects and business outcomes is established. Through optimization a company is able to produce continuously previously unattainable business value through Big Data and Analytics.

It is our understanding that organizations usually start BDA with projects focused on data. In most cases, initial investment is not connected to any specific business problem. Therefore, data mining and analytics do not start to play important roles until subsequent stages of maturity. However, and this is consistent with this paper's findings so far, data mining and analytics are an essential element to derive business value from investments in Big Data technical infrastructure. It is important to underline that from the perspective of maximizing economic impact, mining and analytical activities seem to play a crucial role. For many organizations become disillusioned at the experimental stage and stop BDA projects at the beginning of the maturity curve. Introducing more mining and analytics has the power to serve as an important element affording them the capacity to see the business value offered by Big Data.

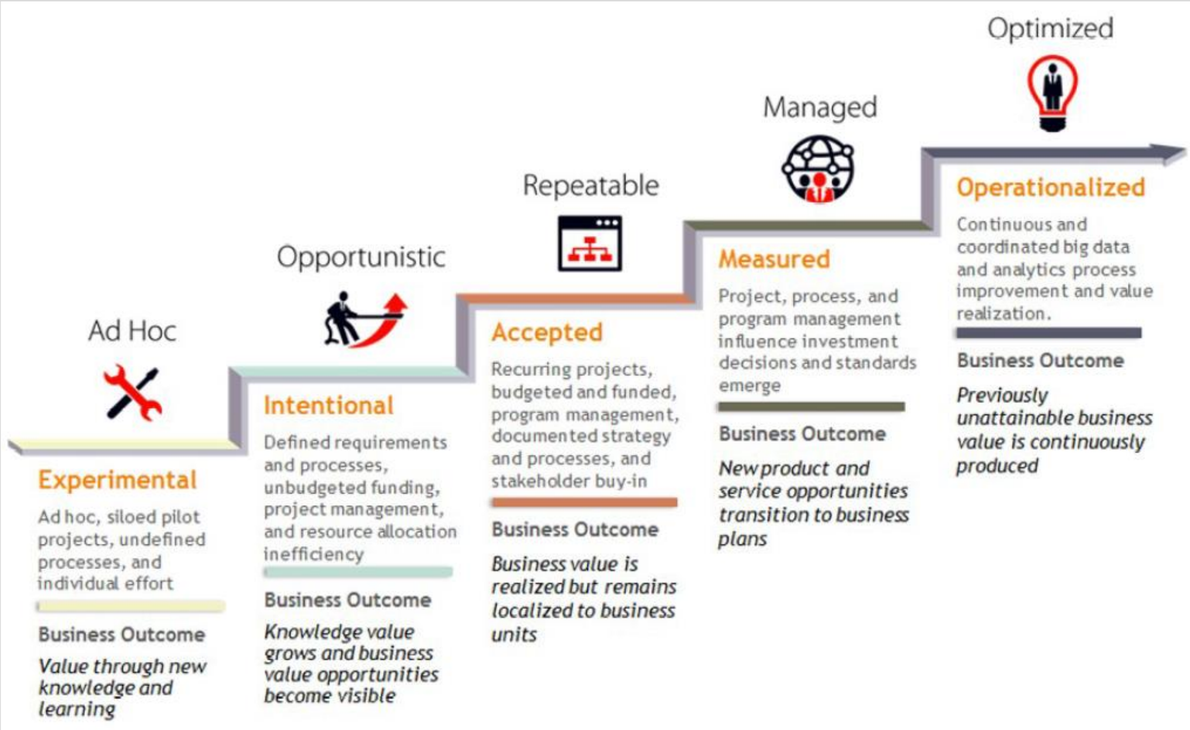


**Figure 19: BDA Maturity stages[31]**

What is interesting is that IDC uses these 5 stages of maturity to assess the 5 organizational dimensions in which an organization needs to excel if it wants to thrive with Big Data and Analytics: Vision, Data, Technology, People and Process. The "People" dimension involves "Recruiting and developing skills for data and content collection, integration, preparation and management in addition to skills for performing multidimensional analysis, predictive analytics and data discovery"; essential skills for performing TDM. At the same time, one may recall that acquiring these skills is one of the top challenges companies currently identify, especially as they mature along the implementation curve.

[31] IDC's Big Data and Analytics MaturityScape Benchmark Survey, 2015.

| Maturity Dimension | Key Traits Differentiating High Achievers |
|---|---|
| Intent | Fund and budget BDA program at the enterprise level. |
| Data | Ensure timely, trusted, and clean data. |
| Data | Integrate customer-interaction data with other data types. |
| Technology | Balance development of structured reporting tools and advanced analytics (predictive analysis and data mining). |
| Technology | Enable mobile decision making and data consumption. |
| Process | Address strategy and planning, governance, application development, and data analysis and information consumption. |
| People | Recruit and develop skills for data and content collection, integration, preparation, and management in addition to skills for performing multidimensional analysis, predictive analytics, and data discovery. |
| People | Build culture of collaboration, communication, and coordination across analytics groups and lines of business. |
| People | Recruit and development skills for administration and maintenance of big data and analytics–related IT hardware. |
| People | Engage executive management in BDA efforts and extend BDA insights to customer-facing associates. |

**Figure 20: BDA Maturity Stages in Detail[32]**

It should not come as a surprise that when IDC performed research on 150 US organizations that practice Big Data and Analytics it showed that "thrivers gravitate toward higher maturity levels. Almost half of the thrivers (47.4%) fell into the managed and optimized stages of BDA maturity, while the organizations classified as survivors had 39.4% in the managed or optimized stages. Conversely, about one-third (34.2%) of the survivors were classified as *ad hoc* or opportunistic, while just one-fourth (25.5%) of the thrivers were in the ad hoc and opportunistic stages. When comparing the thrivers and survivors for the five dimensions of BDA maturity, we see that the thrivers have a higher representation in the managed and optimized stages, with a lower representation in the ad hoc and opportunistic stages. These organizations had a quantifiable advantage when they actively pursued BDA solutions with a long-term vision and a well-defined process." (Vesset and Xiong 2015: 7).

IDC's findings are optimistic. Among the BDA practitioners the company surveyed, only 3.8% of them were at the experimental level. Moreover:

- More than half (53.6%) of the respondents indicated that their organizations were at the opportunistic and repeatable stage. These organizations have already had some defined requirements and processes in BDA projects. Data collection, monitoring and integration processes are in place at the department level, but consistent data governance and security practices have not been established. Business value realized remains localized because of the lack of coordinated use of BDA capabilities. There needs to be enterprise support for measurement tools and methods to drive outcomes and effectiveness, which can be achieved when organizations move to higher maturity levels.
- A small percentage of respondents (19.1%) are at the managed stage, where their organizations had experienced the emergence of BDA program standards. These organizations

[32] IDC Retail, 2014.

have developed a cross-business unit-level BDA strategy and there is an enterprise-wide budget with upper management support. IDC believes that these organizations are starting to enjoy the competitive advantage brought by BDA.

- About one-fourth of the survey respondents (23.4%) are at the optimized stage. Organizations at this stage have already established an enterprise-wide, documented and accepted BDA strategy. We suspect that the percentage of US -based organizations at this level of BDA maturity is this high because of the way in which Big Data and Analytics were defined in this research. In other words, a comprehensive definition encompassing decision support and decision automation processes, technology, people and data was used (rather than a narrow definition focused only on the extreme cases of Big Data). BDA processes are classified by performance management, operational intelligence and exploration and discovery, with appropriate support, staffing, technology and funding for each. These organizations leverage BDA solutions and their predictive capabilities to capitalize on new opportunities and to mitigate risk.
- Very few (3.8%) organizations were still at the least mature, ad hoc, stage. These organizations lack basic BDA strategy and management interest or support. BDA processes in these organizations are focused primarily on creating information repositories with access only to siloed information therein, which results in narrow knowledge and limited effect on business outcomes." (*Ibidem: 4*)

The above findings suggest that most (96.2%) of the Big Data and Analytics practitioners manage to implement Text and/or Data Mining techniques on top of the newly constructed Big Data infrastructure. Only a small percentage stays at a level dedicated only to building and maintaining big data sets. However, 53.6% of them still need a talent boost to link the department level of BDA to the overall company's business model. This requires a particular set of skills in coordinating mining and analytical ventures run by different business units:

"Successful deployment and the use of BDA solutions depends on a multipronged approach guided by a strategy that accounts for not just technology but also talent and capital resources, business and IT processes and the data." (*Ibidem*)

Similar studies were not done for the European market. However, since the European market is much less mature (as demonstrated below), more organizations are still in the stage of experimentation. Also, one has to remember that the Gartner research shows that only c.a. 14%[33] of the surveyed companies have deployed (B)D&A solutions.

## 2.4 Talents – Vital Component for TDM

As BDA matures, it is becoming clearer that although it represents a technological revolution, talents are needed to materialize its promise. The most up to date studies suggest that the challenges associated with Big Data in its very early days (like access to data, or IT infrastructure) are not as vivid

[33] However, this rate has to be interpreted with an understanding that Gartner's research encompasses companies that are much more interested in IT solutions. And, thus, one can conclude that in reality many fewer companies are making real use of (B)D&A. For instance, the European Parliament stated in 2016 that only 1.7% of EU companies "make full use of advanced digital technologies despite the benefits that digital tools can bring in all economic sectors" (European Parliament resolution of 10 March 2016 on 'Towards a thriving data-driven economy' (2015/2612(RSP))

for companies that approach the topic as they strive to attract data scientists responsible for mining and analyzing the data. One, recent study by MIT discovers that:

"*Companies today have more access to useful and cheaper data than ever before, but their ability to generate business value from data is showing signs of strain. More is not always better. Even with increases in useful data over the past two years, our survey suggests that companies might be getting less effective at using analytical insights to guide strategy, despite continued increases in investments in analytics technology. This suggests that managers are struggling to handle increasing flows of data. In addition, the percent of organizations reporting competitive advantage from analytics is on the decline.*

[…] *deriving business value from analytics depends in important ways on building strong internal capabilities that link insights with business outcomes. Half of the survey respondents (50%) cite turning analytical insights into business actions as one of their top analytics challenges. Difficulty managing the vast amounts of ever-increasing data from multiple sources is also an issue. And four in ten (43%) companies report their lack of appropriate analytical skills as a key challenge.*" (Ransbotham, Kiron and Prentice (2015: 3)

Talent gap is also invoked as one of the six major trends in Deloitte's "Analytics Trends 2016: The Next Evolution" (2016). Not surprisingly, the authors cite the IDC study that foresees the US economy facing a shortage of 180 thousand people with profound analytical skills by 2018 (*Ibidem: 7*). No similar study has been done for the European economy. This study also claims that companies can approach the lack of talent either by developing their own training programs or by outsourcing:

"[…] *analytics talent doesn't have to be directly employed by the organization. Some companies are consciously developing ecosystems of external providers. One, for example, has selected multiple services partners in the areas of business intelligence, predictive analytics, data science and cognitive technology. The company continually monitors the efforts of these partners to recruit and develop qualified people and to keep up with new technologies and methods.*

*These are by no means extreme steps. Smart companies are realizing that analytical talent is critical to their success and in short supply. They know they must get serious about preparing or partnering with this strategic workforce if they hope to execute their strategies successfully.*" (*Ibidem*)

Below, we present a brief summary of MIT's research on the use of data analytics in business underlining the necessity of analytical talent acquisition and development.

**Box 7. Analytical Talents Needed – MIT's Research[34]**

MIT Sloan Management Review has been preparing annual research on the use of data analytics in business since 2010. In 2014 it entailed 28 interviews with executives and thought leaders and a survey distributed to 2,719 respondents from a variety of countries, industries and organizations of all sizes. In this research project "analytics" covers the range of applied analytical disciplines like statistical, contextual, quantitative, predictive, cognitive and other models that provide insights into driving fact-based management. It presents the necessary capabilities for Text and Data Mining. Here are the main findings of MIT's review.

**MIT's years of research into data analytics makes it clear that business leaders have noticed that data access is getting easier each year.** From 2010 to 2014 the availability of useful and cheap data to companies surged upward. Just in the period from 2012 to 2014 the percentage of respondents reporting greater access to useful data from the previous year trended up from 70% to 77%. **However, in this same period managers began to struggle to keep a firm grip on higher flows of data. More and more companies are aware of the potential offered by analyzing their data.**

Managers are especially aware that client data – if properly used – may enhance marketing effectiveness. In turn, although the percentage of respondents who contend that business analytics create a competitive advantage in their organization climbed rapidly from 37% in 2010 to 67% in 2012, this trend reversed and dipped to 61% in 2014. This may mean that some companies are suffering a slight disillusionment with the data analytics hype.

Although data analytics is becoming more widespread, it often fails to provide companies with business value, as the conversion of analytical insights into business actions is still very limited. **Half of the respondents admitted that their major data analysis challenge was to turn data-based insights into business actions. What is important though is that the major barrier to capitalizing on data is no longer technological in nature. Rather, it is the shortage of companies' internal capabilities to aggregate multiple data sources into a communicable form, which requires appropriate mining and/or analytic skills.** This problem even affects business giants such as Coca-Cola and General Mills. They have since recognized their failure to exploit the full potential represented by their own data.

MIT's research shows new roles have emerged in organizations recently: **instead of just a simple "analyst" position, *chief data officers* and *data scientists* have appeared, as have supporting positions of *data stewards* and *data visualizers*. However, in reality, they are often left vacant or filled by employees whose skills are not aligned to contemporary challenges.** Four in ten survey respondents said their companies had difficulties attracting new employees with sufficient analytical skills. In addition, 40% admitted struggling to retain staff members on board. Currently, data analysts are often recruited from groups of operations research specialists and graduates in statistics and mathematics. Substantial investments are also being undertaken to run in-house training courses: in 2014, 63% of the companies invested in upgrading the analytical skills of their employees through formal on-the-job training.

The first bottle-neck in the process of "consuming" analytics is properly understanding and interpreting data – a key element of suitable TDM. Communication issues commonly referred to as the "last mile" problem in analytics are a major drawback. Specialized data analysts are frequently unfamiliar with the vernacular of executives and other specialists and it takes time and effort to convince regular employees of the importance of cooperating with data analysts. Only 27% of the respondents declared that they have successfully integrated new analytics talents with other important business departments. **To solve this problem, 49% of the respondents reported that their organizations were training managers to become more analytical and 34% trained their analytics professionals to understand the business and organization better.**

The second area of failure in integrating data analytics into the business decision-making process is the general approach companies take to the possibilities offered by data analytics. Most companies still consider data analytics to serve only operational goals: from 2012 to 2014 the percentage of respondents using data based insights to influence strategic decision-making dropped from 56% to 52%.
MIT's research segmented companies into three categories of analytic maturity with the key underlying factor being the company's access to skilled analysts.

**Analytically challenged** – (34% of the respondents) this group is characterized by relying more on management experience than data analysis. The use of data was of a descriptive nature and focused on cost-cutting. This group reported having the greatest problems with acquiring and integrating new analytical talents. They often preferred outsourcing analytic activities to specialized analytics vendors and only 13% of them gave hiring and promotional preference to persons with analytical capabilities.

**Analytical practitioners** – (54% of the respondents) companies in this group place a high priority on becoming a more data-driven organization. This group applied analytics in operating activity, but also made efforts to use data for predictive purposes.

**Analytical Innovators** – (12% of the respondents) these entities are much more strategic in their application of analytics findings. In this group, senior managers were much more involved in data-driven decision making. Data was used for predictive analytics (preparing various forward-looking scenarios). 75% of the group declared satisfaction with the current analytic talent bench and 65% would prefer analytically-talented employees in their hiring and promotion decisions.

**The main conclusions drawn by the 2014 MIT Sloan Management Review point to human resources being the key to derive benefits from flows of corporate data, reflecting the need to build the company's analytical talent bench.** Data access less problematic than the ability to translate it into business decisions. MIT recommends that companies:

1. Give hiring and promoting preference to people with analytical skills
2. Develop analytical skills through formal training
3. Oversee the integration of new analytical talents with other employees

### 1.1.1. Adoption goals and type of data used

The main macroeconomic effects of introducing Big Data and Analytics are: (1) resource efficiency improvements, (2) product and process improvements through innovation, (3) management improvements (Buchholtz et al. 2014: 11). These effects may be attained only if companies are able to implement effective (B)D&A solutions at the microeconomic level. The microeconomic perspective varies by company size and industry.

