

Communication

Big Data Use and Challenges: Insights from Two Internet-Mediated Surveys

Elisa Rossi ¹, Cinzia Rubattino ¹ and Gianluigi Viscusi ^{2,*}

¹ GFT Italia S.r.l., 16129 Genoa, Italy; elisa.rossi@gft.com (E.R.); cinzia.rubattino@gft.com (C.R.)

² École Polytechnique Fédérale de Lausanne - College of Management of Technology—Chair of Corporate Strategy & Innovation (EPFL-CDM-CSI), 1015 Lausanne, Switzerland

* Correspondence: gianluigi.viscusi@epfl.ch

Received: 11 August 2019; Accepted: 19 September 2019; Published: 24 September 2019



Abstract: Big data and analytics have received great attention from practitioners and academics, nowadays representing a key resource for the renewed interest in artificial intelligence, especially for machine learning techniques. In this article we explore the use of big data and analytics by different types of organizations, from various countries and industries, including the ones with a limited size and capabilities compared to corporations or new ventures. In particular, we are interested in organizations where the exploitation of big data and analytics may have social value in terms of, e.g., public and personal safety. Hence, this article discusses the results of two multi-industry and multi-country surveys carried out on a sample of public and private organizations. The results show a low rate of utilization of the data collected due to, among other issues, privacy and security, as well as the lack of staff trained in data analysis. Also, the two surveys show a challenge to reach an appropriate level of effectiveness in the use of big data and analytics, due to the shortage of the right tools and, again, capabilities, often related to a low rate of digital transformation.

Keywords: big data; big data and analytics; big data technologies; big data use

1. Introduction

In this article we discuss the results of two multi-industry surveys carried out to understand the needs, requirements, and use of big data and analytics by public and private organizations [1–3]. In particular, we are interested in companies whose use of big data and analytics may have also a social value in terms of, e.g., public and personal safety, along the big data information value chain [4]. It is worth noting that the results of the first surveys have been presented by the authors in [5]—whereas in this paper they are further complemented by the insights from a second survey, eventually reinforcing some of their suggestions for practice. However, the purpose of this paper is mainly exploratory and descriptive, aiming to support further reflections on the use of big data and analytics in organization, thus providing insights that can be eventually developed in further research.

As to these issues, the two surveys were conducted for the purpose of understanding market needs and requirements for the platform (<https://platform.aegis-bigdata.eu/>) eventually developed by the AEGIS (Advanced Big Data Value Chain for Public Safety and Personal Security) project consortium. The AEGIS project was funded as European Commission H2020 Innovation Action, aiming at “creating an interlinked ‘Public Safety and Personal Security’ Data Value Chain, and at delivering a novel platform for big data curation, integration, analysis and intelligence sharing.” As for the research presented in this paper it is worth mentioning that the AEGIS big data value chain includes the following steps, as presented in [6]: data acquisition (“the process of gathering, filtering and cleaning data, before any data analysis can be carried out”), data analysis (“concerned with making the raw data acquired amenable to use in decision-making as well as domain-specific usage”), data curation (“the active

management of data over its life cycle to ensure it meets the necessary data quality requirements for its effective usage”), data storage (“the persistence and management of data in a scalable way that satisfies the needs of applications”), and data usage (“‘data-driven’ business activities that need access to data, its analysis, and the tools needed to integrate the data analysis within the business activity”) [6].

The paper is structured as follows. First, we provide a summary of the related work on big data and analytics, providing a motivational rather than theoretical background to the mainly industrial research presented in this paper. Then, the main results of the two surveys are outlined and discussed, before conclusive remarks end the paper.

2. Related Work

Considering the growing literature on big data [4,7–11], we can nonetheless still take, as a proxy for the common understanding of big data, the following definition which appeared in 2013 in the first issue of *Big Data*, one of the first journals on the topic published by Mary Ann Liebert, Inc: “Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn’t fit the structures of your database architectures. To gain value from this data, you must choose an alternative way to process it” [1]. Since then, one of the main barriers for laypersons and businesses has remained an understanding of the capacity [12] of current technological infrastructure, business process, and business models to maintain, produce, and use big data [13], especially considering not only their economic value, but also their potential public and social value [14–18]. These issues are actually related to most of the steps in the above mentioned big data value chain (data analysis, data storage, and data usage). Besides them, it is worth mentioning the need for dedicated approaches and methods to maintain big data quality [17], specifically concerning the steps of data acquisition and data curation of the value chain. However, these challenges are also strictly connected to the fact that, nowadays, the first asset of digitalization are data [12,19] and they are the source for the generativity related to the resulting information [20,21].

Considering now healthcare as an example of a key domain for public as well as personal safety and security, the information needs, characteristics and factors having a potential impact on the use of big data analytics have been extensively studied in the public health and communication literature as well as in the information systems (IS) research domain. Xiao et al. [22], for example, identified a set of factors (individual characteristics and health factors, situational factors, information carrier factors, channel related factors, and perceived information gathering capacity) for cognitive and affective information seeking needs. The authors have also noted how the IS scholars’ contributions have paid limited attention to the users’ online health information search behaviors (p. 418) compared to the interest the phenomenon raised when considering marketing and consumer-related domains [23–25]. Although identified and discussed for a specific domain, those factors are worth considering and investigating for their potential application to other areas, especially within the spectrum of the ones referring to public as well as personal safety and security.

