Exploring Qualitative Data: the use of Big Data technology as support in strategic decision-making

Manuela Bertei. Università di Pisa. Italy. manuelabertei@virgilio.it

Luciano Marchi. Università di Pisa. Italy. l.marchi@ec.unipi.it

Diego Buoncristiani. Big Data Consultant. Italy. diego_buoncris@yahoo.it

Abstract. This paper is a preliminary study that investigates whether the Big Data technology contributes to making the strategic decision-making more objective. Decisions depend on the decision-makers´ knowledge and intuition, the characteristics of decision task as well as the quality of the data analysed. Focusing on the decision objective component, it appears that the contemporary business context is characterised by: (i) a vast amount of data, but not all of this is meaningful; and (ii) an unexplored kind of data, namely qualitative data. The aim of the paper is twofold. Firstly, to elaborate a theoretical framework integrating Big Data technology into strategic decision-making. And secondly, assessing the feasibility of the elaborated model on a strategic variable in the hi-tech field.

Keywords: Strategic decision-making, Big Data, Business Performance Management, Unstructured data

1. INTRODUCTION

In the contemporary business context - characterized by dynamism, variety and variability of scenarios - a crucial role is played by information as a product of agent data elaboration. Several studies on contingency theory, applied to decision-making, have highlighted that the quality and significance of decisions is dependent on: (i) decision-makers' knowledge and intuition (Vroom and Jago,

1974; Beach and Mitchell, 1978; Shah *et al.;* 2012); (ii) characteristics of decision task (Beach and Mitchell, 1978) and; (iii) data collection (Blackwell, 1953; Galbraith, 1974). To react to external environment changes, decision-makers have a need for selective and well-timed but also relevant and accurate information (Miller and Mork, 2013; Bertei and Marchi, 2014).

Today, there is a vast amount of data, but not all of this is meaningful. Nevertheless, there is an unexplored kind of data, namely qualitative data. Qualitative data is typically unstructured data and constitute the main source of information (Forrester, 1975; Sterman, 2000; Luna-Reyes and Andersen, 2003). This data is mainly stored in the agent's head (mental database) and just a part of that is translated in the form of written texts (Forrester, 1994).

In the recent years, Big Data technology has become popular. Like Business Intelligence tools, the purpose of Big Data is to translate data into business advantages. The differences between the two technologies are represented by the well-known three Vs: Volume, Velocity and Variety (McAfee and Brynjolfsson, 2012).

Our research totally embraces on one hand the literature about the three Vs´ benefits; on the other hand the use of these new technologies as a strategic decision-making process to better understand customers, competitors, market features, products and services. In literature, there is already a plethora of contributions on the analysis on those dimensions (Xiang *et al.;*2015).

Our paper takes into account the contingency theory and investigates the potential of Big Data technology in order to: (i) decode and use the ensemble of tacit unstructured data which is not taken into account with the purpose of improving the strategic decision-making process; and (ii) find a right balance between the subjectivity and objectivity information component at the strategic level.

This research is a preliminary study and is divided into two steps. First, to elaborate a theoretical framework integrating Big Data technologies into strategic decision-making in business firms. Second, to assess the feasibility of the elaborated model on a strategic variable in the hi-tech field.

Before introducing the topic, we review literature about the decision-making process and on the evolution of IT tools to support decision-making at every level. Next, we illustrate a theoretic framework to investigate the objectification of strategic decision-making. Added "new" information at the traditional sources of data, then we test the framework on a strategic variable in the hi-tech industries. Finally, we provide our conclusions.

2. LITERATURE

Decision-making coupled with Big Data is not new in literature. Although such integration has already been put in practice for operational decision-making, it is still a new issue for strategic decision-making (Wladawsky-Berger, 2013, Provost and Fawcett, 2013). Authors as Brynjolfsson *et al.* (2011), McAfee and Brynjolfsson, (2012) and Provost and Fawcett (2013) have stated that data-driven decision-making processes are better because they are based on evidence rather than intuition.

To better understand this integration, literature review will be divided into three sections: (i) the decision-making process analysis, (ii) the relationships between decision-making level and information, (iii) the evolution of Information Technology (IT) tools as decision-making supports

2.1 The decision-making process analysis

By decision-making, we refer to *de-caedere*, that is, cutting off some choices between those available.

First of all, two decision-making paradigms were distinguished, the rational choice paradigm and the bounded rationality paradigm. The former defines a decision as a cost-benefit maximization process. The decision-makers know their targets and classifies them depending on their priorities. They also know all the possible alternatives and can choose the solution maximizing their utility. This paradigm is based on the superiority of intellect over experience. The bounded rationality paradigm was introduced by Simon (1979). This paradigm originated in opposition to the previous one and is based on the assumption that the decision-makers´ rationality is limited by a number of factors, e.g. the information available, personal limits, the time available to take decisions, etc. Differently

from the previous paradigm, here the decision-maker might not know all the possible solutions and cannot choose the alternative maximizing his/her utility, but the most satisfying one among those available.