Gartner's research performed over the years shows that companies look mostly for insights: (1) to improve the customer experience derived from pre-existing offerings, (2) optimize operations and (3) undertake marketing initiatives by improving targeting. This shows that companies aim to integrate data sources related to customer interactions and as production-related data. These trends are firmly

[34] S. Ransbotham, D. Kiron and P.K. Prentice (2015), "The Talent Dividend: Analytics talent is driving competitive advantage at data-oriented companies," *MIT Sloan Management Review*, April 2015.

entrenched and will probably continue to develop. However, as companies try to understand their customers better, they might want to look for third party socio-economic data sources.

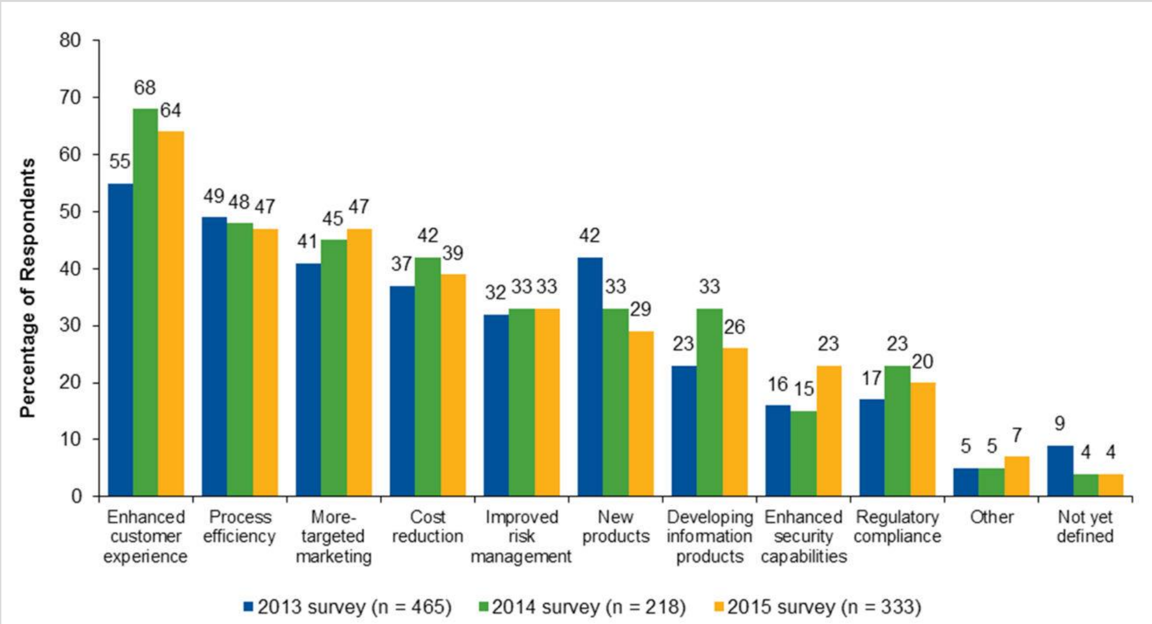


**Figure 21: Trends in using (B)D&A[35]**

One trend that seems to be worrisome in the development of TDM is that organizations seem to have troubles using BDA for developing new products. They have reported a drop in such applications from 42% to 29% since 2013. This finding should be highlighted because one of Big Data's big promises was to enhance product development. It seems though that BDA leads to incremental, not disruptive innovation. Customer insights generally point to measures to improve pre-existing offerings. Companies' top three priorities seem to be: (1) perfecting their pre-existing offerings, (2) searching for efficiencies and cutting costs and (3) improving targeting in marketing.

These three motivations seem to be priorities for most of industries. However, banking, insurance, health care and utilities are much more interested in risk management (see Figure 22). On average, 56,5% of the companies in these lines of business contend that "risk management" is a top priority for investing in BDA compared to an average of 26% for other industries. It also seems that Utilities, Manufacturing and Government do not care as much about enhancing customer experience by tapping into BDA insights. To be sure, government agencies may be affected by their generally low adoption rates[36].

[35] Gartner, September 2015.

[36] In general governments are criticized for their much slower adoption of BDA solutions than the private sector. See, for instance: Parish & Belissent (2015), Yiu (2012), or Bean (2016)

| | Manu & N. Res. | Media/ Comm | Svcs | Gov. | Edu | Retail | Banking | Insur-ance | Health-care | Trans-portation | Utilities |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Enhanced customer experience | 52% | 78% | 66% | 43% | 76% | 83% | 77% | 77% | 73% | 69% | 44% |
| Process efficiency | 45% | 33% | 35% | 49% | 65% | 43% | 41% | 50% | 73% | 69% | 78% |
| More targeted marketing | 43% | 89% | 53% | 17% | 41% | 78% | 66% | 58% | - | 38% | 17% |
| Cost reduction | 42% | 33% | 35% | 37% | 35% | 30% | 41% | 31% | 45% | 56% | 61% |
| Improved risk management | 14% | 22% | 29% | 29% | 35% | 22% | 52% | 58% | 55% | 31% | 61% |
| New products | 23% | 67% | 37% | 14% | 24% | 35% | 27% | 50% | - | 19% | 33% |
| Developing information products | 26% | 33% | 44% | 31% | 12% | 22% | 23% | 19% | 9% | 19% | 11% |
| Enhanced security capabilities | 17% | 22% | 21% | 34% | 29% | 13% | 27% | 27% | 9% | 19% | 28% |
| Regulatory compliance | 11% | 22% | 18% | 23% | 18% | 9% | 25% | 23% | 27% | 31% | 44% |
| *n=* | 65 | 9 | 62 | 35 | 17 | 23 | 44 | 26 | 11 | 16 | 18 |

**Figure 22: BDA Goals by Sector**[37]

The top industries identified as BDA adopters in 2016 by Capgemini & Informatica are: (1) Retail, (2) Communications, (3) Utilities and (4) Consumer and packaged goods.

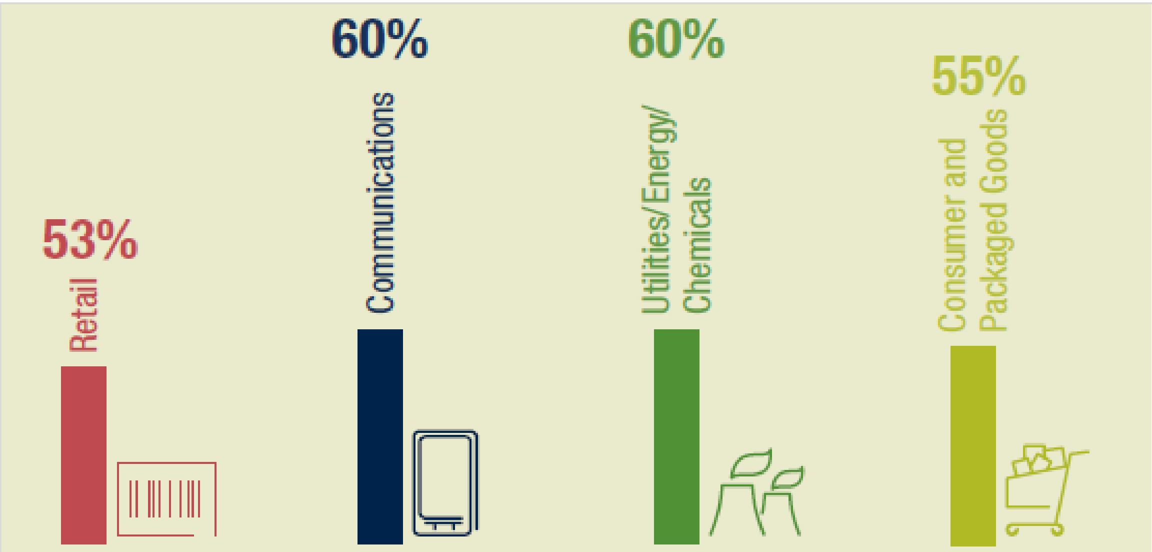


**Figure 23: Recognition of BDA Value Across Industries**[38]

IDG Consultancy proffers another view of the underlying motivation for investing in BDA initiatives. Its research in a global community of companies interested in IT solutions (n=751) in 2014 points to sundry perspectives on corporate motives. This research guides us not only to why organizations get involved in BDA but also where they think they can make an impact.

[37] Gartner, September 2015.
[38] Capgemini & Informatica, 2016: 9.


**Figure 24: BDA Goals**[39]

The data tells us, that in general business leaders feel that using BDA improves the quality of decision-making in organizations. Decisions make the greatest impact when acquiring new customers (marketing) and when "developing new products/services and revenue streams". These findings suggest that the impact exerted by BDA is rather horizontal and, therefore, its exact return on investment may be quite difficult to measure.

The impact IDG has noted on "developing new products…" seems to contrast with Gartner's studies. However, the underlying cause may be purely methodological. We contend that it may relate to how questions are formulated. IDG did not give business leaders the option to distinguish between product development and product enhancement. Thus, the impact reported here may actually entail both categories. It also includes developing new revenue streams. It is also worthwhile to remember that Gartner's research does not posit that BDA fails to exert an impact on product development. However, the trend clearly shows that over the years this motivation for investment has deteriorated substantially, suggesting that the relevant ROI is less pronounced than in other BDA projects.

The underlying reasons for running BDA initiatives give important insight into how this field of business practice will develop in the future. On the other hand, existing research may also provide insights into what types of data companies try to use when attempting to extract business value.

Unsurprisingly, Gartner (2014) shows that companies most frequently use transaction data. This type of data generates the most attention and its importance seems to be on the rise over the years. Transactions allow companies to trace their customers' buying habits and, thus, the usage of this type

[39] IDG, 2014.

of data seems to be easily translatable into courses of actions to reaching higher sales targets. Geospatial data also seems to possess great value. However, Gartner shows three categories (free-form text, email/documents and social media chat/interaction) demonstrating the importance of text mining and natural language processing. Even so, fewer companies reported using this type of data in 2014 compared to 2013.

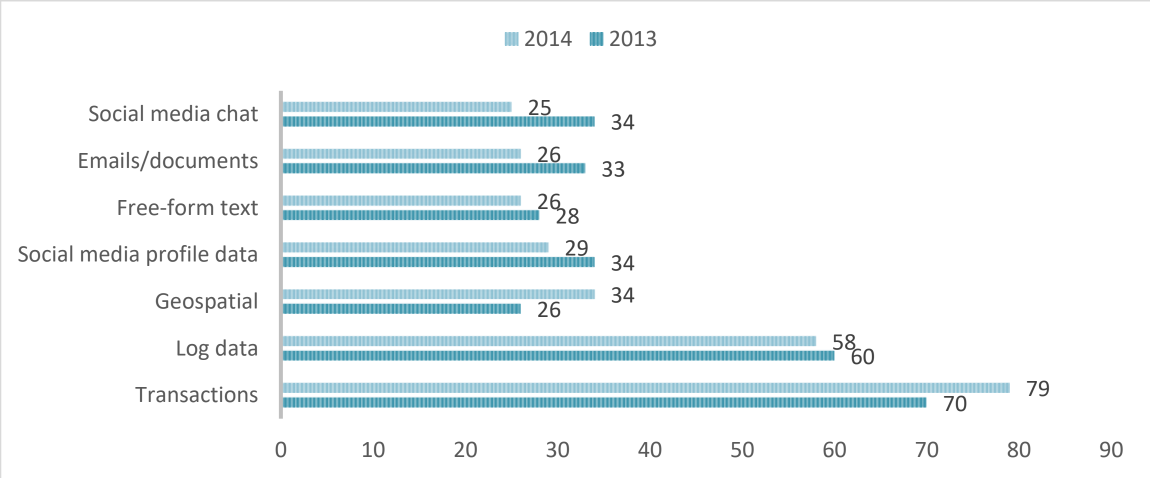


**Figure 25: What type of data do you use?**[40]

When comparing Gartner's findings to the research performed by *Forbes Insight* (2015: 8) it clearly shows that text analysis is an important piece of BDA practice - 48% of companies report using text data in their Big Data projects.

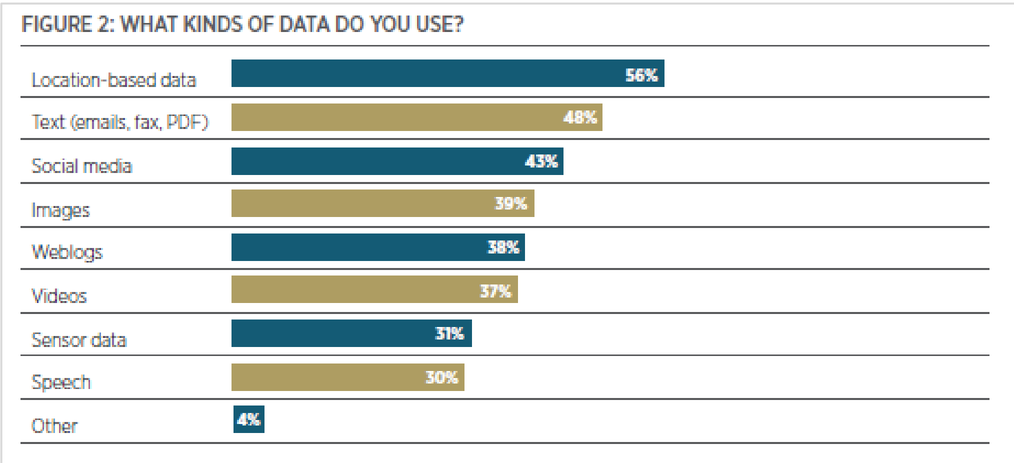


**Figure 26: Types of data used**[41]

The importance of text as a valuable mining asset is even stronger for European IT decision-makers. And it seems that text is gaining prominence among those companies that have already implemented BDA solutions. The authors of "Fujitsu Market Insights: Big Data" (2015: 3) state that "Unstructured data, specifically in the form of text, is considered twice as often by 'Big Data' implementers" compared to ones that have not experimented so far. Text mining is also not as widely recognized as a source of

[40] Gartner, 2014.
[41] Forbes Insight, 2015: 8.

business value compared to more structured data (like transactional data). However, as indicated below, its value is being recognized more strongly when processed in BDA projects.

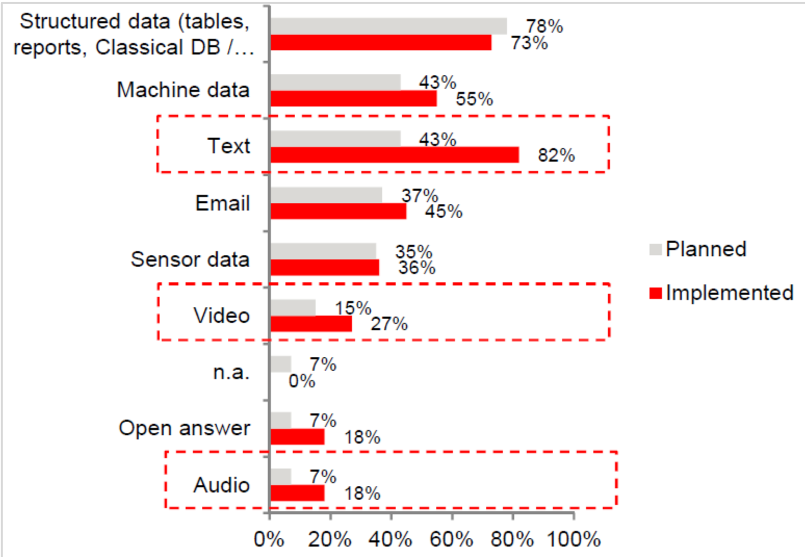


**Figure 27: Types of data analyzed or planned**[42]

### 1.1.2. Main Takeaways from Corporate Data Analysis and Case Study Illustration

The data presented and analyzed in this chapter relies mostly on responses from business practitioners and decision-makers. The main findings suggest that Big Data and Analytics are entering the mainstream. However, the stage of development is still in its infancy meaning that the uptake and usage of TDM techniques is still less than optimal. Many companies are still struggling to grasp how the usage of Big Data can boost their business value. Therefore, many companies invest in developing BDA solutions without clear guidance or strategy. Initially, they mostly invest in technology for Big Data storage and data management. This approach very often turns out to be unsustainable and leads to disillusionment. It also shows the importance of mining and analytical skills. As BDA matures, more companies feel the need to invest in TDM talent and technologies. More companies start to realize that TDM is an essential link to tie Big Data sets to business outcomes. Also, it is worthwhile to remark that BDA (and TDM) are a horizontal investment and more of a soft skill that can boost virtually all business processes. One exemplification is that business leaders find BDA to be less useful in product development than theory tends to promise. It does, however, boost corporate capabilities in the quality and speed of decision-making in various business departments. What is also important for this particular study is that natural text is often undervalued as a source of BDA. However, its value grows once companies start experimenting with it. Therefore, in the future companies may invest much more in natural text processing than they currently do to process more structured transactional data.