Taking the above issues into account, big data and analytics are worth considering as transformational information technology [11,26,27], altering the way of doing business as well as the required capabilities [28]. Thus, these changes require an understanding of how organizations may use those technologies as well as which factors, challenges, and finally benefits their managers and staff state as actually relevant for their activities. This understanding is the motivation for the empirical work we have carried out during the AEGIS project, whose exploratory results we are going to discuss in the following sections.

3. Results

The exploratory research presented in this paper has been carried out through two internet-mediated surveys [29–31]. In what follows, we are going to detail their specific characteristics and present the main results. It is worth noting that we will focus more on the second survey,

providing a summary of first one, being this latter already presented in a former paper [5], to whom we refer the interested reader for further details.

3.1. Survey 1

As discussed in [5], the survey was conducted online on a sample of industries and sectors including, among others, finance and insurance, automotive, information technology (IT), healthcare, research and education (see further details in Table 1). The different groups were selected for their use of big data for creating and capturing not only economic value, but also social value in terms of public safety and security for the final users or customers [5]. Then, the survey was carried out from January to March 2017, using multiple channels such as an online questionnaire form (<https://www.aegis-bigdata.eu/what-is-the-current-and-expected-use-of-big-data-technologies-a-glimpse-to-our-aegis-questionnaire-results/>), face to face, and telephone interviews. In the case of those interviews, the sample was made up of companies from the IT industry (for the big data technological infrastructure [32,33]) and finance (for the degree of information content of their product and information intensity of the value chain [34]), being an example of data-driven industry [35]. Ultimately, we received 77 replies to the questionnaire out of 110 invitations (a response rate of 70%). In what follows we discuss the main results from the survey and the interviews.

Table 1. Distribution of respondents per industry/sector (multiple answers allowed for the field).

Industry/Sector	Respondents
Manufacturing	5
Telcos	5
Transport	5
Academia	5
Automotive	12
Entertainment	1
Retail	1
Internet and social media	18
Computer Software	31
IT services	37
Defense	2
Oil and energy	3
Marketing advertising	3
Public sector	3
Smart Home	9
Health care/Hospital	4
Research	19
Insurance	6
Financial Services	6
<i>Total</i>	175

Most of the respondents came from IT or related industries, as shown by Table 1. As to the geographical distribution, as shown by Table 2, all the countries of the partners of the project were covered with additional replies from Portugal, France, Belgium, Bulgaria, Luxembourg, the Netherlands, United Kingdom (UK), Spain and, outside Europe, Mexico, Argentina, United States (US). Furthermore, there was also a regular distribution of respondents from small and medium-sized enterprises (~75%)

and large entities with more than 1000 employees (~25%) [5]. Generally, while 55.3% of respondents already had a strategy for using big data and analytics, only 34.2% were effectively using them, 35.5% were starting their use, 13.2% were on a planning phase, and only 17.1% had no experience [5]. Concerning the data sources, the most cited sources were logs (~45%), transactions (~32%), events (40%), sensors (32%), and open data (30%). It is worth noting that open data had the higher rate of willingness for exploitation in the next five years together with social media and free-form text [5]. Moreover, the sample has shown little interest in data coming from phone usage, reports to authorities, radio-frequency identification (RFIDs) scans or point of sale (POS), and geospatial data. In general, although ~72.6% of data sources were multilingual, only ~50% of the sample declared to have the needed tools to handle different languages. As for data sources considered as relevant although not yet fully exploited, the main obstacles for their use were related to security, privacy and legal issues, availability and discoverability of data, lack of a common data model and lack of the necessary skills or strategy within the organization. Actually, most of respondents (40%) stated that less than 10% of data collected is further processed for activities connected to value creation and capture, although they also foresaw an increase in the next five years [5].

Table 2. Distribution of respondents per country of provenance of the organization.

Country	Respondents
Greece	25
Austria	12
Italy	11
Germany	10
UK	5
Spain	4
Argentina	1
Belgium	1
Bulgaria	1
Cyprus	1
France	1
Luxembourg	1
Mexico	1
The Netherlands	1
Portugal	1
US	1
<i>Total</i>	<i>77</i>

Thus, the limited exploitation of big data seems associated to a low degree of information capacity of the considered organizations [19] and a gap in analytics capabilities [36,37]. These weaknesses could be also reflected by the fact that more than 60% of respondents had in-house both data collection and data analytics, while only a few were outsourced. In general, it seems that those organizations require an IT transformation rather than a simple reconfiguration or renew of the IT portfolio [5]. As to this issue, the main technologies in use for big data analytics among the respondents were Apache Hadoop (21%) and Microsoft Power BI (17%). Finally, only 36.5% of respondents declared that they share data with other subjects [5].