In this paper, we accept the bounded rationality paradigm and discuss the two schools of thought linked to it, that is, the analyst and the experiential schools. They both share the concept that the decision-making process is composed of three steps (Simon, 1969): (i) problem definition; (ii) identification, evaluation and selection of alternatives; (iii) implementation alternative. The two schools though, weigh the single steps differently. The analytic school gives more emphasis to the first two phases, that is, the problem definition and the identification of alternatives, while the last step is considered as a simple execution of the selected alternatives. The experiential school on the contrary, values the implementation alternative step more than the other steps and introduces the concept of feedback control related to decision-making process to adjust behaviours and actions according to the set objectives.

Starting from the above mentioned schools, different decision-making theories were developed and each one has contributed to define the different features of the current decision-making process, such as the incremental view theory, the garbage can model, the individual and naturalistic perspectives of decision makers, and the speed theory.

One of the first theories on decision-making was the incremental view theory proposed by Lindblom (1959). The author stated that the decision-making was an evolutionary and not a revolutionary process. Sharing the experiential school's view, he stressed the importance of the implementation alternative step as well as of feedback control continuously promoting learning and decision adjustments. The evolutionary feature of decision-making is a compromise of the multiplicity of decision-makers involved on it and the comparison with historical data, which does not contribute to a separation from the status quo.

In the garbage can model, the decision process was compared to a garbage can defined by Cohen *et al.* (1972) as: "the meeting point of a problem in search of a solution, a solution in search of a problem and actors whose attention is divided, who come and go, but who happen to be at the meeting place of the problem and

solution". Here the decision process does not solve problems and seems to be a meeting point with no benefits.

In the theories expounded below, the focus shifts from decision-making to decision-makers. The first authors to highlight the importance of the decision-maker's personality were Keen and Scott Morton (1978) in their individual differences perspective theory. They stated that a unique and standard decision-making process does not exist as it is influenced by subjectivity. Thus, decision-making is strictly linked to the decision-maker's experience and intuition rather than on rational choices.

Also naturalistic decision-making perspective advocates believe a standard decision process is impossible. Decisions are not influenced by the decision-maker's personality but by the context in which organizations operate. The decision-maker's personality and the organization context flow into the multiple perspective approach. This theory claims that the decision-making process, though rigorous, is influenced by the personality of the users involved. This way, the individual and naturalistic theories are developed and connected: the decision process is affected by the decision-maker's personality (individual perspective) but also by the personality of the other internal and external agents (naturalistic perspective) .

Finally, speed theory, focused on the uncertainty characterising modern social contexts. Uncertainty causes the need for additional information, which slows down the decision-making process and alters performances.

2.2 The relationship between decision types and information requirements

Information plays a key role in decision-making. The main difference between data and information is given by the information capability to meet (Isik *et al.;* 2013): (i) decision types, (ii) the decision-maker's information requirements.

The two aspects belong to the decision environment field, which can be defined as "the combination of different types of decisions made and information processing requirements of the decision-maker when making those decisions" (Munron, 1977).

Regarding the first dimension (decision types), a framework was proposed by

Simon (1960). The author classified decisions in programmed and non-programmed ones. A decision is programmed if it is repetitive and used routinely, and when the decision-maker knows all the alternatives; its consequences are predictable with a high level of accuracy. Non-programmed decisions are linked to problems that the organisation cannot clearly define and when alternatives are uncertain. In this case, solutions cannot be found rationally.

The second dimension, information requirement, was analysed by Anthony (1965) in his framework of managerial activity. Three categories of activities were distinguished and grouped in the pyramid: operational control, management control and strategic planning.

At the pyramid base there is operational control that is "the process of assuring that specific tasks are carried out effectively and efficiently" (Anthony, 1965). This type of control is concerned with task and characterized by repetitive and routine decision-making activities; for these reasons the information requires a high level of detailed data on a frequent repetitive basis and is frequently provided by internal sources. Information for operational control is usually short-term and based on historical data.

At the intermediate level of the pyramid there is management control. The activities belonging to this level involve planning and control activities; they focus on the definition of economic and financial objectives and assess if the way chosen to obtain resources follows efficiency and effectiveness criteria of the performance. At this level, the information is mainly internal and is the outcome of elementary data coming from operational control. The information is provided periodically and used to compare the set objectives with the results obtained to adjust them if necessary.