At the end of this section we present a case study to illustrate the approaches, strategies, challenges and complexities related to the deployment of BDA solutions, including TDM in a major European telecommunications provider. This study is based on an in-depth Interview conducted with the director of the company's internal Big Data unit.

[42] Fujitsu, 2014: 3.

**Box 8. BDA Development - A Case Study**

This brief case study presents a real case example of how a major telecom company approaches BDA. It has a strong presence in Switzerland and Italy where it offers an ultra-fast broadband network for mobile and fixed-line telephony. In recent years, the telecom industry has faced stagnant growth, a fact that has posed great challenges to operators and forced them to search for new ways of driving up revenues and profits.

*"[…]because of heavy pressure from competition, commoditization of services and price erosion, telecom companies have to reinvent themselves in order to remain competitive players in the market."*

Also, while the telecom industry enables its customers to digitize their activities, the industry itself has moved very slowly in the same direction of digitalization of its business processes and activities. This process of digitalization is reported as one of the main challenges telecoms face. Traditional corporate cultures became entrenched before the current digital age. Fragmented organizational structures and in many cases, obsolete IT systems are major problems market operators struggle to overcome. These challenges also form obstacles for developing BDA projects and capitalizing on them.
BDA is seen by the industry as a promise of great opportunities for companies whose business model is currently facing major threats. Industry decision-makers believe that by introducing BDA solutions they can circumvent market and organizational challenges by blending large amounts of information that may – ultimately – drive up revenues and profits while they realize even more efficiency gains across various business operations. Thus, the company described here has concentrated substantial efforts over the last ten years to allocate suitable resources and develop capabilities to make the most of its BDA.

***Strategy***
The company described here started to make use of data solutions in 2008, the same year that the company underwent a major general reorganization. At that time, it started to develop a Business Intelligence Center with an integrated BI team from business and IT reporting to a Central Financial Officer (CFO). The main idea was to create high-quality and robust applications by having people allocated to promote BI as a service and to test services based on customer experiences. It is thus interesting to notice that data solutions were implemented here in business intelligence and with the aim of gathering more insights into customer behavior.

As a consequence of this process, after 5 years of betting on Business Intelligence, in 2013 the company initiated a BDA strategy with a strong focus on three main goals: (1) cost savings, (2) product development and marketing improvement and (3) monetization of information. The company has demonstrated great concern to business efficiency and it has used BDA applications to improve services.

When asked about how the company uses BDA applications specifically, our interviewee said that the company adopts blended approaches. In areas in which business goals are clearly defined, the team collects data, analyzes the underlying structure and proceeds to transform this data into insight. However, he also stressed it is very often the case that in practice no previous goals can be defined in many areas. In those exploratory cases, few creative people with strong analytical skills in working with data conduct small-scale pilots.

The company envisions large potential for getting efficiency gains through BDA applications. The company is currently collecting and analyzing data on the performance of services per customer. The experience of customers in each area, time and segment is seen as a valuable source of information for the company to slash data center costs and, as a consequence, to bring down the prices of services, thereby giving the company a competitive advantage.

To explore the potential behind BDA solutions, the company has allocated large sums of capital to infrastructure – mainly on tools and hardware to process and store data and people. The interviewee mentioned that at this point investing in people has become even more important to the company. About 80% of the company's BDA budget is allocated to pay salaries of a staff of 30 professionals working daily on BDA projects.

Indeed, talent acquisition has been one of the company's key concerns. Although it receives some support from outside, most work is done internally. Over the last years, the company has not only successfully built in-house capacities for most projects by hiring skilled professionals but it has also made use of open source tools thereby considerably attenuating the company's dependence on service providers.

***Main Challenges***
Even though this company seems to be far ahead of most other companies in terms of BDA, its decision-makers are aware of the most pressing challenges associated with implementing BDA solutions: (1) talent acquisition and retention, (2) adaption of technological tools and (3) managing expectations regarding BDA.

Talent acquisition and retention are certainly a top priority. As the interviewee stressed, most data solutions need highly skilled people and because of the current "data-hype" these professionals are hard to find on the market. Moreover, most professionals are more attracted to jobs in large digital companies such as Google and Facebook. For this reason, companies operating in more traditional sectors have tried to provide skilled professionals with challenging projects and attractive salary packages as incentives for retention.

Understanding the merits behind each data approach and the corresponding complexities is not an easy task. It requires a lot of creative skill, business and statistical knowledge. This makes talent acquisition even more relevant and more difficult. Companies improve skills through training; however, our interviewee claims that in the BDA domain "training is not enough and it does not necessarily lead to better performance". Instead, the main strategy to put sufficient skill in place is to insource the right talent by offering attractive challenges to engineers and providing them with proper compensation.

The second challenge is related to the capacity of the company to adapt to new technologies. Because of the constant development of tools and the exploratory approach to data, companies in general have had a hard time in choosing the "right" technological tools to meet their needs. The right tool is not necessarily the one to be used in each instance. The interviewee cautions that openness to adapting emerging tools and understanding what is behind the available technologies are crucial if companies want to make the most of BDA.

The last challenge is associated with market expectations vs. the company's strategy. Narrow profit margins have put telecoms under constant pressure to the point of many business leaders expecting too much, too fast from BDA. The projections prepared by consulting firms on BDA's impact have contributed to some extent to creating expectations that some say are inflated and not very realistic. In the interviewee's words "analytics may bring value, but they are not a silver bullet. Time to evolve is needed". In this sense, if companies want to extract value from data, they should embrace a long-term view on the potential benefits they may reap with BDA applications. Rapid monetization may encounter impediments.

Our interviewee also referred to another extra horizontal challenge. Even though many companies still have difficulties in understanding the process of gathering data to solve a problem, businesses should not neglect the organizational part of this process. The respondent says that in general companies focus too much on the more technical parts of the process, i.e. data and software and the organizational aspect is often neglected. However, "gains depend on how the team is aligned to use the technology". Moreover, "it is necessary to transform the organization so people can make proper use of those technologies".

The importance of organizational adaptation and data value awareness becomes more evident when a company decides to introduce a new solution. As an example, the interviewee recalled his team's dedication to develop best in class marketing profiling. Regardless of the efforts of the BDA team, if marketers are ill-prepared to use these applications or have concerns about the impact of automated solutions on their work, resistance will emerge in the organization against adopting these solutions. For this reason, the respondent reminds us that businesses should be consistent in how they treat their potential impact on the organization and the underlying organizational dynamics that affect implementation.

***Risks and the Bottom Line***

The main risks in BDA are more related to a technical aspect – the quality of data from a service provider. The interviewee says that regardless of service provider, data is never of good quality, which means that the team needs to put substantial effort into making it ready for the task at hand.

However, successful outcomes have not been widely quantified by companies. When our interviewee was asked about how the company measures the value extraction from these solutions, he said he goes through this exercise all the time. Moreover, he stressed that it is possible to compare business and analytics and without analytics to obtain rough estimates, especially when one thinks of data monetization. However, he recognizes how challenging this exercise may become. “Sometimes it is almost impossible to measure in advance the exact benefits the company may extract from these solutions. There are cases in which not even rough estimates are plausible.”

He continued his line of thought by adding: “In marketing, for example, which part of the impact is brought by analytics and which part is a result of the genius of the creative director? It is not possible to say.” Experimentation, here, seems inevitable.

This company has adopted an exploratory approach to assessing the return on its investment in talent. The company has divided team activities into seven categories or “focus topics” namely: (1) corporate performance and analytics, (2) sales and services, (3) marketing, (4) product development, (5) network infrastructure, (6) data structure construction and (7) data monetization. In each category, a committee of executives responsible for the outcomes of that particular category has to assess how the capacities assigned to that specific category have been deployed. This assessment is, however, qualitative and it is mostly based on an executive’s impressions.

Although this company has built in-house capacities in recent years, it still allocates part of its BDA budget to contract service providers. When the respondent was asked if the benefits of outsourcing outweighed the costs, he said this calculation has not been done so far. However, he was sure of one thing. “If we would look at the costs of not doing so, it would be much higher than the current costs.”

# 3 MARKET VALUE OF BIG DATA AND TDM

In the previous chapter we underlined how important it is to develop talent in-house and, no less importantly, to transform organizational culture to make it more "data savvy". In its essence, developing a practical ability to translate data into business value continuously means one has to manage substantial change. Many companies use external help to facilitate this change: their demand for knowledge, technologies and skills has forged the Big Data Market, a business-to-business market in which technologies, data and services linked to Big Data are the commodities in question. In this chapter we endeavor to quantify the Big Data Market. Its financial valuation and development prospects allow us to understand how much Big Data is worth in today's economy and how this value changes over time.

Again, some parts of the market are devoted to building and developing proper database infrastructures to store ever greater data pools. However, other parts of the market focus on mining and analytical abilities to convert data into real business value. This is why understanding the Big Data market will help us estimate TDM's financial value.

**Box 9. A Reminder on Big Data and the Big Data Market**

Big data refers to large volumes of data that are structured, unstructured or partially structured. Organizations have started deploying big data solutions to access and analyze huge data sets better. These data sets, when analyzed using big data analytical solutions, can provide actionable insights to organizations. These insights help organizations take quick decisions to provide customer-centric solutions. In the past few years, big data has generated hype in the technological landscape. Moreover, the cost effectiveness of big data solutions and secure data assurance have led to exponential growth in the big data market.

The big data market established its footprint in 1999 when the term first appeared in a publication entitled Visually Exploring Gigabytes Datasets in Real Time. Later, in the 2000s, the volume of data increased to gigabytes and services, carrying the capability to handle this amount of data, came into existence. Moreover, the three Vs were coined to describe data: velocity, volume and variety. Companies started realizing the importance of analyzing the volumes of data generated by social media activities as well […].

*Source: Research and Markets (2015)*

## 3.1 Global and European Estimates

The Big Data Market is growing at a very fast clip. IDC has calculated that in 2010 the global Big Data market was worth $4.5 billion. It also estimated that this market would grow to $23.8 billion in 2016 (IDC 2012).

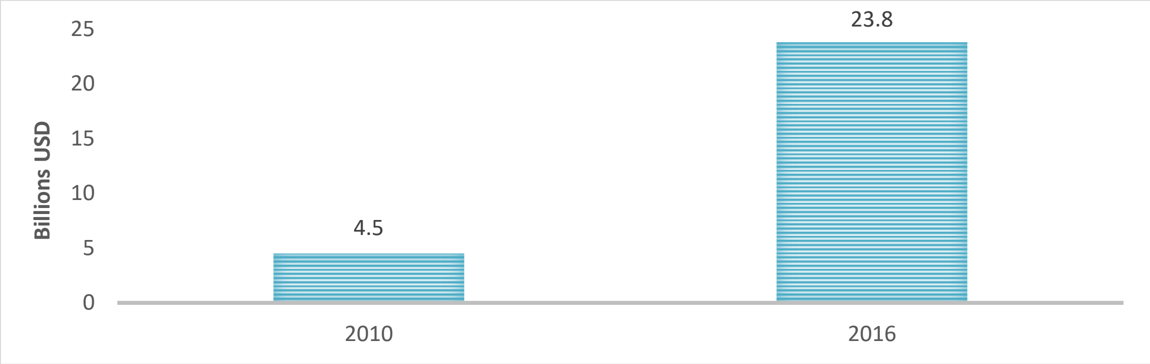


**Figure 28: Global Big Data Market Value[43]**

It is important to know that these estimates do not represent the economic value of data (in the sense of the economic gains derivable from data) but the market value of the technologies and services to manage, analyze and access Big Data. In other words, these figures represent how much global companies are spending externally to store, manage, mine and analyze data to produce valuable insights for business action.

The methodology standing behind these estimates breaks down expenditures into three main segments: (1) infrastructure, (2) software and (3) services. In 2012 companies spent $3.1 billion on the first category, $1.9 billion on the second one and $3.0 billion on the third one. The spending was estimated to grow to $10. 6 and $6.9 billion, respectively in 2016 at a high compound annual growth rate of 31.7% between 2011 and 2016.

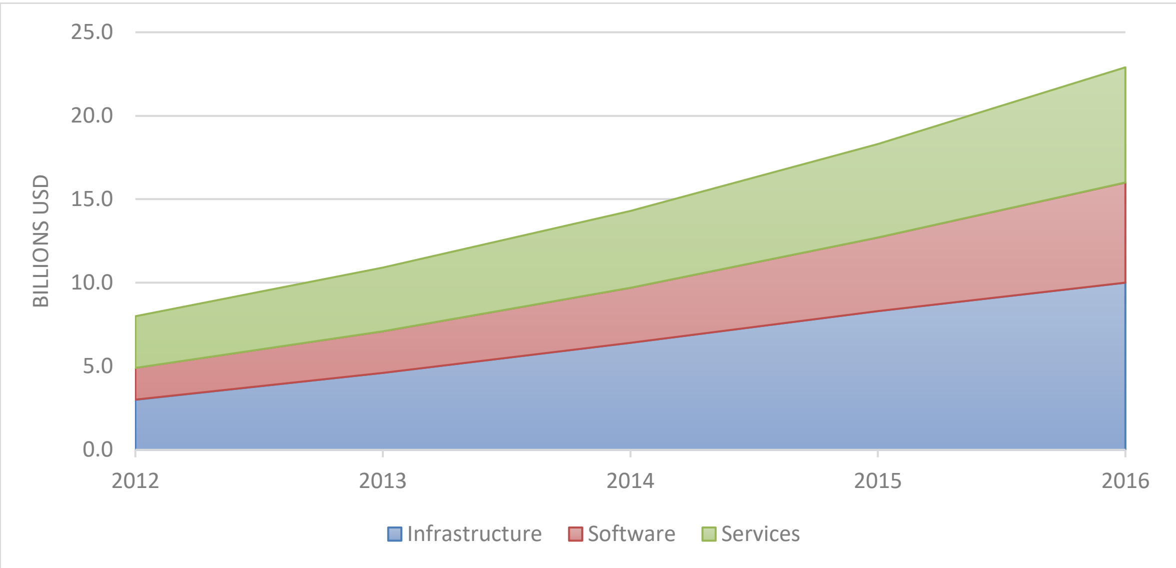


**Figure 29: Big Data Market by Type of Expenditure, 2012-2016[44]**

Infrastructure spending aims at proper storage and maintenance of data. This category has less to do with the data extraction value and more to do with preparation. This category also includes money spent on hardware infrastructure, such as servers. A detailed characterization is provided below.

[43] IDC, 2012.
[44] IDC, 2012: 5-6.

| | |
|---|---|
| 1 | External storage systems purchased by enterprises and cloud service providers and direct purchases of hard drives by large cloud service providers. (This also includes supporting storage software for device, data replication and data protection of Big Data storage assets. Internal storage installed directly on servers is included in the server segment, not the storage segment for market sizing.) |
| 2 | Server revenue (including internal storage, memory, network cards) and supporting system software as well as spending for self-built servers by large cloud service providers. |
| 3 | Datacenter network infrastructure used to support Big Data server and storage infrastructure. (Specifically, this forecast models spending based on IDC's research into the following markets: ethernet switches, fiber channel switches, InfiniBand switches and application delivery. Datacenters owned by enterprises and cloud service providers are counted.) |
| 4 | Cloud infrastructure services that combine server, storage and networking services delivered through public cloud offerings. |

**Figure 30: Types of Infrastructure Investments[45]**

As indicated, in 2012 companies spent $3.1 billion worldwide for the purpose of suitably storing data. It was also foreseen that this amount would grow to $10.8 billion in 2016 at an cumulative annual growth rate (CAGR) of 43% over a period of 5 years. This growth speaks to the vastly growing interest shown by business decision makers in organizing the digital data companies produce. One has to remember, though, that this type of investment is only a prerequisite for being able to translate stored data into actual business value.