3.2. Survey 2

In this section, we discuss the results of the second survey carried out during the AEGIS project to collect further requirements from the potential stakeholders [6]. The questionnaire was submitted between February and March 2018 to all the different sample groups of the first survey, although targeting some specific roles for the participants in their organizations. According to [6], the roles identified were:

- **Manager:** “a person responsible for controlling or administering an organization or group of staff, he/she has a high-level point of view about big data analytics but is the person that could benefit from them. He/she has a focus on business intelligence” (6, p. 18).
- **IT Technical Operator:** “a person responsible for the management of the data storage, curation and collection, he/she knows which could be the critical points of these tasks” (6, p. 18).
- **Data Scientist:** a “person that extracts information from data, using big data analytic tools, for instance following the instructions of the manager. He/she has the proper skills for data analysis and could identify the deficiencies of the existent tools” (6, p. 18).

Table 3 shows for each role the main topics/features investigated through the survey.

Table 3. Overview of the main topics investigated through the survey for each of the defined roles, adapted from [6].

Role	Main Features
Manager	<ul style="list-style-type: none"> • Effort and resources involved in data analytics • Type of data involved in the analysis • Data treatment agreement • Sharing of the analysis
IT Technical Operator	<ul style="list-style-type: none"> • Data collection • Data sources • Hardware • Sharing of the analysis
Data Scientist	<ul style="list-style-type: none"> • Type of data involved in the analysis • Data sources • Analytic tools and algorithms • Sharing of the analysis

An online version of the questionnaire (powered by Easy Feedback) has been provided and it is still Available online: the following link: <https://indivsurvey.com/aegis/117873/8il3tU> (still online: Accessed on 17 September 2019). Moreover, each AEGIS’s partner sent direct email invitations to people in personal networks or on LinkedIn groups and Facebook and we eventually received 37 valuable replies to the questionnaire out of 56 submissions (see Table 4 for the type of organizations of the respondents).

Table 4. Role of the respondents per type of organization (for businesses the size is reported).

Types of Organization					
Role	Academia	Research Center	Non-Profit Organization	Business	Total
Managers	1	5	0	12	18
IT Technical Operators	0	3	1	2	6
Data Scientists	1	6	0	6	13
Total	2	14	1	20	37
Size of Business					
Role	Large Enterprises	Medium Enterprises (<250 Employees)	Small Enterprises (<50 Employees)	Micro Enterprises (<10 Employees)	Total
Managers	5	2	1	4	12
IT Technical Operators	2	0	0	0	2
Data Scientists	3	1	2	0	6
Total	10	3	3	4	20

As shown by Table 5 it is worth noting that 14 out of 37 respondents were from the Information Technology (IT) industry or related sectors (such as “Information Management”, “Statistics and Information Systems” or, generally, information and communication technology—“ICT”). As for the size of businesses, as shown by Table 4, large enterprises and small-medium-micro enterprises were almost equally represented—although it is worth noting that the latter forms of enterprise have to be considered as a single cluster for this survey—otherwise we have an average of ~four enterprises for those that were not large enterprises. As shown by Table 6, considering the country of origin of the organizations of the respondents, the majority of the replies came from Austria (~31%), Greece (~17%), and Italy (~17%), followed by Spain (~13%) (it is worth noting that ~21% of respondents did not mention the country of their organization).

Table 5. Industry/Sector of the respondents—in this case not all of them provide the information.

Industry/Sector										
	Education	Insurance/Finance	Business (Generic)	IT	Statistics and Information Systems	Information Management	ICT	Food, HR, Healthcare, Aeronautical	Automotive	Total
Managers	1	1	1	4	0	1	2	1	1	12
IT Technical Operators	0	0	0	1	0	1	1	0	2	5
Data Scientists	1	1	0	2	2	0	0	0	6	12
Total	2	2	1	7	2	2	3	1	9	29

Table 6. Country of the respondents—in this case not all of them provide the information.

Country										
Role	Austria	France	Germany	Greece	India	Italy	Spain	Uganda	US	Total
Managers	2	1	1	2	1	3	2	1	2	15
IT Technical Operators	2	0	0	1	0	1	1	0	0	5
Data Scientists	5	0	0	2	0	1	1	0	0	9
Total	9	1	1	5	1	5	4	1	2	29

Considering now the use of big data (see Table 7), the 60% of the respondents for the organizations that participated to the survey have declared that they were effectively using big data. It is also worth

noting that here we define ‘big data effectiveness’ as “the capacity to elaborate big data and use them to create value” on the basis of the discussion of the results of the first survey [5] (see also the concluding remarks in this paper).

Table 7. Level of Experience of the organization with big data, adapted from [6].

Level of Experience	%
Effectively using big data	60
Beginning in the use of big data	16
Planning to use big data	14
No experience	10

As said, the survey included the three different types of participants presented above and shown with their related features in Table 3—thus, the answers came from managers (49%), data scientists (35%) and IT technical operators (16%).