Finally, on the top of the pyramid there is strategic planning that can be defined as "the process of deciding on the goals of the organization and the strategies for attaining these goals." The latter may involve changes in these goals, and/or in the policies for the acquisition and allocation of the resources to attain the set goals. The information necessary at this level is both internal and external and based on historical and forecasting data. The purpose of this activity is to understand as soon as possible future developments of the company to adopt the most

advantageous policy. The information produced regards the long period, and its value depends on the decision-maker's ability to interpret it.

The Gorry and Scott-Morton (1971) framework came out from the integration of Anthony (1965) and Simon (1960) frameworks. The authors identified three types of decisions: structured, semi-structured and unstructured decisions. The structured decisions are typically routine activities and follow a standard procedure; on the contrary the unstructured decisions have no fixed methods of handling it and do not follow a repetitive procedure. Decisions falling between the two categories are ranged as semi-structured.

Some final considerations on the frameworks exposed should be given. First of all, there is a difference between the concept of programmable decisions proposed by Simon (1960) and the decision structurability concept proposed by Gorry and Scott Morton (1971). The concepts of non/programmed are focus on the dependence of computer instead of the un/structural concept is focus on the problem solving activity.

The final remark is about the non-complete correspondence between Anthony's classification (1965) and Gorry and Scott Morton's (1971). The structurability/unstructurability level of a decision is not necessarily linked to its operational/strategic level. So, there might be unstructured operational decisions and semi-structured strategic decisions, which vary in the frequency these types of decisions, occur, i.e. operational decisions will be mainly structured whereas strategic decisions will be mostly unstructured.

2.3 The evolution of IT tools as decision-making supports

Companies operate in a complex business context in which the vast amount of information and the variety of data sources produced contrast the need for fast decision-making to keep pace with the dynamism of scenarios. Decision-makers have always had to cope with these aspects, especially the ones coming from the external environment, and they are considered the starting points for IT - Information Technology (IT) as a part of the Information System (IS) - in companies.

Our analysis starts off with the framework proposed by Gorry and Scott Morton

(1971), which highlights that different decisional activities require different information. This idea was ignored for a long time, which resulted in the failure of introducing IT in organisations.

Gorry and Scott Morton's framework identifies nine decisional classes; we suggest re-mapping this framework with the state-of-the-art IT tools currently used to face decision-making aspects (Figure 1).

Figure 1. A representation of integration decision types/IT tools

The information required for structured decisions at every level – operational, management and strategic – and for the semi-structured operational decisions are now managed with the Extended Enterprise Resource Planning (eERP or ERPII). These systems are an evolution of the Enterprise Resource Planning (ERP), that is, software-based modules capable of representing business processes with several implementations applicable at different times.

ERP's traditional functions, i.e. Human Resources, Sales and Marketing, Financials and Logistics, are integrated with more external information related to the operational context and customers with modules such as SCM (Supply Change Management), CRM (Customer Relationship Management), SRM (Supplier Relationship Management) and PLM (Product Lifecycle Management) (Chen, 2001). ERP originated from systems that could focus just on planning and control internal resources, that is, MRP (Material Requirements Planning) and MRPII (Manufacturing Resource Planning). MRP was developed to efficiently

calculate production needs. The evolution of this technique is MRPII which includes more functions, e.g. sales planning (Klaus *et al.;* 2000). The information obtained is mostly for structured decisions at operational level.

Analysing the information already present in an organisation and managed through various systems (i.e. ERP, CRM), has given rise to a new set of applications known as Business Intelligence. BI provides information for managerial and strategic semi-structured decisions and is defined as applications and technologies for data gathering, storing, analysing and visualizing to help decision-makers in their choices (Negash, 2004).

The primary activities of BI are "getting data in" and "getting data out". The former can be obtained by data warehousing and has the aim of moving data from several heterogeneous sources into a repository called data warehouse (DWH). These various sources can have different origins, as instances, they might come from internal or external systems or data can be structured or unstructured. All this mix-up of information coming into a data warehouse generates such problems as data cleansing, integration and transformation, and is one of the most relevant aspects of BI, requiring the 80 percent of the time, efforts and unexpected project costs.

The latter activity is the most interesting part of BI in an organisation. After the data are located on a DWH, data marts are created in order to analyse data and extract useful information. The analysis is supported by query processing techniques and the results are spreadsheets, reports, dashboards, scorecards, interactive visualization, KPIs and the like.

Finally, the main objective of Business Intelligence is enabling managers and analysts to make better and faster decisions in order to get benefits in terms of costs, business performance and all those aspects that can lead to improving decision-making. All in all, the purpose of BI is then to translate data into business advantages.

The production of information for unstructured decisions – operational, management and strategic – is covered by Business Performance Management (BPM) (Frolik and Ariyachandra, 2006). BPM coexists under a twofold aspect: technology and process. Regarding the first aspects, BPM is wrongly considered

as a synonymous of BI (Frolik and Ariyachandra, 2006). BI provides tools and applications to implement BPM. BPM incorporates a business process that support BI. That does not mean that the latter needs of the former to exist, but it can be considered as an enhancement. The relationship between them is univocal, BI can exist without BPM, the vice versa cannot be possible.