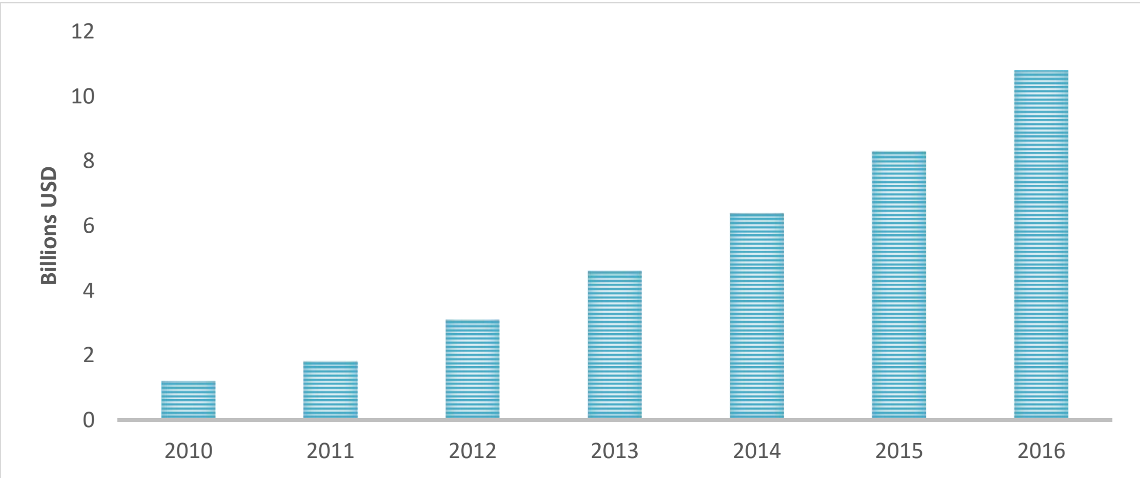


**Figure 31: Global Spending on Big Data Infrastructure[46]**

The second category of Big Data related expenditure (software) is a bit more complex. It encompasses algorithms that help in data organization and management, data mining analytics and discovery as well as industry specific programs. The details are presented below.

[45] IDC, 2012: 4.
[46] IDC, 2012: 5.

| | |
|---|---|
| 1 | Data organization and management software, including parallel and distributed file systems with global namespace, highly scalable (size and structure) relational databases, key-value pair (KVP) data stores, content management systems, graph databases, XML databases, object-oriented databases, dynamic application data stores and caches, data integration, event-driven middleware and others. |
| 2 | Analytics and discovery software, including search engines, data mining, text mining and other text analytics, rich media analysis, data visualization and other related tools. |
| 3 | Applications software including business process or industry-specific applications such as for Web clickstream analysis, fraud detection, logistics optimization and others. |

**Figure 32: Types of Software Investments[47]**

In 2012 companies spent c.a. $1.9 billion on all the above categories and it was estimated that this amount would grow to $6 billion in 2016 globally.

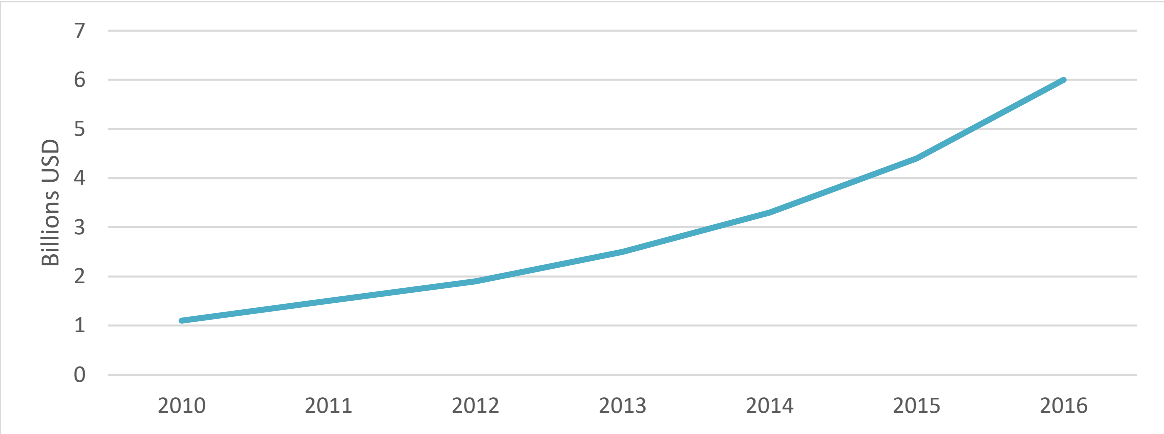


**Figure 33: Global Spending on Big Data Related Software[48]**

One can easily notice that the growth here is rather exponential portending even faster growth in the future. In fact, more recent research suggests that software will be the fastest growing component of the Big Data market between 2016 and 2019 (IDC 2015). What is more interesting, though, is that the above expenditures can be broken down according to three categories of spending allowing us to estimate the market value of TDM-related technologies and activities.

In 2012 companies spent $786.7 million on information and management software (we do not count this as TDM spending because they aim to organize data and databases). $895.2 million was spent on discovery and analytics software (direct TDM spending in our view) and $254.7 million on application software.

[47] IDC, 2012: 4.
[48] IDC, 2012: 6.

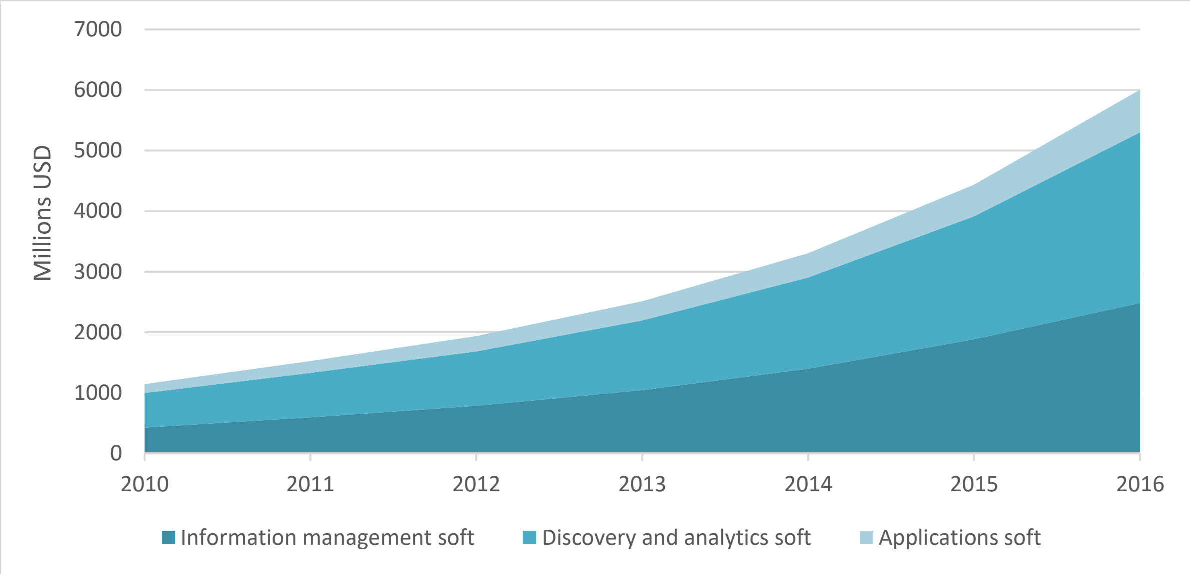


**Figure 34: Subcategories of Software Expenditure**[49]

It is especially tricky to estimate how much of this last category is attributable to TDM, as under IDC's methodology, it encompasses applications for managing business processes (which we do not count as direct TDM specific spending) and industry-specific applications for data discovery (which, in our view, is TDM specific to a given task such as fraud detection through data mining). The data does not allow us to make exact estimates, so we must venture an informed guess. Taking into account the fact that Big Data is still in its infancy and that companies need to transform their organizational culture and processes to extract value from data, we estimate that c.a. 60% of this expenditure is attributable to business process specific applications. To summarize, it may be contended that according to this data $3.2 billion will be spent globally on TDM in the "software" bucket in 2016:

$2.5 billion – information management; $2.8 billion – Discovery and analytics; $0.7 billion – Applications. $2.8 billion of this amount and 40% of $0.7 billion may be counted as direct TDM spending. This means that c.a. $3.1 billion is being spent globally on TDM in the form of software.

| Type of spending | Amount ($) | TDM Specific ($) |
|---|---|---|
| Information management | 2.5 billion | 0 |
| Discovery and analytics | 2.8 billion | 2.8 billion |
| Applications | 0.7 billion | 0.3 billion |
| *Total* | *6 billion* | *3.1 billion* |

**Table 6: TDM Spending on Software (2016)**[50]

Services are the third spending category in Big Data Market estimates. This is the "fuzziest" category encompassing "[b]usiness consulting, business process outsourcing, IT project-based services, network consulting and integration services, IT outsourcing, storage services, security services, software and hardware support and training services related to Big Data implementations" (IDC 2012). It is also a

[49] IDC, 2012: 5.
[50] based on IDC, 2012.

quite substantial category of spending totaling $3.1 billion in 2012 (39% of all Big Data related purchases) and estimated to grow at a CAGR of 21.1% to c.a. $7 billion in 2016.

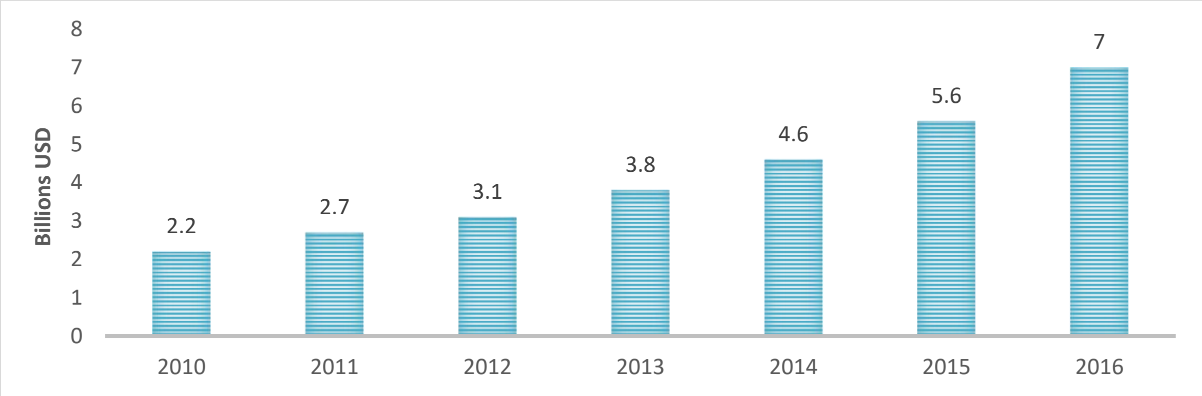


**Figure 35: Expenditure on Big Data Related Services**[51]

It is clear that some services are bought to support mining and analytics initiatives; however, because of the way this data is collected it is difficult to estimate the exact TDM expenditure. Definitely some software support and training aim to boost TDM related actions. By looking at the "software" category of spending we estimated that c.a. $2.8 billion will be spent in 2016 for purchasing mining and analytics software globally, i.e. roughly 15% of the spending for infrastructure development, information management and TDM software (stuff that needs servicing). It is thus not irrational to estimate that 15% of $7 billion is spent on extra support for TDM-related software totaling c.a. $1 billion in 2016.

Also, based on the information provided in the second chapter we know that companies in general still invest more in infrastructure and database maintenance than in developing TDM. However, they are starting to feel considerable demand to mine talent and tweak their organizational culture, thereby calling for proper training. We estimate that up to $2 billion (of $7 billion) is being spent (in 2016) on training on mining and analytics. We assess that in the "services" category of spending, roughly $3 billion is TDM specific.

2016 estimates suggest that a total of $23.8 billion is spent globally on the Big Data Market, with $6.2 billion specifically for TDM (software purchases, support and training). $10.8 billion is dedicated to secure proper infrastructure for storing the data, $2.5 billion is devoted to information management and the remaining $4.3 billion is used for other types of services.

[51] IDC, 2012: 6.

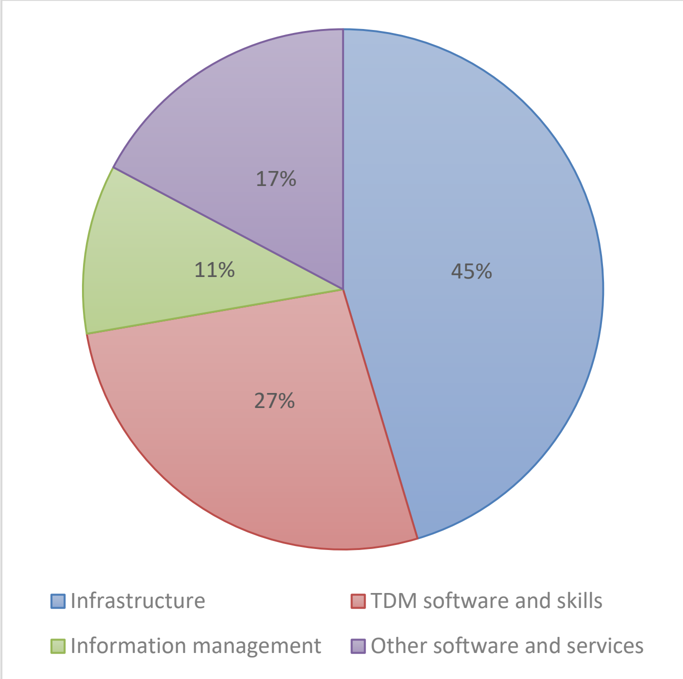


**Figure 36: Big Data Market Showing TDM Related Expenditure[52]**

At this point, an important remark must be made; namely, these calculations are based on estimates made in 2012 due to the dearth of more recent data to extract TDM's value. However, IDC recently amended its estimates showing that the value of the Big Data Market will be $26.1 billion in 2016 (IDC 2015) and will rise to $48.6 billion in 2019. This means that IDC underestimated its 2012 calculations and that the Big Data Market is picking up the pace. **Using the calculations based on 2012 estimates and the adjustments made in 2015, the market value of global TDM spending will be c.a. $7 billion (27% of the total) in 2016.**

More recent data does not give reveal market structure. The data does suggest that from 2016 "software" will drive the market's spending. This also means that TDM will gain in importance *vis-à-vis* other types of Big Data expenditures. These findings are consistent with more qualitative information about the Big Data usage we presented in Chapter 2 – companies become more fluent in building and managing databases (data lakes) and are in more need for mining and analytical capabilities. Thus, more money will be spent on these aspects in the future.

From 2014 to 2019 software is expected to grow at a CAGR of 26.2%, services at a CAGR of 22.7% and infrastructure at a CAGR of 21.7%. Previous estimates suggested that infrastructure spending would grow at a CAGR of 43% from 2011 to 2016. This has been adjusted to 21.7% in the next period of Big Data Market development. **It is thus not unreasonable to predict that TDM-related spending will account for 35% of the overall Big Data Market in 2019, giving it a price tag of $17 billion.**

***Europe***

The European Big Data Market is expected to have a value of $9.37 billion in 2016 and is projected to grow at a CAGR of 25.7% to $29.43 billion in 2021 (Research and Markets, 2016). Research shows that

[52] Proprietary estimates based on IDC, 2012.

it is the second largest Big Data Market after North America, valued at c.a. $16 billion. It should be noted that the European Market is expected to grow faster than the North American market (see the Figure below).

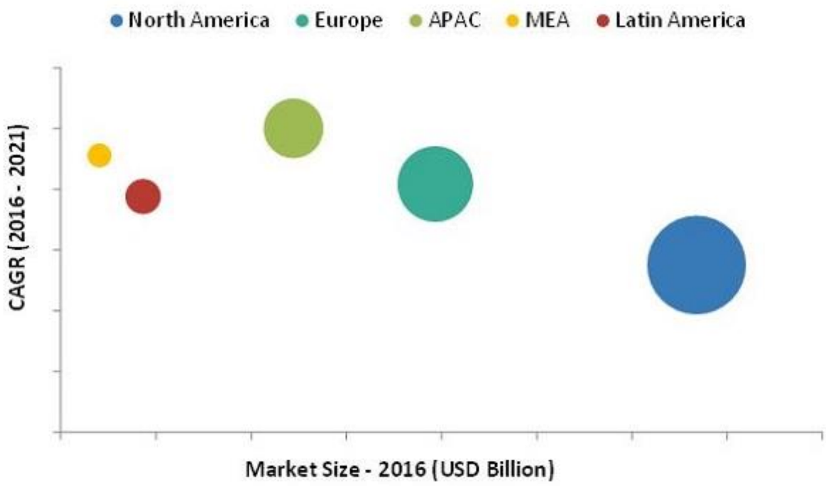


**Figure 37: Big Data Market by World Region[53]**

Using global data calculations (to extract and estimate TDM-related expenditures) we can roughly estimate that European TDM will be worth c.a. $2.5 billion in 2016 and it could grow to c.a. $10.3 billion in 2021.