Considering managers, there were eighteen replies (see Table 4)—four of them came from small, medium, and micro organizations planning to use big data, although not having a designed team (internal or external) to perform data analysis. Moreover, considering the organizations that declared to be beginners in the use big data, they preferred to perform analysis through external consultants or, in some case, an internal team, but not as a main activity. Instead, the organizations that were effectively using big data (the majority in the IT and automotive sectors) had an internal team of data scientists. Also, it is important to point out that 18% of the participants declared that even if they have a dedicated budget for big data and analytics, the investment was not adequate. Accordingly, only 2% of participants declared to have the proper hardware to manage big data [6].

Table 8. Added values of big data analysis from the perspective of managers and data scientists (only one IT technical operator replied to the question, thus we have decided to omit it from the analysis), adapted from [6].

Big Data Added Value (Areas)	Managers (<i>n</i> = 18)	Data Scientists (<i>n</i> = 13)
Predictive analysis	77%	71%
Cross-domain analysis	61%	57%
Real time analysis	38%	21%
Fast decision making	31%	43%
Improvement of the offered services	24%	43%
Cost reduction	31%	36%
Better customer service	24%	36%
More effective marketing	15%	43%
Competitive advantages over rivals	23%	21%

Moving now to the added value of big data and analytics, Table 8 reports the perspective of managers and data scientist, while Table 9 shows the main issues that the managers pointed out as key to the use of big data. Here, it is worth noting that contrary to the first survey the heterogeneity of available data was not considered an issue by managers and IT technical operators; the reason for this result could be related to the different degree of experience with big data exploitation required of the participants of the second survey.

Table 9. Big data issues from the perspective of managers and IT technical operators, adapted from [6].

Big Data Issues	Managers (<i>n</i> = 18)	IT Technical Operators (<i>n</i> = 6)
Difficulty of handling Big Data	44%	66%
High management cost	39%	33%
Difficulty of finding trained staff in Data Analysis	38%	33%
Legislation about privacy and security	27%	50%
Lack of performance of the available tools	16%	0%
Heterogeneity of available data	0%	0%
Lack of confidence in the real benefit	16%	16%

Considering specifically the AEGIS big data value chain, one of the questions regarded which of its steps were actually implemented in the respondents' organization: 62% of the respondents declared to carry out "Data Analysis", 48% "Data Acquisition", 43% "Data Storage", 37% was actually using the results of the analysis ("Data Usage"), and finally 18% performed "Data Curation". It is worth noting that the respondents that declared that they were 'effectively using big data' were also already implementing all the steps of the big data value chain identified in the AEGIS project.

As for the origin of the data involved in the analysis by the various organizations in their activities, Table 10 shows the main sources as identified by managers, data scientists, and IT technical operators that replied to the survey (question: 'Which are the data involved in the analysis of your organization?'). As to this issue, the respondents that used external or purchased data also declared to have an agreement with the providers, which included a reference to further processing of previously collected personal data. Furthermore, all the participants agreed on the importance of linking datasets from different domains/data sources for their analyses, although real time data were used by a limited number of participants (~15% for the three roles considered in the survey). Also, only the 10% of the overall respondents declared to use alerts, warnings or monitoring systems based on big data analytics as a support after an event, while 10% did not know, and 80% did not use that kind of automated feedback. Finally, as shown in Table 11, the sharing of data and the related analyses was mainly with the customers and with colleagues of the same team.

Table 10. Origins of the data according to managers, data scientists, and IT technical operators (multiple answers allowed), adapted from [6].

Origins of Data	Managers (<i>n</i> = 18)	Data Scientists (<i>n</i> = 13)	IT Technical Operators (<i>n</i> = 6)
External, customers (e.g., data from social media, sensors)	27%	53%	33%
External, open data	22%	38%	50%
Internal of the organization	22%	30%	50%
External, real-time data	16%	15%	16%
Internal, related to customers	16%	46%	57%
Purchased data	11%	7%	0%

Table 11. Data sharing target according to managers, data scientists, and IT technical operators (multiple answers allowed), adapted from [6].

Data Sharing Target	Managers (n = 18)	Data Scientists (n = 13)	IT Technical Operators (n = 6)
With customers	27%	46%	33%
With colleagues of the same office/department/team	33%	69%	66%
With colleagues of other offices of the same organization	33%	38%	16%
As open data	5.5	8%	33%
I don't share analysis	-	8%	-
With external, entities	5.5%	30%	33%

Focusing now on the data scientists that participated in the survey, six respondents out of thirteen declared that there were different restrictions about data visibility in their organization, while only two declared that there were no restrictions about data visibility (four respondents replied that they did not know about it). Then, considering again the results shown by Table 10, the data scientists pointed out that data processed came mainly from external sources related to customers (53%), or from internal data related to customer (e.g., contracts)—46%. The other categories (open data—38%; data internal of the organization—30%; real-time data—15%; and purchased data—7%) scored percentages considerably lower. Also, all of the participants of the survey asserted that they used to acquire data only when needed, and through scheduled streaming. Furthermore, the types of data mainly used for the analysis were logs and sensors data, while, contrary to the first survey, the data types not yet exploited but that the participants would have liked to use were: geospatial data, phone usage, email, transactions, social media, audio, radio-frequency identification (RFID) scans or point of sale (POS) data, and earth observation. Considering now the tools used, 46% of data scientists answered that they had proper analytic tools for their needs, the most popular tools being R, Matlab, and Python, while other tools mentioned were Pandas, Microsoft Excel, Spark and SAS Base. The algorithms actually adopted by the respondents or that they would have liked to adopt for the analysis are reported in Table 12, while the main output format identified for the results of the analyses were the tabular one (69%).