BPM as a process can be defined as a set of processes helping organizations optimise business performance and with the aim to better understand one's own company and establish the strategy, actions and decisions to be taken (Golfarelli *et al.;* 2004). BPM process is divided into four steps: (i) Strategize; (ii) Plan; (iii) Monitor and Analyse and; (iv) Take correct action (Figure 2).
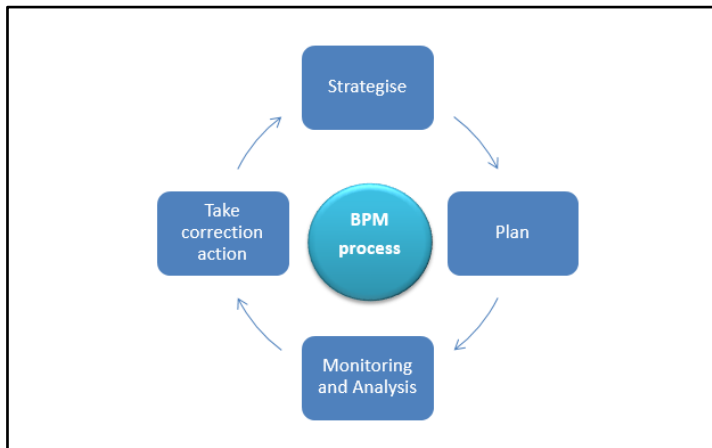
Figure 2. A representation of BPM process

The definition of the strategy is the first step in this process. An organisation focuses on what to achieve (strategies). In the plan step the action to be taken are developed with tools of planning and control such as business planning, budget and the like. The last two steps (monitoring and analysis) are responsible for the execution of the strategy. In the monitoring and analysis step the performance are compared with the targets defined into the plan step. Corrective action can be generated by feedback and/or feed forward mechanism to establish the accuracy of the planned actions. Finally, the last step implements the corrective actions with the aim to re-establish the right direction to achieve the strategy (taking corrective action).

The vast amount and variety of data, together with the requested time of analysis are considerable limits of traditional Business Intelligence tools. Hence a new research field has been originated in order to expand the BI's boundaries: Big data is considered a suitable solution. This field of studies will be analysed in the next paragraph.

3. BIG DATA AS STRATEGIC DECISION-MAKING SUPPORT

Literature review has shown as decision can be classified into several decisional categories supported by different IT tools. Regarding strategic decisions, Business Performance Management does not taken into account unstructured data that can contribute to generate value. In this paragraph, we focus on the characteristics of Big Data features and we show a framework on which BPM process and Big Data are linked.

3.1 Big Data Technologies

Social networks, sensors, mobiles, Internet clicks generate a multitude of data that can hide relevant and fundamental meanings for organisations.

The aim of Big Data is still to translate data into business advantages, which represents the heterogeneous and vast amount of data that can be collected and analysed. Big Data unveils the trends and patterns of these data with powerful computational techniques (George *et al.* 2014).

Figure 2 shows as Big Data pervades each decisional level, as well as, contributes to explore "new" data sources. Big Data does not substitute, but rather integrates the previous technologies, such as ERP II, BPM the like.

In spite of the "Big" label, Big Data is not only data volume but it coexists in the three Vs aspects of "Volume", "Variety" and "Velocity" (McAfee and Brynjolfsson, 2012; Mayer-Schönberger and Cukier, 2013; Sagiroglu and Sinanc, 2013; Kwon *et al.*,2014; Wu *et al.,* 2014; Waller and Fawcett 2013).

According to Russom (2011), "Volume" is considered the primary attribute of Big Data. Organisations are dealing with terabytes and petabytes of data and this outstrips the traditional databases (Thusoo *et al.*,2010; Madden, 2012; Shah *et al.*, 2012; Provost and Fawcett, 2013; Chen *et al.*, 2014).

The "Variety" attribute of data sources can make Big Data very big (Russom, 2011; Sagiroglu and Sinanc, 2013). The datasets available for analysis are not only from structured data − databases − but also can be found in unstructured data − texts, organisations documents, emails and more − and semi-structured data − xml and RSS feed.

"Velocity" allows obtaining information in real time or nearly real time or process the data more quickly that helps organisations to move faster into the business context and make them more agile (Davenport *et al.;* 2012; McAfee and Brynjolfsson, 2012).