## 3.2 Broader Perspective on Data Value

The calculations and estimates presented above are specific to Big Data and Analytics. One can, however, assume that data mining technologies also apply to more traditional datasets. Thus, taking this broader perspective on data usage by companies may highlight the even greater importance posed by data (regardless of whether it is big or not) and TDM.

Europe has seen the rise of "data companies" in recent years, companies "whose main activity is the production and delivery of digital data-related products, services and technologies" (IDC 2015b). They still form a niche that is mostly present and active in the UK (46%) and Germany (11%). It is estimated that in 2014 c.a. 243 thousand companies existed in the EU, accounting for 14% of the 1.7 million enterprises operating in the ICT and professional services sector (IDC 2015b: 18).

On the other hand, there is a growing population of data users, companies that strongly rely on data to run their operations. The total number of such EU-based companies in 2014 was estimated to be 642 thousand. This number constitutes 6.3% of the 10.3 million potential data users. It is apparent that there is much room for improvement. It is also worthwhile to remark that large companies are driving the usage of data as a business asset and they are mostly located in the UK and Germany (39%). These two countries will drive the growth of data usage in the near future.

The value of the data market was estimated to be $66 billion in 2014 ("meaning the aggregate value of the data products and services exchanged in the EU bought by European businesses and

[53] MarketsandMarkets Analysis.

consumers") (IDC 2015b: 19). It is much harder to estimate the role TDM plays in the overall data market. Many products and services traded on this market are more traditional than automated discovery and pattern recognition technologies. With certainty we can claim that roughly 12% (c.a. $8 billion) of the data market was (in 2014) devoted to Big Data and 27% of that is attributable to TDM, i.e. c.a. $2.2 billion. Nevertheless, what portion of the remaining $59 billion of the overall European data market was spent on TDM is still shrouded in mystery because of the lack of detailed data.

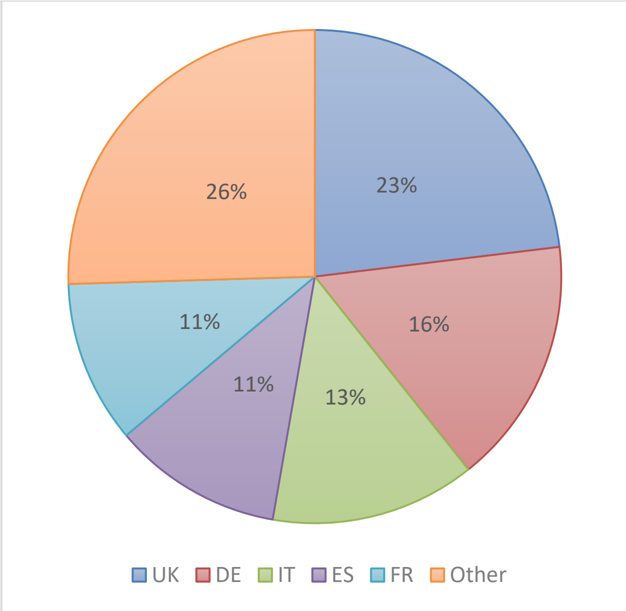


**Figure 38: Share of Data Users by Country (total = 642 K)[54]**

What we can say is that the importance of Big Data related to TDM will grow over time. The growth rate of the overall data market was 6.4% (between 2013 and 2014) while the projected pace of growth on the Big Data Market is estimated to be 25.7% from 2016 to 2021. The Big Data Market is expected to grow four times faster. Therefore, a rough estimate would allow us to assume that by 2021 the Big Data Market will constitute c.a. 35% of the EU's overall data market.

The value of data analytics may also be expressed through the importance of the data workforce. In 2014 the EU had 6.1 million data worker, i.e. professionals who "collect, store, manage and analyze data as their primary activity" (*Ibidem: 17*). This figure constituted more than 3% of the EU's total employment. Its growth rate was 5.7% compared to the EU average of 2.2%. This rate shows that data workers are driving EU job rates and this industry embodies the bulk of the European growth opportunity.

## 3.3 Economic Impact

The interesting part in showing the importance of TDM (and data in general) is its impact on the economy. The calculations presented above show only direct impacts "generated by the [big] data industry itself; they represent the activity engendered by all businesses active in the data production. The quantitative direct impacts will then be measured by the revenues from data products and services sold, i.e. the value of the data market." (IDC 2015b: 145). However, it is worthwhile to recall that both

[54] IDC, 2015b: 103.

Big Data and All Data markets are B2B markets. Their existence and growth rely purely on the demand created by companies. This indicates directly that companies treat those expenditures as investment to optimize their operations.

Therefore, we can speak of "forward indirect impacts". The consumption of data products and services will cause "such impacts including economic growth depending on data uptake by downstream industries. For user companies, data is now a relevant factor of production; the adoption of data products and services by downstream industries provides different types of competitive advantage and productivity gains to user industries) (*Ibidem*).

One can also cite reverse indirect impacts. "These impacts represent business growth resulting from changes in sales from suppliers to the data industry. In order to produce and deliver data products and services, data companies need inputs from other stakeholders" (*Ibidem*).

The last type of economic impact to be taken into account refers to induced impacts. "These impacts include economic activity created by the additional payment of wages to staff in the data industry and its direct supply chain. Some of this will be spent on consumer goods and services. This leads to further business growth in the EU economy. The additional consumption of data workers and data companies' suppliers will in fact support economic activity in various industries such as retail, consumer goods, banks, entertainment, etc." (*Ibidem*).

IDC presented a model that calculated the overall economic impact of data on the European economy to be $321.1 billion in 2013, accounting for 1.8% of European GDP at that time. In 2013 the value of the overall data market was $61.6 billion. This means that for each $1 invested to buy data-related products and services, $5.2 of extra value was produced. To estimate TDM's financial value, one may use this general data-to-value ratio – as TDM constitutes an element of the data-to-value story. **One could therefore claim that $2.5 billion spent on TDM in 2016 will produce an economic impact of $13 billion on the European economy. This impact will grow to $53.6 billion in 2021.**

Now, one has to remember that these estimates are very rough and merely present a starting point for much more detailed calculations requiring a labor intensive gathering of TDM data from European companies. To do this, the approach to researching Big Data and/or Data Markets would have to change dramatically.

However, we estimate, that the impact exerted by Big Data related to TDM may be much higher than the impact exerted by a more traditional usage of data. Buchholtz et al. appears to voice this argument by presenting a stochastic model calculating the impact of Big and Open Data on the European economy. The authors claim Big Data would bring c.a. €200 billion of extra value to the European economy. Previous estimates show that the European Big Data Market will be worth c.a. €18.7 billion in 2020. This means that according to the BOUDICA model €1 spent on Big Data products and solutions will translate into €10.7 of overall economic value. This also means that the economic impact of Big Data is estimated to be to be over twice as powerful as the economic impact of more traditional data. Taking all this into account one could claim that TDM's impact on the European Economy will range from $13 billion (conservative calculation) to $26.7 billion (optimum scenario) in 2016. Additionally, its minimum value will range from $53.6 billion to $110.1 billion in 2020. In fact, we believe, that the latter value is more accurate.

# 4 QUANTIFICATION OF BIG AND OPEN DATA ECONOMIC IMPACT

## 4.1 Introduction

In the previous chapter, we presented an estimate of TDM's financial value and its impact on the European economy. To do this, we used the work of Buchholtz et al. (2014) including their calculations of the impact exerted by Big Data on the European economy. Because their calculations form an important assumption for estimating TDM's economic value, we present a discussion of BOUDICA in this chapter. This stochastic model (Big and Open Data Universal Impact Assessment) was developed by Buchholtz et al. BOUDICA authors estimated that the economic gains derived from Big and Open Data will give the European economy an incremental boost of 1.9% by 2020 as a result of economic gains at the micro level that affect the macro level. Through their work, it is also possible to conclude that **€**1 invested in Big Data solutions will bring **€**10 of value to the European economy.

**Box 10. BOUDICA**

BOUDICA is a bottom-up macroeconomic model developed by Buchholtz, Bukowski & Śniegocki from the Warsaw Institute for Economic Studies (WISE) to estimate the economic potential of Big and Open Data in all 28 EU member states in the next decade.

It was presented in a report entitled "Big and Open Data in Europe: a growth engine or a missed opportunity?" (2014) under a study commissioned by the Center for European Strategy Foundation (demosEUROPA) and sponsored by Microsoft.

This study aimed to estimate the European ability to exploit the economic potential of data-driven solutions through BOUDICA by decomposing the expected economic effect of Big and Open Data to the European GDP into 21 sectors in 28 countries.

Before we explain our main reasons for adopting BOUDICA and present the model and its main aspects in more detail, we believe that some key points on BOUDICA are worth clarifying:

**1. Composition**: Although BOUDICA consists of two parts, namely Big Data and Open Data, the model focuses mainly on Big Data and presents Open Data as an addition. As we will see in this chapter, the economic impact predicted by Buchholtz et. al with regard to Open Data is low in comparison with Big Data.

**2. Conceptual Understanding**: It is also important to note BOUDICA's authors' understanding of Big Data. As mentioned above, there are two ways of interpreting Big Data: as a descriptor of Big Data sets and as a metaphor encompassing the use of data to produce business value. BOUDICA was built under the latter broader understanding of Big Data and, thus, encompasses TDM and other processes needed to translate Big Data into actionable intelligence and later into value.

It is important to remember that we approach TDM as a key part of Big Data. The reasoning behind the relationship between Big Data and TDM is that Big Data is a *sine qua non* condition for TDM activity to exist. It is through and because of large data sets that the analytical process of exploring data to detect patterns and relationships can be carried out. It therefore becomes clear that the fast and ever-growing pool of varied data constituting Big Data is more than a resource for TDM; it is also a dimension in which TDM has the potential to create value if it discovers new data applications. In other words, TDM is the next step in the Big Data revolution as it not only presupposes the existence of data infrastructure and databases containing large datasets but also offers organizations the possibility of extracting value from abundant data sources.

In reality, it is our argument that the full value of Big Data can only be extracted by linking datasets through data mining and analytics. As more organizations mature in deploying BDA applications, i.e. move from pilot experimentation to sound presence and effective use of data storage and infrastructure, the role of data mining and analytics tools to translate data into value tends to become even more pronounced and this trend is expected to strengthen in coming years. Indeed, more and more companies recognize the need for proactive action to connect their data infrastructure to business problems and consequently business outcomes. In that sense, data mining and analytics offer great opportunities for value creation at the micro level that may produce positive effects for the whole economy.

For these reasons, our study relies on BOUDICA to estimate TDM's influence as a subset of the economic impact exerted by Big Data. BOUDICA is the most accurate and updated model capable of capturing ICT developments and their adoption in Europe. It estimates the overall economic impact of Big and Open Data by encompassing the potential behind data mining and analytics and aspects related to data infrastructure building and maintenance, such as processing power, larger transmission and storage capacity. All these components are crucial in the ICT revolution. To reach this end, the bottom-up BOUDICA model analyzes the economic gains derived from those ICT developments in organizations hailing from 21 sectors and translates them into three key forces affecting the economy on an aggregate level: (1) productivity uplift in manufacturing and services, (2) competition improvements and (3) improved allocation of production factors.

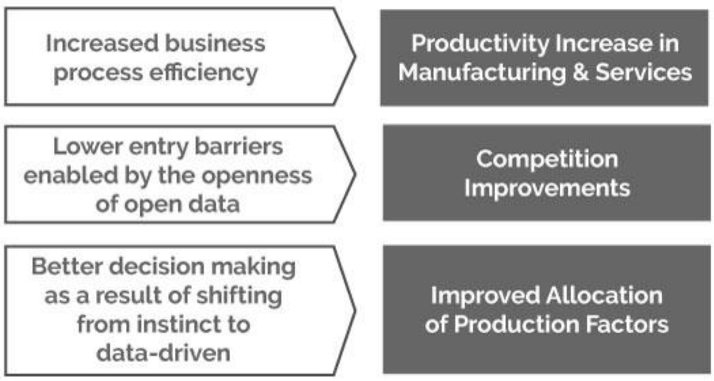


**Figure 39: Key forces analyzed in BOUDICA**[55]

Another element in the BOUDICA model is company size. Researchers posit that company size is a variable that measures a company's likelihood to exploit ICT solutions. The assumption is that the larger a company, the higher its propensity to engage in Big Data projects. As we mentioned above, a Big Data program demands dedicated resources and in general it is easier for large companies to access resources. Furthermore, the model assumes that the adoption rate of ICT solutions in companies tends

[55] Buchholtz et al., 2014.

to be higher over time as more companies start to grasp the potential of Big and Open Data applications for their businesses.

BOUDICA has proven itself in capturing the latest ICT advancements in European economies, sectors and company sizes. It serves as a consistent basis for us to produce estimates of TDM's impact as a subset of the overall economic impact exerted by Big and Open Data. We would like to explain in greater detail how BOUDICA may be used to measure the economic impact exerted by Big and Open Data in Europe. After that, we will present the study's key findings. Finally, we will summarize our understanding of the BOUDICA model.

## 4.2 Model Structure and Scope

The bottom-up BOUDICA model estimates the impact exerted by Big and Open Data on companies in various sectors.

The detailed data on the European economy and the classification of economic activities are taken from Eurostat's database. The base year for all calculations is 2010 because it is the last year available on the database for taking a cross-country sample and results are presented for years from 2013 to 2020.

To measure the total impact exerted by Big and Open Data on the EU-28 GDP in a given year, the model sums the impact exerted by Big and Open Data solutions verified using three key forces that should affect the economy at the macro level, namely: (1) data-driven management, (2) fiercer competition and (3) efficiency improvements. It multiplies this sum by the total value added. The key figures describing the effects and the total value added are compiled from different sized companies operating in one of the sectors in each EU member state a number of years after the first year of calculations.

The model is flexible and allows for adaptations to obtain more detailed estimates by sector, company size and country. All output data from the sectors was taken from the Structural Business Statistics and National Accounts Detailed Breakdowns in the Eurostat database.

| Accommodation | Financial and insurance activities | Information and communication technology |
|---|---|---|
| Wholesale and retail trade | Manufacturing | Electricity, water and other utilities |
| Transportation and storage | Administrative services | Real estate |
| Activities of households | Construction | Professional, scientific and technical activities |
| Public administration and defense | Activities of extraterritorial organizations and bodies | Human health and social work activities |
| Agriculture | Arts and recreation | Mining |
| Education | Manufacturing and other services. | - |

**Table 7: Sectors Analyzed in BOUDICA**[56]

For the purpose of this quantitative analysis, the three groups of company sizes were defined according to the following number of employees – (1) small companies with fewer than 50 employees, (2) mid-

[56] Buchholtz et al., 2014.

sized companies with 50 to 249 employees and (3) large companies with more than 250 employees. Countries are gathered in three different buckets. However, Big and Open Data implementation effects were calculated for each country separately.

| Northern Europe | Southern Europe | New Member States |
|---|---|---|
| Belgium | Portugal | Bulgaria |
| Denmark | Spain | Czech Republic |
| Germany | Italy | Estonia |
| Ireland | Greece | Croatia |
| France | | Cyprus |
| Luxembourg | | Latvia |
| The Netherlands | | Lithuania |
| Austria | | Hungary |
| Finland | | Malta |
| Sweden | | Poland |
| United Kingdom | | Romania |
| | | Slovenia |
| | | Slovakia |

**Table 8: Country Groups in BOUDICA[57]**

## 4.3 Big and Open Data Impact Assessment - Key Effects and Indicators in BOUDICA

The BOUDICA model assumes that the impact of Big and Open Data will gradually spread across the economy in so far as more companies start to take advantage of the opportunities created by innovative data solutions and realize the benefits in productivity growth on the microeconomic level.

The estimates for the impact of Big and Open Data were calculated as follows:

The percentage of enterprises altered by one of the macroeconomic factors multiplied by the impact parameters related to the expansion of a certain technology across time. To be more specific, the impact parameter is a time function that incorporates the differences in technology adoption between the year to be analyzed and the year in which a group of enterprises was affected by that technology for the first time.

Below we verify in detail the impact parameters from the literature used in BOUDICA for each macroeconomic factor and the underlying determinants. Then, we present the factors related to ICT usage in companies leading to the ICT index.