Table 12. Algorithms used or willing to use by the data scientists participating to the second survey, adapted from [6].

Algorithm	I Use	I Would Like to Use
Linear regression	61.5%	30.77%
Predictive analysis	46.1%	30.77%
Clustering algorithms	46.1%	38.46%
Simulations	46.1%	30.77%
Estimation of correlation between variables	38.4%	30.77%

Taking the above issues into account, only three out of thirteen data scientists declared that their organization have scheduled automated analysis of data, while six declared that their organization didn't perform scheduled automated analysis of data, and four did not know about it. The last question for each participant was aimed to understand the main features/functionalities for a potential big data and analytics platform. To this end, a set of features/functionalities were listed (see Tables 13 and 14) and the respondents could assign to each of them a level of interest ranging from 'Not at all' ("0") to 'Very' ("3").

Table 13. Features and functionalities for potential big data and analytics platforms: median for the replies from managers, data scientists and IT technical operators.

Features/Functionalities	Managers (Median)	Data Scientists (Median)	IT Technical Operator (Median)
Online and free	1.5	2	2
Where you can buy and sell assets	1.5	1	1
With a set of open assets	2	2	1
Where you can connect in-house streaming datasets	2	1	1.5
Where you can store your analysis and assets	2	1	2
Where you can manage the metadata related to your data	2.5	2	2
Where you can query your datasets and access a set of related visualizations	2	2	2
Where you can set and save the steps of your analysis	2	2	1.5
Where you can share the information with a selected group of users	1	2	2
Where you can set different restrictions about data visibility	2	1	1.5

As shown in Table 13, two features/functionalities with the highest median for the three categories of respondents are the ones related to metadata management, queries, and visualizations (cells in grey in Table 13). Here, it is worth noting the interest of managers in metadata, usually a more technical subject than a business oriented one, yet one reason for this result could be related to the already mentioned significant presence in the survey population of companies from the IT industry. Also, it is worth noting that the median value for managers' preference for a feature/functionality as "where you can manage the metadata related to your data" is 2.5, thus closer to 3 ("Very") than the median values for all the other features/functionalities, which are in a range between 1 ("Slightly") and 2 ("Moderately").

Those results are reflected also by the values shown in Table 14, where metadata management, queries, and visualizations received the highest percentage in the survey for features and functionalities that were considered as "very" interesting (grey cells in Table 14), especially by managers and IT technical operators. However, it is also worth remarking that 'being online and free' is one the key features that interested more data scientists in a platform for big data and analytics. This could be potentially interesting for further research, considering that, according to the results in Table 10, data scientists were less oriented or more cautious with regard to openness in data sharing; thus, they seem to be more oriented to openness when they need to access tools than when they have to share results with subjects external to their organization. Looking now at the features that were evaluated as not interesting at all (grey cells in Table 14 for "not at all" answers), a high percentage of managers and IT technical operators considered not worth having features for buying and selling assets, as well as the provision by an eventual platform of a set of open assets. On the contrary, data scientists were not interested in connecting in-house streaming datasets and being able to store analyses and data assets.

Table 14. Features and functionalities for potential big data and analytics platforms: percentage of replies for each question by preference per role: managers (M), data scientists (DS), and IT technical operators (IT).

	Roles	“Not at All” (0)%			“Slightly” (1)%			“Moderately” (2)%			Very (3)%		
		M	DS	IT	M	DS	IT	M	DS	IT	M	DS	IT
Features/ Functionalities	Online and free	33	39	17	17	0	17	17	15	33	33	46	33
	Where you can buy and sell assets	39	38.5	33	11	38.5	33	28	15	17	22	8	17
	With a set of open assets	28	38	33	6	0	50	33	54	17	33	8	0
	Where you can connect in-house streaming datasets	22	46	17	17	8	33	33	31	50	28	15	0
	Where you can store your analysis and assets	22	46	0	11	8	33	33	38	67	33	8	0
	Where you can manage the metadata related to your data	17	31	16.6	11	0	16.6	22	54	50.2	50	15	16.6
	Where you can query your datasets and access a set of related visualizations	22	31	17	22	8	33	17	46	0	39	15	50
	Where you can set and save the steps of your analysis	28	38	17	6	0	33	22	31	17	44	31	33
	Where you can share the information with a selected group of users	22	38.5	16.5	33	0	16.5	17	46	67	28	15.5	0
	Where you can set different restrictions about data visibility	28	38.5	17	17	15	33	28	38.5	17	28	8	33