The framework (Figure 3) shows how Big Data technologies process data in a decisional business context.
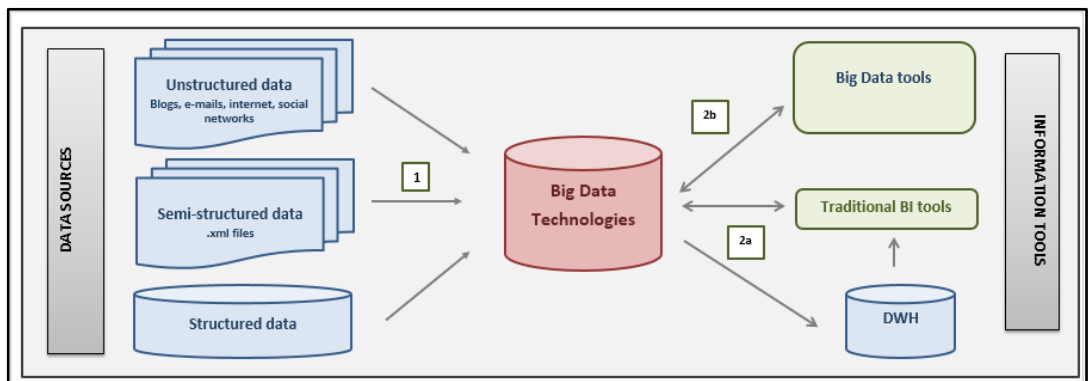

Figure 3 Big Data framework

After the dataset has been defined, the Big Data technology can be fed with various types of data: unstructured data, semi-structured data and structured data.

Dataset is now available into the Big Data technologies and it can be elaborated, processed and analysed with several tools, not only coming from the Big Data area, but also from the Business Intelligence field. The best approach, at this stage, depends on the analysis to be performed and the competencies available in an organization.

3.2 The Big Data integration into BPM process: the framework

The previous framework - about the features of Big Data technologies and strategic decision-making process - has highlighted as the lack of information is

falling down with Big Data technologies and that leads to focus on data-driven instead of intuition-driven approach (McAfee and Brynjolfsson, 2012, Provost and Fawcett, 2013).

Integrating Big Data as support for strategic decision-making process resides into the contribution of the three Vs: (i) data volume is a critical factor in the current challenging business context. On one hand, data are everywhere and organizations need for new systems to store them. On the other hand, there are many unexplored data sources that organisations want to discover for better predicted outcomes with greater precisions (George *et al.;* 2014); (ii) Velocity allows to take decisions timely that can lead organization to wear a market leader role and; (iii) Variety lets to endorse the available information sources. Strategic information is not strictly represented in database formats but can be also located in sources hard to access, i.e. e-mails, internal companies' documents, mental databases (Forrester, 1994; Madden, 2012; Lin *et al.;* 2014, Kwon *et al.;* 2014).

The framework proposed for business firms focuses on how the identification, evaluation and selection of data sources impact on the strategic decision-making process.

The process starts when the decision-makers establish what the organisation wants to achieve – goals, policies and directions (Step 1).This step is fundamental to understand the information needs and constitutes the first meeting point between Big Data and BPM. This integration allows to enlarge the data sources taken into account, starting from structured to unstructured ones (Variety). The availability of more data sources (Volume) and their timely processing (Velocity) is considered hard with the traditional technologies (Figure 4).
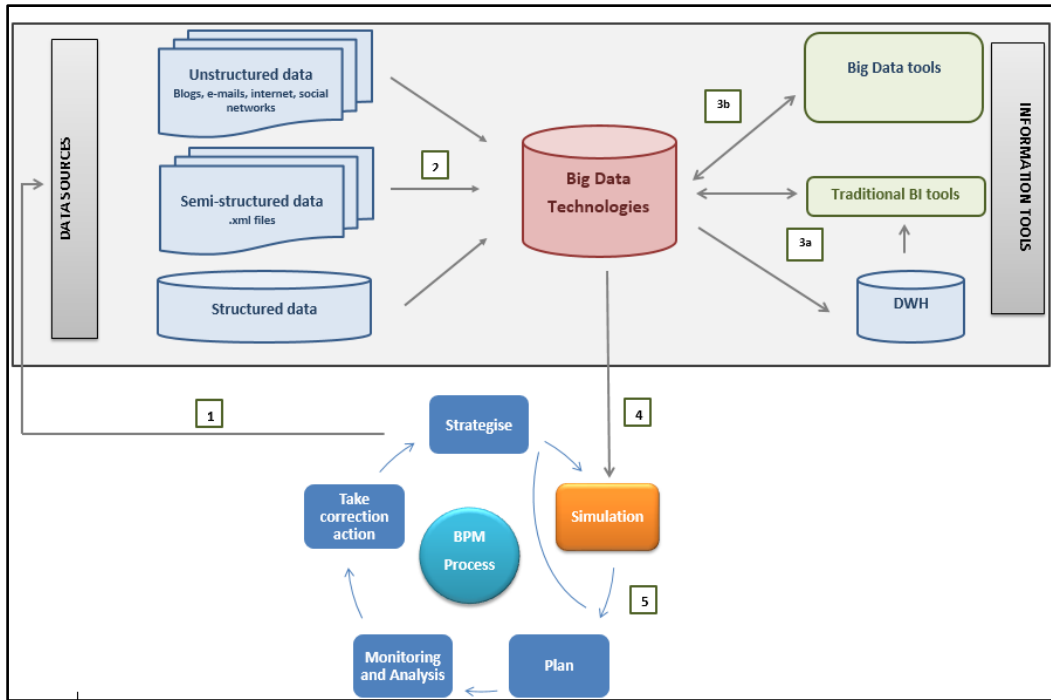
Figure 4. The Big Data/BPM framework