### 1.1.3. Impact Parameters

**Data-driven decision-making**

The impact parameter used in the analysis for the data-driven decision making (DDD) effect was based on Brynjolfsson et al. (2011) studies – one of the first large-scale studies on the relationship between DDD and company performance. By applying a standard econometric method to survey and financial data on large cap publicly-traded firms, the authors came to the conclusion that companies that adopt

[57] Buchholtz et al., 2014.

DDD technologies sport a 5-6% increase in output and productivity above and beyond what can be explained by traditional inputs and IT usage.

To reach this productivity growth rate, researchers relied on business practices and information systems' measures estimated after a survey conducted with large cap publicly-traded companies in 2008. The questions pertained to IT usage and workplace organization, innovative activities, the usage of information for decision-making and the consistency of their business practices. To explore the impact of DDD, researchers first used the survey response to build measures of organizational practices in companies. Then, these measures were combined with publicly available financial data. As a result, 179 companies from all major sectors provided complete data for the research analysis on company productivity from 2005 to 2009.

| | |
|---|---|
| **Data-Driven Decision-Making** | This key variable was built from three survey questions: (1) the usage of data to create new products or services; (2) the usage of data for business decision-making in the entire company and (3) the existence of data for decision-making in the entire company.<br><br>DDD was produced by first standardizing each factor with a mean of zero and a standard deviation of 1. Then, the sum of these factors could be standardized to produce an index. |
| **Adjustment Cost** | This measure was created from 6 survey questions in which companies describe the degree to which the following factors facilitate organizational changes: (1) financial resources, (2) skill mix of current staff, (3) employment contracts, (4) work rules, (5) organizational culture, (6) customer relationships and (7) senior management involvement.<br><br>The composite index was built similarly to DDD. |
| **Consistency of Business Practices** | This was built using six survey questions on the consistency of business practices: (1) across units, within business units, across functions and across geographies; (2) the effectiveness of IT for supporting consistent practices; and (3) consistency of prioritization of projects.<br><br>The composite index was built similarly to DDD. |
| **Exploration (EXPR)** | It was created using eight survey questions on a company's likelihood to explore a new technology or market and engage in radical innovation. Because a company's age is one of the instruments in the study that may be correlated with a company's innovative activity, EXPR was used as a control variable.<br><br>The composite index was also built similarly to DDD. |
| **Human Capital** | The average worker's wage and the common employee's education were used as a proxy for the company's human capital. |

**Table 9: Measures of Companies' Business Practices[58]**

Apart from measures concerning business practices, the study also includes other data to obtain measures for production inputs and performance, company age and information technology staff.

[58] Brynjolfsson et al., 2011.

| | |
|---|---|
| Production Inputs and Performance | Measures of materials, physical assets, employees and sales and operating income. When it comes to materials, it must be added that this measure was estimated by subtracting operating income before tax and labor cost from sales. |
| Company Age | Collected from a semi-structured data website. If multiple companies emerged, the company's year established was used assuming name retention or the founding year of the oldest company in a merger. |
| Information Technology Staff | Obtained from survey questions on IT budgets, outsourcing, change of IT budgets from 2008 to 2009 and full time IT employment. |

**Table 10: Additional Data**[59]

By using existing performance measures from the literature in combination with additional instruments, namely company age and consistency of business practices, the authors of the study developed tests on productivity, business profitability and companies' market value.

Concerning productivity, production function variables were analyzed through a multifactor productivity approach, that is, relating a measure of company output – sales or value-added, to company inputs, such as capital, labor and capital and IT-related labor. The key factors treated as company inputs were materials, physical capital, the number of IT employees, the number of non-IT employees and data-driven decision-making variables. With regard to business profitability, three performance measures were used: return on assets (ROA), return on equity (ROE) and asset utilization. Companies' market value through financial market-based measures was included in the analysis because it could provide more realistic results in combination with business profitability measures, basically accounting measures that may not on their own capture trends, intangible assets and time lags necessary for realizing the potential of organizational change.

The study's findings show that incremental productivity may be explained by the DDD differences among companies with the same level of IT use. Moreover, the business profitability test shows that IT seems to be significantly correlated with ROA and asset utilization, but not with ROE. Concerning the company market value test, analysis suggests that companies that adopt DDD have a higher market value, which is most closely related to their level of IT capital. On this point, the result shows that each employee is associated with $8.2 thousand of IT capital stock and that companies that invest in DDD have an additional value of $3.1 thousand per employee for each standard deviation of DDD above the mean.

However, it is interesting to see other research analysis results. The study found that the relationship between companies' IT staff and DDD applications explains only 14% of business performance improvement and the level of correlation between IT budget and DDD is also very low. It accounts for a mere 13% of companies' performance improvement. In fact, DDD usage and other organizational elements tend to have a higher impact on performance along the order of 20%.

Overall, this impact parameter suggests that companies' DDD capabilities can be modeled as intangible assets and their performance may improve further as a result of adjustment costs and consistency of business practices. From this remark, it is reasonable to conclude that TDM activities (used as a means

[59] *Brynjolfsson et al., 2011.*

to translate data into decisions) find greater potential to expand depending on the consistency of internal processes across the whole business.

**Competition Improvements**

With regard to competition improvements from lower entry barriers with the opening of public sector data, the impact parameter developed by WISE was based on the Etro study (2009) on the economics of cloud computing.

In his study on cloud computing, Etro analyzes the economic impact exerted by the gradual introduction of cloud computing by adopting a simple endogenous market structure approach to macroeconomics. The experiment covered most companies domiciled in the EU in terms of their numbers, accounting for more than 17 million companies and focused on a few aggregate sectors in the EU, namely manufacturing, wholesale and retail trade, hotels and restaurants, transport storage and communication and real state renting and business activities.

Cloud computing has a key role in fostering business creation and competition as a result of lower ICT fixed costs. This may reduce the initial fixed investments made by a new business in hardware, software and labor. Once a business is active on the market, each company competes with rivals at a given production level. On this last aspect, an increase in the marginal cost of production is treated as endogenous and depends on companies' ICT choices. To be more specific, it depends on how much hardware and software a company will use to match its necessities and on the rate of return generated on ICT capital measured in units of final output.

The model captures the process of cost reduction enabled by ICT by searching for the fixed cost of entry that constraints business creation in a given sector and at a given time. By assuming that changes in fixed cost and the speed of adoption of a new technology depict the absolute impact exerted by cloud computing, the author calculated a range of fixed cost savings in the long run ranging from 1% to 5%.

The expansion of cloud computing means that storage capacity and infrastructure may no longer constitute significant barriers for SMEs. In fact, centralized software management and data storage enabled by cloud computing will eventually lead SMEs not only to enter the market, but also to pursue, TDM activities to a greater extent.

The Buchholtz (et al.) study focused mainly on public and community clouds to verify the impact exerted by cloud computing on sectors. Although the authors recognize that the impact parameter for this effect varies by sector, their estimates point to an increase of 0.2% on average in the value added in sectors.

It seems that, regardless of sector, once an SME becomes active in the market due to lower entry costs, it may quickly adopt data-driven decision-making solutions as a result of more flexible processes. As a consequence, these applications may contribute to a greater capacity to extract value from data, which may produce even larger disruptive effects in the market, especially in more traditional sectors.

**Business Efficiency**

The impact estimates for this macroeconomic factor were based on the McKinsey Global Institute (2011) study and then assigned by business process. The study estimates a 10% productivity increase in retail and wholesale trade from efficiency developments in supply chain management and in manufacturing with regard to improvements in industrial processes. Efficiency improvements in fraud and error detection produce impact estimates of 2.5% productivity growth in the financial industry and public administration.

The percentage of enterprises affected by these three effects takes into account differences in adoption rates between the year to be analyzed and the base year and these values were modeled for enterprises of different sizes and operating in different sectors and countries. Adoption rates may be calculated for each year in every country and sector and take into account company size. To verify how the adoption of ICT solutions has evolved, the study used survey data from Information Society Statistics available on the Eurostat database on the adoption of previous ICT innovations by companies. ICT innovations in the survey data include homepages, DSL, LAN and Intranet/extranet, online banking and financial services usage and fixed broadband access.

The pace of adoption was then scaled to account for different saturation levels across companies of different size, companies and countries. This endeavor was facilitated by an additional impact-specific index extracted from the relevant literature and an ICT index was derived by blending data in surveys on ICT adoption.

**Additional Indices**

The additional impact indices used in the analysis for data-driven decision making and efficiency improvements were based on the McKinsey Global Institute (2011) and Cebr (2012) studies on sectoral readiness for big data solutions. The increased competition emerging from opening access to data as an additional index was based on the Deloitte study (2013) and the ePSI Scoreboard on sectoral and country-level differences. For data-driven decision making data the additional index was adjusted upwards after giving consideration to the additional motivational effects produced by public sector open data in health and education.

**ICT Index**

The indicator values in the ICT index are averages of available data from 2010-2012 standardized by WISE in such a way that the average for every indicator for all enterprises in the EU is 100%. Indicators are grouped into the following categories:

| Category | Eurostat Indicator | Weight |
|---|---|---|
| Online Business Activity | Enterprises purchasing online (at least 1% of orders) | 10% |
| | Enterprises selling online (at least 1% of turnover) | 10% |
| Internal Infrastructure | Enterprises where persons employed have access to personal human resources services electronically | 10% |
| | Enterprises which electronically share information on purchases with the software used for any internal function | 10% |
| | Enterprises which electronically share information on sales with the software used for any internal function | 10% |

| | | |
|---|---|---|
| Data Exchange | Enterprises which electronically share information suitable for automatic processing with external business partners or on the SCM with suppliers or customers | 10% |
| | Enterprises sending/ receiving e-invoices in a standard structure suitable for automatic processing | 10% |
| | Enterprises using automated data exchange with other ICT systems outside the own enterprise | 10% |
| Data Collection | Enterprises using Radio Frequency Identification (RFID) | 20% |

**Table 11: ICT Index**

The ICT index is applied to all countries, company sizes and most sectors under the classification of economic activities of the European Community – NACE. Sectors not covered by NACE were estimated based on average index values for a given country. The ICT index varies by company size and tends to be higher in large companies, especially in northern Europe. The ICT index values differ significantly by country. Since sectoral values vary significantly, we conclude that some sectors invest in infrastructure and adopt ICT solutions more than others and for this reason they benefit more from potential positive returns related to Big Data.

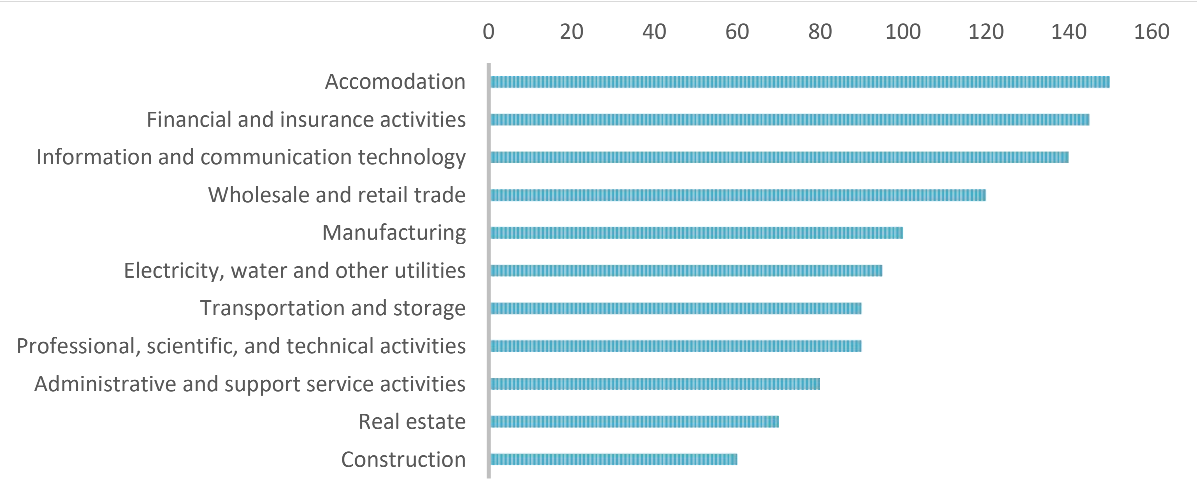


**Figure 40: ICT index values per Industry and Service (%)**

## 4.4 Research Findings

According to the estimates presented by Buchholtz et al. (2014), the total annual benefits generated by Big and Open Data in Europe will be €206 billion by 2020. Although this total economic impact comprises both Big and Open Data[60], the study analyzed those sources of data together and separately to verify the impact of each source and the benefits of combining them. Undoubtedly, one of the main

[60] The report draws a very simple distinction between Big and Open Data: Big Data is owned by private companies while Open Data is understood to be Public Sector Information available to the public. We do not think that this simplified classification is correct. As we have indicated in this report, we follow the definition in this part in accordance with the quoted report.

conclusions of the study is that the largest contribution to EU's GDP will come mainly from companies' proprietary Big Data. The extra value boost Big Data affords to the European Economy is expected to be c.a. €196 billion by 2020. Therefore, the contribution made by Open Public Sector Data to the total economic impact will be very modest but no less important. It will accrue a total extra value of c.a. €10 billion by 2020. In reality, Open Data will have a complementary role in ICT development when it is used in combination with Big Data.

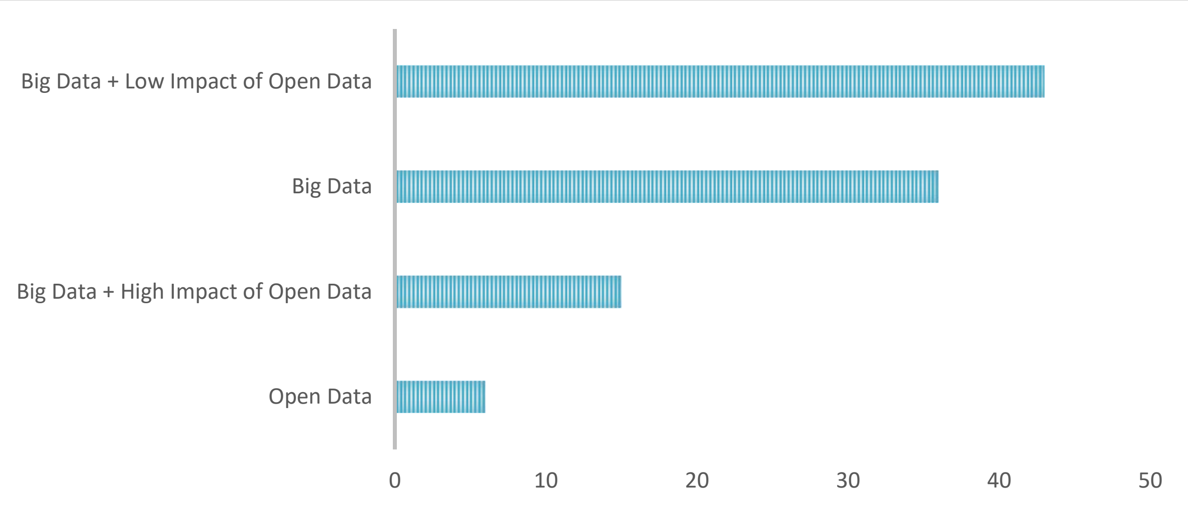


**Figure 41: Additional GDP by Source of Impact - Economic potential of Big and Open Data (%)[61]**

Through scenario-based value estimates, the authors of BOUDICA concluded that Big Data tends to produce larger gains when compared with Open Data. The broad productivity increases from Big Data are expected to be much higher than the direct impact of the fiercer competition stimulated by opening public sector data. Moreover, Open Data from the public sector, according to estimates presented in the Buchholtz (et al.) study can be broadly applied to roughly 10% of the economy and significant effects may materialize in another third of the economy. This means that the indirect impact generated by opening public sector data may allow for higher economic gains especially when coupled with big data. The authors of the study point out that Open Data may create opportunities for sectors recently affected by financial and public debt crisis, such as the finance, real estate and public administration sectors, as well as better prepare Europe for challenges related to its aging population and building a knowledge economy. In this sense, health services and knowledge-intensive sectors may highly benefit from Open Data.

Besides those key findings on impact sources, the study concluded that the additional value of Big and Open Data to the European economy tends to vary significantly by macroeconomic force, sector and country group as we will see below in more detail.

### 1.1.4. Macroeconomic Forces

The study shows that the three major economic forces are predicted to generate gains from better analytics and applications. By far, the largest contribution at the aggregate level will come from

[61] Buchholtz et al., 2014.

improved allocation of production factors, followed by the impact caused by productivity increase due to improved business efficiency and competition improvements, respectively.