4. Discussion

Considering the results of the two surveys presented in this paper, it is worth noting that the second iteration was aimed to understand the actual usage of big data and analytics by organizations from different countries and industries concerned by public or personal safety and security, with a particular focus on the steps of the AEGIS big data value chain [6]. Compared with the first questionnaire, the questions of the second survey were more specific, targeting people with experience with big data and analytics. For that reason, most of the participants of the first iteration has not been involved in the second iteration, as well as most of the questions has been changed. It is worth focusing now again on Table 8 that shows the benefits of data analysis from the point of view of the managers and the data scientists that participated to the survey. In general, we can observe that the replies that could be associated with a business view (i.e., cost reduction, fast decision making, more effective marketing and improvement of the offered services) received higher scores among the data scientists than among the managers. This discrepancy could be related to a greater confidence in the value of big data and analytics from their daily use by data scientists and the consequent degree of acquaintance with those technologies. Considering the data sources used for the analysis, both the surveys identified logs, sensors, and events as the relevant ones. However, the second survey revealed a greater interest for the future, among others, on geospatial, audio, earth observation and space, phone usage, social media, emails, and transaction data (see Table 15).

Table 15. Summary of key data types for data scientists that actually use and/or would like to use them, adapted from [6].

Data Type	I Use	I Would Like to Use
Log, sensors, events	40–60%	40–55%
Open data/public sector information, external feeds, free-form text, emails	10–25%	65–75%
Transactions, social media, phone usage, reports to authorities, Earth observation and space, audio	10%	85–90%

Due to the role-based survey developed for the second iteration, the results for the question “Does your organization have the right analytical tools to handle big data?” were fairly different: In the first iteration only the 24.3% of the participants declared to have the proper tools, while in the second the overall percentage reached 53.8%. The main big data analytic tools identified in the first survey were Apache Hadoop (21%) and Microsoft power BI (17%). In the second questionnaire, each of the respondents among data scientists has indicated more than one tool, the most popular being Python and R (50%), Panda and Matlab (33%). The features that mainly attracted the respondents were the possibility to have a tool to manage metadata, to set and save the steps of the analysis, as well as the fact that the tools would be online and free. Particular attention was also received by a potential new tool with a set of open assets, where it would be possible to connect in-house streaming datasets, store the analysis performed together with the related assets, allowing to query datasets as well as to access a set of related visualizations, likewise. The functionality that scored the lowest level of interest was the possibility to buy and sell assets—this lack of interest could be explained by the low percentage of purchased data (10%) and with the tendency to share the analysis within the organization or at least with customers.

Taking these issues into account, the two surveys and their results presented in this paper have limitations, related among other issues to the high percentage of the participants belonging to the IT industry and the number of respondents, which would require further sampling at country level for each industry among the ones considered as well as a more granular consideration of demographic data. However, the main goal of the article is to provide insights from two surveys that involved a sample of real organizations for questioning their current uses and needs as well as value perspectives on big data and analytics. Also, although they were related to the development goals of the AEGIS project, the two surveys originally had somewhat different goals, with the second one more oriented to requirements and functional needs elicitation for the three macro roles identified as relevant for the use and adoption of big data and analytics, especially in organizations whose activities were concerned with public or private safety and security. These limitations point out directions for future research where research questions and hypotheses will be tested on the basis of a theoretical framework that would eventually also consider the insights presented in this paper.

5. Conclusions

The paper presented the results from two multi-industry and multi-country surveys aimed to understand the needs, requirements, and actual use of big data and analytics by public and private organizations, with a specific focus on public safety, security, and social value. Already preliminary presented for the first survey in a previous article [5], those findings have been complemented by the results of a second survey carried out by the authors, early presented in a deliverable of the AEGIS project [6], and eventually further elaborated in this article. Some of them partially resonate with state of the art challenges in the management information systems literature, that we have formerly categorized [5] in terms of ‘big data efficiency’ (the capacity to simply elaborate big data), ‘big data effectiveness’ (the capacity to elaborate big data and use them to create value), and ‘big data accessibility’ of a given organization (the capacity to elaborate big data and use them to create and capture value from their sharing).

As for big data efficiency, the two surveys showed a challenge in the low utilization rate of the data collected, related to, among other issues, privacy, security, and the claimed ‘difficulty of handling

big data', especially due to the lack of trained staff in big data analysis as well as the various constraints raised by the current legislation on data management and access. As to data analysis capabilities and the information capacity of an organization [19], data heterogeneity (i.e., structured and unstructured data as well as the use of different languages) has been considered an issue mainly by the respondents from the first survey, while it was not relevant in the second one, where participants were selected for their higher degree of acquaintance with big data exploitation. Moreover, the two surveys showed a challenge to reach an appropriate level of big data effectiveness, due, again, to the lack of capabilities, where the available tools were considered difficult to use, especially in the case of multilingual data, also in this case, mainly by the respondents from the first survey. Finally, the two surveys showed a challenge for big data accessibility in the low rate of digital transformation of the company, which could improve, on the one hand, the data collection and analysis (still mainly in-house) and, on the other hand, the sharing of data with other external entities.