Now, the data collected are not in a `prematerialised´ view (step 2) and they can be analysed with the traditional BI tools (step 3a) or with some specific Big Data technologies (step 3b). After the various elaborations, the results are available in the databases and decision makers can use them, in order to plan actions. In addition to the traditional BPM steps, in this framework we take into account simulation tools in order to improve the results of the analysis with the decision-maker knowledge and experience. Apart for the simulation step, the BPM cycle can keep the traditional route (step 4). Nevertheless, some corrections are necessary. Firstly, an absolute balance between the subjectivity and objectivity information component does not exist. For this reason a right balance based on the costs/benefits evaluation should be found for each decision. Secondly, the framework is a preliminary output and cannot be considered as a best practice but an alternative approach to enlarge the strategic decision-maker process based on Business Performance Management. Finally, the Big Data technologies can improve the accuracy of decision-making process, but it cannot be considered as a replacement for the decision-makers (Shah *et al.;* 2012).

4. RESEARCH DESIGN

4.1 Research setting

Emotions, as well as additional information, can play a key role in human decision-making (Choo, 1993). This additional information can be found everywhere. The study focuses on a popular data source, namely the social network.

Authors as Mishne (2007), Davenport *et al.;*(2012) and Lin *et al.* (2014) stated that crucial information to support strategic decision-making can be extracted from social networks (Twitter, Facebook, blogs and the like). Also Xiang *et al.* (2015) have already explored the utility of online user reviews on a website in the hospitality management, analysing the customer satisfaction and the guest experience dimensions. Duan *et al.* (2008) have examined the effect of online user reviews in the area of movie box office sales predictions. On this ground, the framework elaborated is being tested to support a new product development within strategic decision-making in the high-tech industry (Figure 5). The hi-tech industry features –negative free cash flow at early stages, availability of little historical financial data, lower generated value from physical assets, main role played by human factors, market sentiment influence, high default risks and severe volatility of stocks – make this decision more critical and lead to a set of new information needs in addition to the traditional ones (Bollen *et al.;* 2011).

In particular, we will focus on the development of a new version of a famous smartphone and want to assess if the framework can contribute to make the strategic decision-making more objective with the support of Big Data tools.

Strategic decision-making starts with the problem definition (Step 1) that in our case is the development of a new smartphone. This step is fundamental to understand the information needs in order to lead the business strategy.

The data source explored through Big Data tools is a well-known social network. The firm that produces the smartphone has created a social network page about the prototype of the new smartphone version and has posted images, potential technical features asking the viewers for comments.
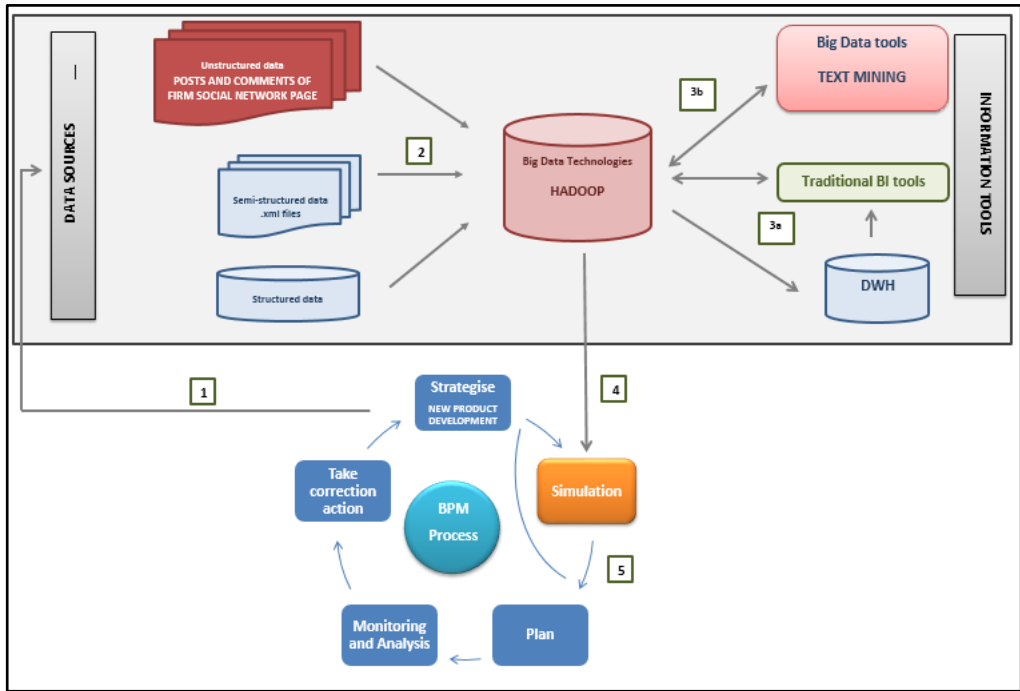
Figure 5. The Big Data/strategic decision-making framework application