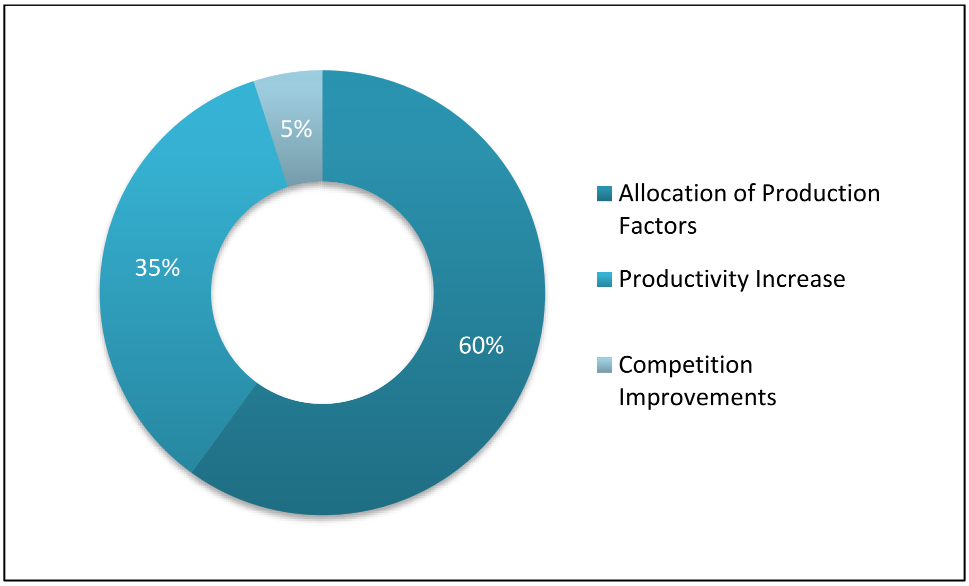


**Figure 42: Impact of each force at the aggregate level in 2020**[62]

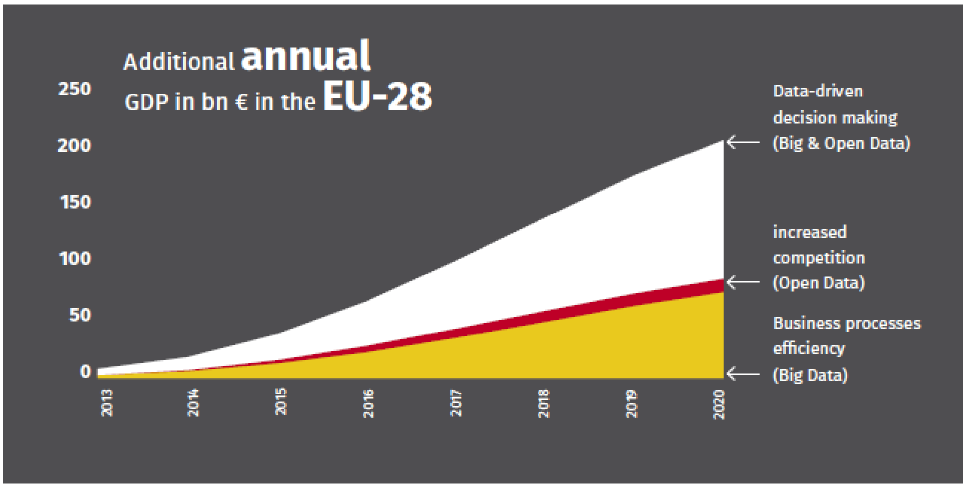


**Figure 43: Additional value from each macroeconomic channel (2013-2020) – In € billion**[63]

The allocation of production factors and their economic impact are expected to produce annual benefits exceeding **€**123 billion by 2020. It is, however, important to remember that this figure may only be realized if real value can be extracted from data.

The Buchholtz (et al.) study predicted that data-driven decision-making applications alone will account for half of all the gains produced by productivity growth and contribute to a 20% increase in a company's performance. Data-to-Management occurs at the intersection of tangible and intangible parts of the economy and this is where TDM offers the greatest potential for expansion as data may be transformed into actionable intelligence. In this sense, TDM becomes a vital activity for companies in their decision-making processes. More and more companies identify the necessity of making

[62] Buchholtz et al., 2014.
[63] Buchholtz et al., 2014.

investments in the most analytical aspects of working with data. Notwithstanding the large initial investments in network storage and infrastructure, the costs associated with developing these technologies tend to diminish significantly once they are incurred. For this reason, one may assume that companies that have developed their database and network infrastructure may subsequently have more latitude to concentrate investments in mining and analytical activities.

Another important conclusion of the study is that challenges in Big Data go beyond its technological aspect. In reality, European companies should mobilize substantial efforts to make organizational culture work in parallel with the pace of technological adoption. In other words, this means that the extraction of productivity gains from data-driven decision-making applications depends largely on the capacity of employees to adapt to new technologies and develop new organizational processes. What can be inferred from this remark is that productivity gains from evidence-based decision-making applications may produce better results if combined with honing human skills. It is important to add that larger investments in mining and analytical activities may allow companies to expand in-house capacities, including the hiring of new talents. As a result, companies may become less dependent on external service providers and this may reduce their cost base.

| **Data-to-Product/ Process** | **Data-to-Management** | **Data-to-Information** |
|---|---|---|
| >> Insights from data need to be implemented in the physical world to bring value to an enterprise. | >> Data-based information becomes part of a company's decision-making process | >> Efficiency improvements and innovations occur entirely in the intangible part of the economy. |
| >> The most straightforward way for industry and physical stores to increase their efficiency or introduce valuable innovations. | >> The most challenging and rewarding way of bringing value to companies | >> Data mined for valuable information |
| | >> Open throughout the economy | >> Companies with information as their core business |

TANGIBLE ← → INTANGIBLE

**Figure 44: Forms of Value Extraction from Data**

The second largest macroeconomic force is the productivity increase from improved business efficiency and it will be responsible for more than €72 billion of the overall incremental GDP by 2020. Since a lack of information often leads to squandering resources in production, distribution and marketing activities, Data-to-Process solutions, such as real-time sensor data analysis and advanced supply chain management may help European companies slash operating expenses. One example is the use of internal and external databases to increase accuracy in the detection of fraud. In these cases, TDM solutions may be valuable for companies in their efforts to solve common business problems. More than that, companies may derive competitive advantage through actionable intelligence as it may allow them to optimize resource inputs in the value chain and also introduce valuable innovations.

However, it is important to remember that large companies may easily and more rapidly exploit opportunities arising from Data-to-Product and Data-to-Process applications in the next decade. Larger companies have fewer constraints concerning capital and skills than small and mid-sized enterprises. For this reason, they are more capable of making the extensive investments ICT requires in intangible assets, such as strategy, skills and organizational structure. Moreover, these applications may assist established companies to segment markets and define or reshape products and services.

When it comes to competition improvements, the last macroeconomic channel, the expected economic impact of fostered competition with the opening of public sector data has been estimated to exceed €10 billion by 2020. Although these annual benefits seem far smaller than the productivity gains originating from company's proprietary Big Data, they are no less important. The reduction of information gaps using Open Data may provide companies with a substantial competitive edge if they are able to adapt to new technological tools. Moreover, it may spark new businesses to emerge and have a disruptive effect in different sectors. It is reasonable to mention at this point, as an example, opportunities in healthcare. Through data-driven solutions, companies have been able to improve health systems by comparing symptoms, causes, treatments and even patients' distance to hospitals to drive therapeutic strategy. In many cases, these improvements are offered by new companies with ample technical skills and data structures to create and maintain services. TDM has a wide range of possibilities to wield a real impact and extract value from Data-to-Process/Product and Data-to-Management solutions, from diagnosis by doctors in making their clinical decisions to the discovery of the most effective forms of treatment. Thus, to conclude this example, companies may cut costs and medical errors, thereby boosting patient satisfaction and, consequently, business reputation.

#### 1.1.5. Sectoral Impact

As stated above, the economic potential of Big and Open Data in terms of the incremental GDP to be produced in the next years will also vary substantially by sector. About half of the incremental GDP up to 2020 will be due to data-driven applications in trade, manufacturing and logistics. Together, finance, public administration and ICT will be responsible for almost 30% of the economic impact of Big and Open Data. Other sectors, including construction and health and social care, will account for 24%.

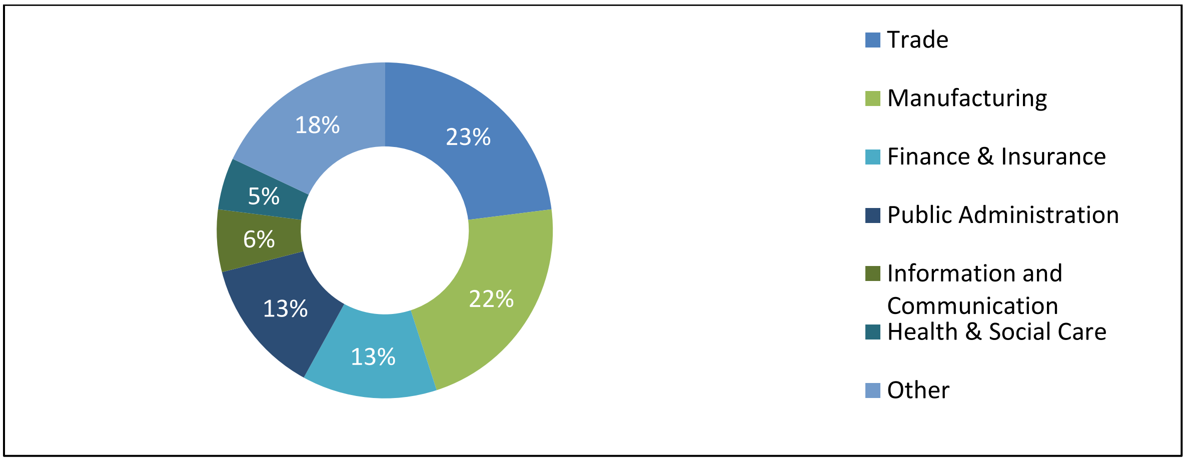


**Figure 45: Economic Impact of Big and Open Data by Sector (%) by 2020**[64]

**Box 11. Economic Impact by Sector Explained**

Buchholtz (et al.) study states that differences in the impact of sectors are explained by their relative size, technological characteristics and distinct economic features. Additional remarks on potential explanations for impact distribution of big data across sectors in Brown, Chui & Manyika (2011):

*"(...) while it (Big Data) will be important in every sector and function, some industries will realize benefits sooner because they are more ready to capitalize on data or have strong market incentives to do so."*

To analyze the dynamics behind Big Data adoption in 20 sectors in the US economy, Brown (et al.) elucidate their idea behind market incentives and readiness to capitalize on data through the use of two indicators: (1) a sector's potential for value creation through Big Data and (2) the ease of capturing Big Data's value.

The first indicator captures market incentives by taking into account a sector's competitive conditions, such as market turbulence and performance variability; structural factors, such as transaction intensity and the number of potential customers and business partners; and the quantity of data available.

The readiness-related indicator (encompassing TDM) considers the number of employees with deep analytical talent in an industry, the baseline IT investment, the accessibility of data sources and the degree to which managers make data-driven decisions.

*Source: McKinsey, "Are you ready for the era of big data?". Retrieved from www on 26th August 2016, http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data#0*

In this respect, it is important to note that the sectors expected to have the greatest economic impact, namely trade and manufacturing, are also the sectors with a greater potential of materializing data insights. Besides the fact that these sectors are responsible for the largest contribution to European GDP, this impact is also related to the readiness to capture value through big data mainly incentivized by e-commerce companies. In recent years, the rising economic impact of e-commerce – and the adaptability and efficiency of this model as a result of customer intelligence solutions – has forced traditional companies to adapt and embrace big data flows (Buchholtz et al., 2014). For these reasons one may assume that, at least in the short run, companies in these sectors may be the most likely to

[64] Buchholtz et al., 2014.

benefit from data solutions to improve operational efficiency. An essential economic feature of trade and manufacturing, i.e. gains from economies of scale, can be highly transformed by TDM solutions as they may produce positive repercussions across the entire value chain. For example, the integration of data streams on trade from multiple companies may eradicate the sector's current market fragmentation and enable new collaboration and services. As a consequence, companies may cut costs and grow their businesses quickly. Apart from operational efficiency, TDM offers the potential to transform customer experience by enabling companies to engage in more precise customer segmentation and targeting. For this reason, we conclude that TDM's potential in these sectors is still to be realized.

Greater data-driven transformation and value creation opportunities are also expected in finance and insurance. In this field, mining and analytics may be used to detect fraud, enhance retail customer service and improve operational efficiency. In fact, TDM may allow financial institutions to reduce risk exposure across different financial instruments as a result of lower information gaps and improved decision-making processes. Data solutions may be used to tackle one of the main challenges the financial sector faces, namely operational "silos". Still, many financial institutions, for example, fail to share data among their lines of business, including lending and money management. In this respect, TDM may be an essential activity for financial institutions to provide customers with more personalized services and get a more precise understanding of financial market operations.

The ICT contribution to the overall economic impact of Big and Open Data is smaller in comparison with other sectors, such as finance, manufacturing and trade, but there is an explanation for this fact. This smaller result comes from the fact that the available statistics present the macroeconomic effects of Big and Open Data applications in specific sectors and not necessarily the value-added growth within those sectors in which ICT innovations are part of the core business of companies. In this sense, the results presented for the ICT sector stands for the gains that the sector will have for investing in innovative data applications in other sectors. Moreover, it is also related to ICT's capacity to absorb some of the higher value generated by data-driven productivity expansion in those sectors. As an example, this report cites cases in which Big Data analytics are implemented in retail companies to improve performance in their supply chain management. Companies must incur expensive investments to put in place ICT infrastructure and software, but at the macroeconomic level this cost is translated into the incremental value generated by the ICT sector and the overall economic impact of larger productivity facilitated by Big and Open Data. Nevertheless, the study underlines that the extent of this economic effect depends largely on the level of monopolization of big data solutions on the providers' side.

With respect to value transfer between sectors, the study suggests that the competitiveness between local Big and Open Data solution providers, i.e. data owners, actors in global ICT supply chain and analytic hubs, exert a great impact on the final macroeconomic outcome. Because the positive impact of innovative ICT solutions is greater for net exporters than net importers, the EU should invest in developing competitiveness in the data analytics industry.

Indeed, and as it will be further explained, countries are expected to benefit differently from Big and Open Data developments. Noticeably, the study's remarks on this aspect offer valuable insights into how competition problems should be addressed to ensure fair and equal conditions for businesses in the EU.

**Country Groups**

As the results of the BOUDICA modeling suggests, the contribution made by countries to the EU's incremental growth is expected to vary in composition and economic terms.

Although countries were analyzed individually, the study presents its findings in country groups. Northern Europe, except France, presents a stronger than average incremental GDP and may benefit mostly from Big and Open Data developments from having highly developed ICT and more global companies. In contrast, most of the economies in Southern Europe and the New Member States should benefit less. However, it is important to note expressive exceptions in the country group comprised by New Member States. Poland should gain from the increased productivity of large companies in the retail trade and wholesale sector as both of them have a high share in its economy and the Czech Republic. This is expected to improve its manufacturing sector by integrating it even more with international supply chains.

In general, countries may rely on different sectors to benefit from data-based innovations. In spite of the fact that some countries find more potential in one sector or another, they may also benefit from opportunities across different sectors.