Notwithstanding the exploratory nature of research presented in this article, we believe that the results, although mainly descriptive, have at least the point of strength of highlighting some issues that would be worth investigating, especially for understanding the under exploitation of big data and analytics by companies that are not yet fully advanced in the digital transformation of their activities [38–41] due to their history, information systems legacy, or characteristics of the environment where they are based or actually operating.

Author Contributions: all the authors contributed equally in all the other sections of the paper.

Funding: This research was funded by the AEGIS project, which has received funding from the European Union's Horizon 2020 Framework Programme, grant number 732189.

Acknowledgments: This work was partially supported by the AEGIS project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 732189. The document reflects only the author's views and the Commission is not responsible for any use that may be made of information contained therein.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dumbill, E. Making Sense of Big Data (Editorial). *Big Data* **2013**, *1*, 1–2. [CrossRef] [PubMed]
2. IBM What Is Big Data? Available online: <http://www-01.ibm.com/software/data/bigdata/> (accessed on 7 May 2018).
3. Kitchin, R.; McArdle, G. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data Soc.* **2016**, *3*, 2053951716631130. [CrossRef]
4. Abbasi, A.; Sarker, S.; Chiang, R.H.L. Big data research in information systems: Toward an inclusive research agenda. *J. Assoc. Inf. Syst.* **2016**, *17*, 3. [CrossRef]
5. Rossi, E.; Rubattino, C.; Viscusi, G. For What It's Worth: A Multi-industry Survey on Current and Expected Use of Big Data Technologies. In *Proceedings of the 15th European, Mediterranean, and Middle Eastern Conference, EMCIS 2018, Limassol, Cyprus, 4–5 October 2018-LNBIP 341*; Themistocleous, M., Rupino da Cunha, P., Eds.; Springer: Cham, Switzerland, 2018; pp. 72–79.
6. AEGIS. D1.3—Final AEGIS Methodology. Available online: <https://www.aegis-bigdata.eu/wp-content/uploads/2017/03/AEGIS-D1.3-Final-AEGIS-Methodology-v1.0.pdf> (accessed on 23 September 2019).
7. Lavallo, S.; Lesser, E.; Shockley, R.; Hopkins, M.S.; Kruschwitz, N. Big Data, Analytics and the Path from Insights to Value. *MIT Sloan Manag. Rev.* **2011**, *52*, 21–32.
8. Lohr, S. The Age of Big Data. *New York Times*, 11 February 2012. Available online: <https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html> (accessed on 23 September 2019).
9. Phillips-Wren, G.; Iyer, L.S.; Kulkarni, U.; Ariyachandra, T. Business analytics in the context of big data: A roadmap for research. *Commun. Assoc. Inf. Syst.* **2015**, *37*, 448–472. [CrossRef]
10. Goes, P. Editor's Comments: Big Data and IS Research. *Manag. Inf. Syst. Q.* **2014**, *38*, iii–viii.
11. Agarwal, R.; Dhar, V. Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Inf. Syst. Res.* **2014**, *25*, 443–448. [CrossRef]