Textual data are imported to the Hadoop Distributed File System (HDFS) through the Hadoop Ecosystem tools, and other techniques in the Hadoop Ecosystem are available, in order to import data into the Hadoop system (Step 2). Differently from the data warehouse approach, the data introduced into Hadoop system are adaptable, which means they don't appear in a pre-materialized view.

Here data are ready for analysis that can be accomplished with Hadoop Ecosystem tools and some Text Mining techniques (step 3). The information produced should be translated into simulation variables and their manipulation will then allow planning different simulation scenarios (Step 4).

This is the starting point for the planning process. Decision-makers discuss and evaluate threats, opportunities and coherence of each scenario in order to develop an action plan on how to carry out the business strategy (step 5). After this step the actual plan definition is accomplished by the traditional BPM, starting again with the monitoring and taking action phases.

Our analysis focuses on the conversion process from unstructured data to

relevant information for the strategic decision-making.

Before the pilot study started off, we conducted an exploratory questionnaire within an Italian smartphone firm. The questionnaire was addressed to all managers involved in the product development process. The aim was to understand the context of our research more in depth; identifying the strategic resources and the data sources.

The questionnaire is organised into two sections. The first part focuses on the importance of the information for the business context taken into account and the IT infrastructure available. The second one investigate whether the social network is a relevant component and how to use it.

4.2 Data sampling

We collect data from the social network page of a world famous smartphone company. This setting is particularly suitable for our study, as the social networks include a huge amount of unconsidered unstructured data– textual posts and comments, photos and videos, emoticons, hypertextual links and the like – and they allow interacting with existing and potential customers.

The analysis of data sampling is divided into two phases. In the first one, we conduct a posts analysis with the aim to understand more in depth the information need of the firm. In the second one, we focus on the comments left by the viewers. They are in the form of unstructured data; our aim is to convert them into information valuable for strategic decisions.

4.2.1 Posts analysis

We start our research applying the method of genre analysis (Figure 6). Genre analysis is used to classify the post dataset with the aim to understand and identify patterns and logic in the published posts.

In literature, there are several definitions of genre; in this work we adopt the definition proposed by Swales (1990) who defines genre as "a class of communicative events, the members of which share some set of communicative purposes".

Authors as Herring *et al.* (2004) have applied genre analysis to identify the features of weblogs. Riemer *et al.* (2012) have already used this methodology to investigate communication on Yammer platform to establish relationship and practices in this social network. We obtained a collection of 49 posts from the smartphone social network page.

As a result of the analysis we identified 3 genres categories divided into in maximum three subcategories each one. For our research, we have identified four categories of posts: (i) technology; (ii) design; (iii) comparison and; (iv) other.
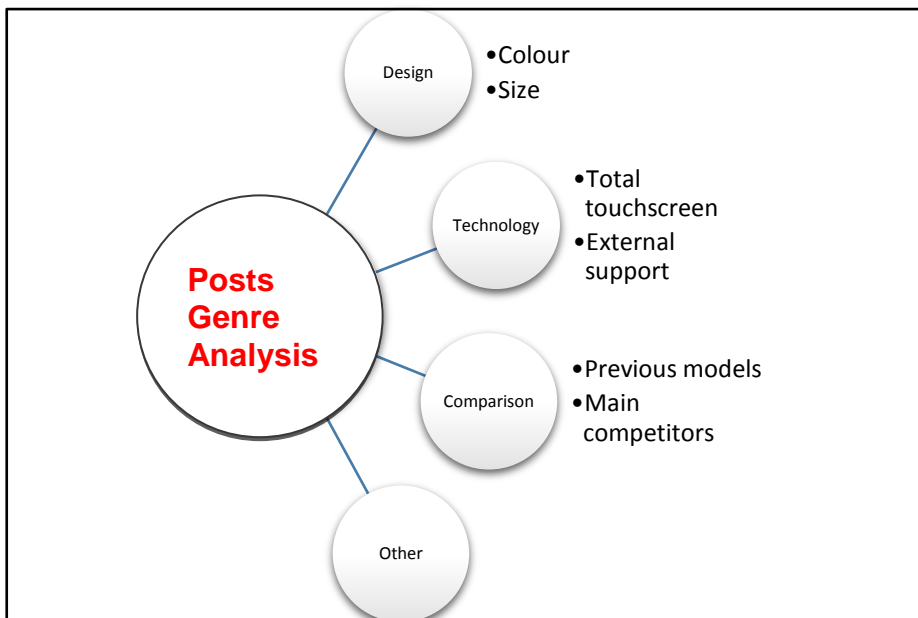