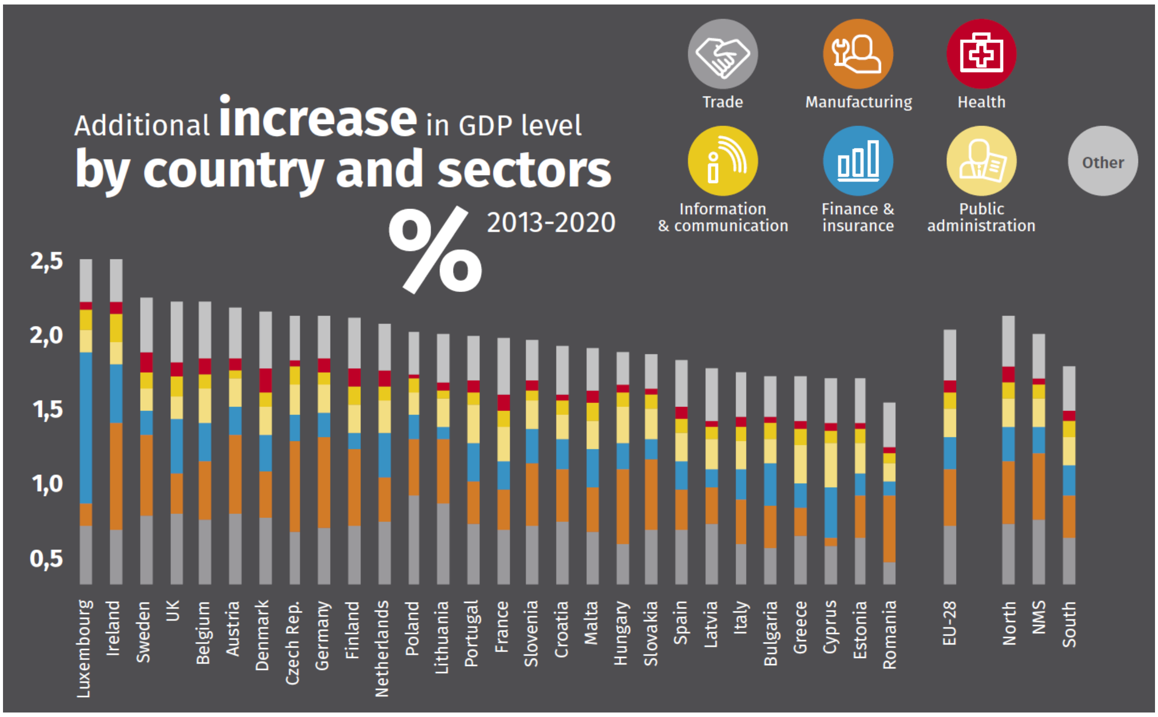


**Figure 46: Incremental GDP growth by country and sector (%) by 2020[65]**

The substantial gains countries may reap from Big and Open Data in specific sectors is worthy of noting. In Sweden, manufacturing will highly depend on data-driven solutions, while in Luxembourg, Ireland, UK and Cyprus, the highest data-driven growth potential comes from the financial sector. Another important remark from the study is that the broad macroeconomic effects exerted by Big and Open Data, i.e. the overall benefits to be delivered to the European economy will mainly depend on the broad ICT adoption rate, company size and structure.

[65] Buchholtz et al., 2014.

**Box 12. Differences in impact by country**

Outcomes will vary substantially in individual sectors and in countries where such data applications commence implementation. These outcomes mainly depend on company size, the country's economic structure and ICT penetration. The study underlines that countries whose industries are predominantly comprised of small and medium businesses and that lag behind in ICT penetration will face more obstacles in extracting economic benefits from Big and Open Data. In those countries, it is expected that the overall positive economic impact will be smaller due to lower adoption rates of data-driven applications and management practices.

Because countries seem to find larger potential for extracting value in different sectors, one can assume that the opportunities and challenges to be addressed by countries are also of a different nature. The structured economies of Northern Europe, considering their sector composition and average company size make it favorable to deploy ICT data-driven innovations. In fact, this country group seems to be more apt to extract value from data in the short run as the essential technological conditions are already in place for TDM to be further explored. About 60% of incremental GDP will come from data-driven decision-making applications. It is worth noting that this country group should be responsible for the largest gains in the finance & insurance and health sectors. However, the biggest challenge for those countries is to remain globally competitive, which includes securing talent and reshaping their industries through data-driven solutions.

In the New Member States, the greatest opportunities reside in manufacturing and trade. Nevertheless, countries should focus on diminishing their innovation lag, enacting effective open data policies and conducting large investments in knowledge-intensive services and R&D to realize the potential of Big and Open Data applications. Although these countries should expect extensive gains through business efficiency, they tend to be smaller than in Northern Europe.

The last country group, Southern Europe, seems to be an exception concerning sectoral impact. The economic contribution made by this country group to incremental GDP appears to be evenly distributed across different sectors. This country group may identify growth opportunities in improving public sector efficiency and fostering economic restructuring. Although economic restructure may be hard in the short run, these countries can increase their potential returns of investments in Big Data by focusing on developing data-based managerial skills and establishing linkages between large Big Data providers and SMEs. However, these countries must overcome certain challenges to extract benefits from these opportunities, such as create conditions to achieve economies of scale, secure financing for innovative data solutions and close the ICT gap.

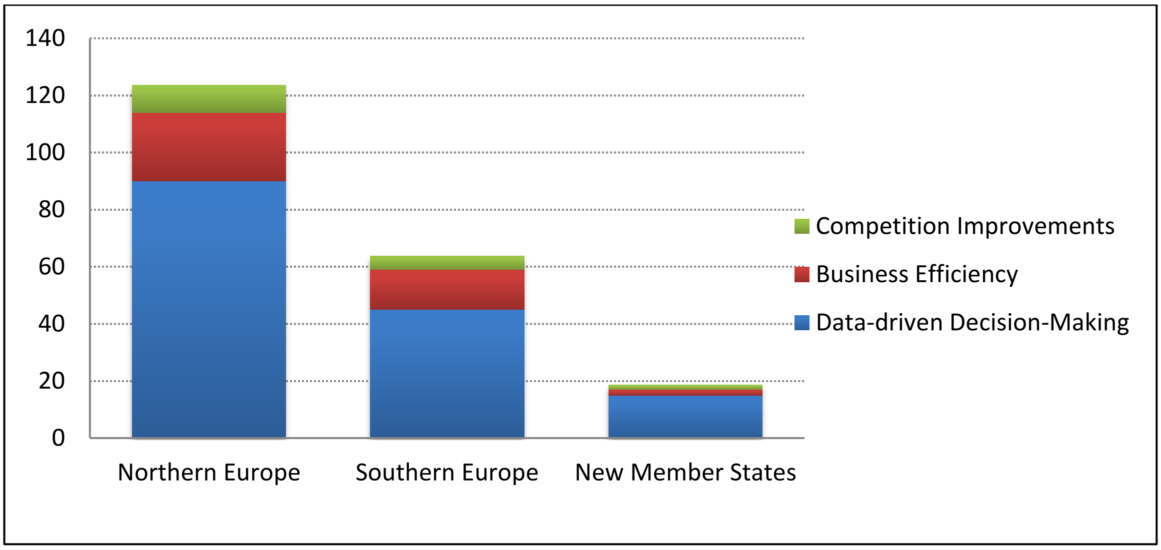


**Figure 47: Incremental GDP growth by country group by 2020 (in € billion)[66]**

Regarding the economic impact contributed by each country group, the largest contribution will definitely be seen In Northern Europe. About 60% of incremental GDP produced by this country group will come mainly from data-driven decision-making applications enabled by Big and Open Data, followed by business process efficiency as a result of Big Data applications with more than 30% and the remaining part due to increased competition evinced by Open Data. Although the impact of ICT solutions should not be realized in its full potential as in Northern Europe, the incremental GDP from data solutions in the New Member States will be quite extensive. About 60% of the additional growth in this region will result from data-driven decision solutions and 35% will be due to business process efficiency. In the southern region, the level of competition improvement from Open Data will be larger than in the New Member States. However, the incremental GDP due to business efficiency and data-driven decision making solutions from 2013-2020 will not be as large as the percentages produced by the New Member States.

We may conclude that the level of readiness to overcome challenges and take advantage of the opportunities fostered by Big and Open Data applications are not spread evenly across Europe. Although some similar challenges may transpire in all the countries, such as the capacity of companies to adapt to new technologies and extract value from data, TDM cannot be tackled through a one-size-fits-all policy during ICT development, especially in terms of infrastructure, as the economic strengths and weaknesses of countries are not symmetric.

In this sense, it is evident that country-specific policies to enable the expansion of TDM processes and applications are indispensable. The distinct economic features and levels of technological adoption across countries point to the fact that TDM has greater potential to expand in Northern Europe than in Southern Europe or in the New Member States. In the latter two country groups, it is essential that policies address ICT gaps while recognizing the necessity to make larger investments to overcome the innovation lag.

Although the authors of BOUDICA emphasize the capacity of large companies to invest in people and

[66] Buchholtz et al., 2014.

infrastructure, which are preconditions for TDM, SMEs may also become more and more capable of deploying data-driven innovations as a result of cloud computing expansion. Cloud computing seems to be a promising catalyzer for SMEs because it significantly diminishes the costs related to data storage and infrastructure capacities. Thus, one can assume that cloud computing may not only enable new business to emerge but also expand the opportunities companies have to engage in data mining and analytics.

Notwithstanding the fact that market dynamics in Europe are, to some extent, captured through the impact parameters used in the model, a thorough analysis on the exchanges between large and small companies should be explored further. This would not only allow for ways to diminish data fragmentation across sectors but also lead to more detailed market trends, including the disruptive effects of SMEs in traditional sectors as a result of access to data-driven decision making applications after those companies kick off their market operations. Furthermore, the analysis on market dynamics could also contribute to the identification of potential market distortions and ways to cope with them in order to guarantee fair competition conditions.

Technically, it is important to add that data accessibility and data quality are key issues. Regardless of country or company size, both internal and external data sources should offer a capacity of producing feedback loops fast enough so as to guarantee reliable and up-to-date data. This is especially important with regards to public administration data where the adoption of data solutions seems to occur at a slower pace vis-à-vis private sector developments.

For the abovementioned reasons, it is possible to conclude that policy recommendations on Big Data development and more specifically on data mining and analytics practices, should be country specific. However, it must be said that some challenges are visible in all these countries, including fragmented data sources and uses. Because fragmentation creates silos that prevent the extraction of actionable intelligence and also limits the capacity of companies to attain economies of scale and generate ample returns on their investments, the coordination of a common digital framework and strategy at the European level is essential to spark the expansion of data mining and analytics. It must allow for fair competition to prevent suppliers and vendors from colluding to segment pricing and enable innovation through financial support for startups, tax breaks for big data services and specific programs intended to boost digital skills.

In this sense, a clear strategy on Big and Open Data envisioning the potential of data mining and analytics is of paramount significance for European countries to seize opportunities and tackle its current socioeconomic challenges. That being said, the development of a Big Data ecosystem capable of generating positive economic repercussions can only be sustained using policies that simultaneously address legal, business, technological and social dimensions.

## 4.5 Summary on TDM

Based on previous macroeconomic studies, the BOUDICA model mirrors the impact of the latest ICT developments in the main sectors of the European economy. Big and Open Data are recognized as impact sources as well as the environment in which TDM has its role to play as a catalyst of actionable intelligence. Data value extraction through TDM may exert a significant impact on the economy as a whole, depending on the capacity of countries, sectors and more specifically, companies to deploy ICT innovations and create internal processes to allow for broader business efficiency, decision-making improvements and entry cost cutting. Thus, policy recommendations on TDM should be country specific and cover human capital, organizational practices and market dynamics. Dialogue with the various economic structures in the EU is also advisable. On this last aspect, it is important to add, however, the need of devising a robust strategy at the European level in order to highlight the necessity of developing TDM as a process to safeguard the extraction of value from data.

# 5 LITERATURE

Bean, Charles. 2016. *Independent Review of UK Economic Statistics.*

Barel, Yannick. *Big Data. Big Opportunity*, IBM: Power Systems.

Bradshaw, David, Gabriella Cattaneo, Rosanna Lifonti, John Simcox. 2014. *Uptake of Cloud in Europe. Follow-up of IDC Study on Quantitative estimates of the demand for Cloud Computing in Europe and the likely barriers to take-up*, Framingham: IDC Analyze the Future.

Brown, Brad, Jacques Bughin, Angela Hung Byers, Michael Chui, Richard Dobbs, James Manyika, Charles Roxburgh. 2011. *Big data: The next frontier for innovation, competition and productivity*, McKinsey Global Institute

Brynjolfsson Erik, Hitt Lorin M. and Kim Heekyung Hellen (2011), *How Does Data-Driven Decision-making Affect Firm Performance?*

Buchholtz, Sonia, Maciej Bukowski, Aleksander Śniegocki. 2014. *Big and open data in Europe. A growth engine or a missed opportunity?*, Warszawa: demosEUROPA, Warsaw Institute for Economic Studies.

Buytendijk, Frank. 2014. *Hype Cycle for Big Data*, Stamford: Gartner.

Capgemini and Informatica. 2016. *The Big Data Payoff: Turning Big Data into Business Value.*

Cattaneo, Gabriella, Mike Glennon et al. 2015. *European Data Market SMART 2013/0063, D6 – First Interim Report*, Framingham: IDC Analyze the Future.

Cebr. 2012. *Data Equity. Unlocking the value of big data,* London: Centre for Economics and Business Research Ltd.

Danson, Forrest et al. (red.). 2016. *Analytics Trends 2016. The Next Evolution*, New York: Deloitte.

EMC Corporation. 2014. *Redefine. EMC Big Data: Become Data-Driven*, Warszawa: EMC Forum 2014.

Etro Federico (2009), *The Economic Impact of Cloud Computing on Business Creation, Employment and Output in Europe: An Application of the Endogenous Market Structures Approach to a GPT innovation*

European Commission, *Digital Market Factsheet (2015)*

European Parliament. 2016. *Towards a thriving data-driven economy*, Strasburg.

EY Building a better working world, Europe Unlimited, IDC Analyze the Future, tech4i2. 2013. *Business Opportunities: Big Data.*

Fayyad, Usma, Gregory Piatetsky-Shapiro and Padhraic Smyth. 1994. *From Data Mining to Knowledge Discovery in Databases*, „AI Magazine", nr 17/3: 37-54.

Forbes Insights. 2015. *Betting on Big Data. How the right culture, strategy and investments can help you leapfrog the competition.* Washington.

Fujitsu. 2014. *Fujitsu Market Insights. Big Data*, Fujitsu Technology Solutions GmbH.

Georghiou, Luke. 2015. *Value of Research. Policy Paper by the Research, Innovation and Science Policy Experts (RISE)*, Luxembourg: Publications Office of the European Union.

Girard, Greg. *From Big Data to Real-Time Business Value,* Framingham: IDC Analyze the Future.

Heudecker, Nick, Lisa Kart. 2014. *Survey Analysis: Big Data Investment Grows but Deployments Remain Scarce in 2014*, Stamford: Gartner.

Heudecker, Nick, Merv Adrian. 2015. *Survey Analysis: Hadoop Adoption Drivers and Challenges*, Stamford: Gartner.

Heudecker, Nick, Lisa Kart. 2015. *Survey Analysis: Practical Challenges Mount as Big Data Moves to Mainstream*, Stamford: Gartner.

Hexagon Mining, Level 3 Communications and Liquid Telecom, Maptek, Reflex SAP. 2015. *Mining Big Data*, London: Aspermont Media

Intel IT Center. 2012. *Peer Research. Big Data Analytics. Intel's IT Manager Survey on How Organizations Are Using Big Data*, United States.

Kart, Lisa, Nick Heudecker, Frank Buytendijk. 2013. *Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype*, Stamford: Gartner.

Kart, Lisa. 2015. *Big Data Industry Insights*, Stamford: Gartner.

Kelly, Jeff. 2015. *Bod Data Vendor Revenue and Market Forecast. 2011-2026,* Wikibon.

McKinsey (2011), *Are you ready for the era of big data?* Retrieved from www on 26th August 2016, [http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data#0](http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/are-you-ready-for-the-era-of-big-data#0)

Pietruszyński, Piotr. 2014. *Big Data+. Systemy analityki wielkich zbiorów danych w polskich organizacjach.* Computerworld.

Ransbotham, Sam, David Kiron, Pamela Kirk Prentice. 2015. *The talent dividend. Analytics talent is driving competitive advantage at data-oriented companies.* MIT Sloan Management Review.

SNS Research Market Intelligence & Consultancy Solutions. *The Big Data Market. Opportunities, Challenges, Strategies, Industry Verticals & Forecasts*.

SVB Analytics. 2015. *Big Data Next: Capturing the Promise of Big Data,* United States: Silicon Valley Bank.

Vesset, Dan et al. 2012. *Market Analysis. Worldwide Big Data Technology and Services 2012–2015 Forecast*, Framingham: IDC Analyze the Future.

Vesset, Dan et al. 2015. *IDC FutureScape: Worldwide Big Data and Analytics 2016 Predictions*, Framingham: IDC Analyze the Future.

Vasset, Dan, Suya Xiong. 2015. IDC MaturityScape Benchmark: Big Data and Analytics in the United States, Framingham: IDC Analyze the Future.

Watson Healy Leigh. 2016, *Information Industry Outlook 2017: All Data, Nothing But Data.* Outsell

Zwakman, Greg, Matt Aslett, Jason Stamper, Krishna Roy. 2015. *Total Data market expected to hit $115bn by* 2019. 451 Research LLC.