12. Viscusi, G.; Batini, C. Digital Information Asset Evaluation: Characteristics and Dimensions. In *Smart Organizations and Smart Artifacts SE-9; Lecture Notes in Information Systems and Organisation*; Caporarello, L., Di Martino, B., Martinez, M., Eds.; Springer International Publishing: Cham, Switzerland, 2014; Volume 7, pp. 77–86.
13. Buhl, H.U.; Röglinger, M.; Moser, F.; Heidemann, J. Big Data—A Fashionable Topic with(out) Sustainable Relevance for Research and Practice? *Bus. Inf. Syst. Eng.* **2013**, *5*, 65–69. [[CrossRef](#)]
14. Benington, J.; Moore, M.H. *Public Value—Theory and Practice*; Palgrave Macmillan: Basingstoke, UK, 2011.
15. Cordella, A.; Bonina, C.M. A public value perspective for ICT enabled public sector reforms: A theoretical reflection. *Gov. Inf. Q.* **2012**, *29*, 512–520. [[CrossRef](#)]
16. Morris, S.; Shin, H. Social value of public information. *Am. Econ. Rev.* **2002**, *92*, 1521–1534. [[CrossRef](#)]
17. Batini, C.; Rula, A.; Scannapieco, M.; Viscusi, G. From data quality to big data quality. *J. Database Manag.* **2015**, *26*, 60–82. [[CrossRef](#)]
18. Viscusi, G.; Castelli, M.; Batini, C. Assessing social value in open data initiatives: A framework. *Future Internet* **2014**, *6*, 498–517. [[CrossRef](#)]
19. Batini, C.; Castelli, M.; Viscusi, G.; Cappiello, C.; Francalanci, C. Digital Information Asset Evaluation: A Case Study in Manufacturing. *Sigmis Database* **2018**, *49*, 19–33. [[CrossRef](#)]
20. Kallinikos, J. Information out of information: On the self-referential dynamics of information growth. *Inf. Technol. People* **2006**, *19*, 98–115. [[CrossRef](#)]
21. Kallinikos, J.; Aaltonen, A.; Marton, A. The ambivalent ontology of digital artifacts. *MIS Q.* **2013**, *37*, 357–370. [[CrossRef](#)]
22. Xiao, N.; Sharman, R.; Rao, H.R.; Upadhyaya, S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. *Decis. Support Syst.* **2014**, *57*, 417–427. [[CrossRef](#)]
23. Kuruzovich, J.; Viswanathan, S.; Agarwal, R.; Gosain, S.; Weitzman, S. Marketspace or Marketplace? Online Information Search and Channel Outcomes in Auto Retailing. *Inf. Syst. Res.* **2008**, *19*, 182–201. [[CrossRef](#)]
24. Browne, G.J.; Pitts, M.G.; Wetherbe, J.C. Cognitive stopping rules for terminating information search in online tasks. *MIS Q.* **2007**, *31*, 89–104. [[CrossRef](#)]
25. Ghose, A.; Goldfarb, A.; Han, S.P. How Is the Mobile Internet Different? Search Costs and Local Activities. *Inf. Syst. Res.* **2013**, *24*, 613–631. [[CrossRef](#)]
26. Lucas, H.; Agarwal, R.; Clemons, E.K.; El Sawy, O.A.; Weber, B. Impactful Research on Transformational Information Technology: An Opportunity to Inform New Audiences. *MIS Q.* **2013**, *37*, 371–382. [[CrossRef](#)]
27. Bharadwaj, A.; El Sawy, O.; Pavlou, P.; Venkatraman, N. Digital Business Strategy: Toward A Next Generation of Insights. *MIS Q.* **2013**, *37*, 471–482. [[CrossRef](#)]
28. Dehning, B.; Richardson, V.J.; Zmud, R.W. The Value Relevance of Announcements of Transformational Information Technology Investments. *MIS Q.* **2003**, *27*, 637–656. [[CrossRef](#)]
29. Dillman, D.A.; Smyth, J.D.; Christian, L.M. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2008.
30. Hesse-Biber, S.; Griffin, A.J. Internet-Mediated Technologies and Mixed Methods Research: Problems and Prospects. *J. Mix. Methods Res.* **2013**, *7*, 43–61. [[CrossRef](#)]
31. Hewson, C. Internet-mediated research as an emergent method and its potential role in facilitating mixed methods research. In *Handbook of Emergent Methods*; Hesse-Biber, S.N., Leavy, P., Eds.; Guilford Press: New York, NY, USA, 2008; pp. 543–570.
32. Tsai, C.-W.; Lai, C.-F.; Chao, H.-C.; Vasilakos, A. V Big data analytics: A survey. *J. Big Data* **2015**, *2*, 21. [[CrossRef](#)]
33. Chen, M.; Mao, S.; Liu, Y. Big data: A survey. *Mob. Netw. Appl.* **2014**, *19*, 171–209. [[CrossRef](#)]
34. Porter, M.E.; Millar, V.E. How Information Gives You Competitive Advantage. *Harv. Bus. Rev.* **1985**, *63*, 149–162.
35. Zillner, S.; Becker, T.; Munné, R.; Hussain, K.; Rusitschka, S.; Lippell, H.; Curry, E.; Ojo, A. Big Data-Driven Innovation in Industrial Sectors. In *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*; Cavanillas, J.M., Curry, E., Wahlster, W., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 169–178.
36. Sambamurthy, V.; Bharadwaj, A.; Grover, V. Shaping Agility through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms. *MIS Q.* **2003**, *27*, 237–263. [[CrossRef](#)]

37. Chen, D.Q.; Preston, D.S.; Swink, M. How the Use of Big Data Analytics Affects Value Creation in Supply Chain Management. *J. Manag. Inf. Syst.* **2015**, *32*, 4–39. [[CrossRef](#)]
38. Chalias, S.; Myers, M.D.; Hess, T. Digital transformation strategy making in pre-digital organizations: The case of a financial services provider. *J. Strateg. Inf. Syst.* **2019**, *28*, 17–33. [[CrossRef](#)]
39. Matt, C.; Hess, T.; Benlian, A. Digital Transformation Strategies. *Bus. Inf. Syst. Eng.* **2015**, *57*, 339–343. [[CrossRef](#)]
40. Subramaniam, M.; Iyer, B.; Venkatraman, V. Competing in digital ecosystems. *Bus. Horiz.* **2019**, *62*, 83–94. [[CrossRef](#)]
41. Reis, J.; Amorim, M.; Melão, N.; Matos, P. Digital Transformation: A Literature Review and Guidelines for Future Research. In *Trends and Advances in Intelligent Systems and Technologies*; Rocha, Á., Adeli, H., Reis, L.P., Costanzo, S., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 411–421.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).