Figure 6. A representation of post genre analysis

The design category is divided into two subcategories: (i) Colour; and (ii) Size. The former represents the subclass of posts related to new colours or materials for the smartphone. The latter refers to the dimension of the product.

The technology category is related to posts describing new technology features and functionalities for the product. We pinpoint two subcategories: (i) Total Touch-Screen and (ii) External Support. The former refers to the smartphone which does not integrate physical components, but it is only based on the touch-screen. The latter refers to the chance to use external components as keyboards or various plug-ins.

The Comparison posts focus on the differences with other products: (i) Previous Model and (ii) Main Competitor. The first subclass identifies the posts that discuss on the evolution respect to previous models and the last one on the distinctions with products of the company's main competitor.

The latter category-Other post - contains: (i) Residual, that is, posts that are not included in the previous categories and; (ii) Irrelevant, that in our opinion are insignificant as to build a specific category in which insert them. For the purposes of this work, the "Other" posts category is not take into account.
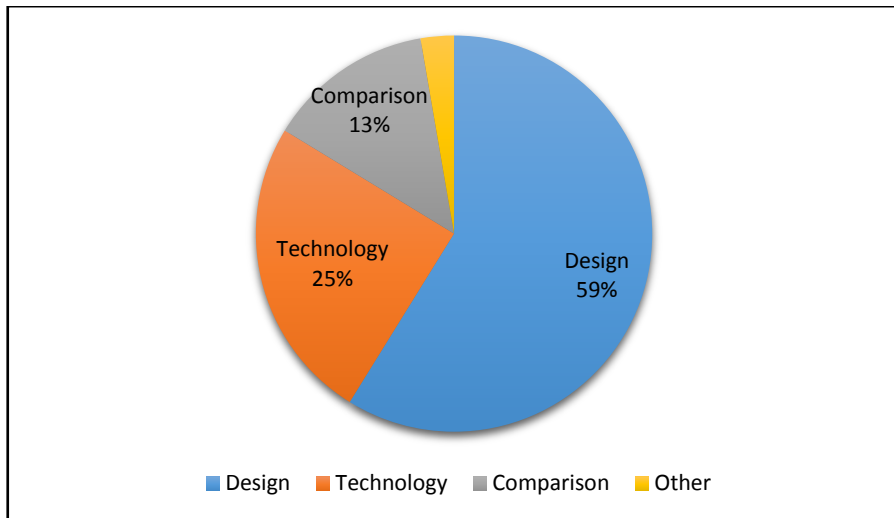

Figure 7. Distribution of posts´ genre analysis

Figure 7 shows the proportional distribution of posts´ genre categories. The most common genre found in the data sampling is the Design posts category with up to 59% of the total amount of posts, followed by technology (25%) and Comparison (13%). This result shows as the company is careful on the design and it prevails on technology features.

4.2.2 Comments´ analysis

We identified a total of 59.500 comments and, in the table 1, we provide their overview grouped for each category as pinpointed for the posts with the genre analysis.

Categories	Sub-categories	Posts	Comments
Technology	Total Touch-Screen	7	8.550
	External Support	4	10.115
Design	Colour	18	15.764
	Size	6	6.439
Comparison	Previous models	3	9.235
	Main competitor	3	9.324

Table 1. Overview of comments.

The administrator, who can ask for 'likes', otherwise in case of dislikes comments are asked, manages opinions. In this way, the stakeholders can indirectly collaborate on the development of a new product.

The comments' analysis has been realised with some Text Mining techniques. Text Mining is the process of extracting interesting, novel and valid patterns or knowledge from unstructured textual data (Ah-Hwee Tan, 1999). The analysis process is rather complicated due to the use of slang, emoticons and languages other than English. All the comments and posts were translated into lower case because we reckon that does not necessary mean an emphasis.

At the early stage we applied a text normalisation, in order to clean up the text from stop words, transcription and typing errors, punctuation, emoticons. The text normalisation is a crucial step in these analysis inasmuch the statistics, indexes or the like need to be computed on a well-established form. This process can facilitate the investigation of the outcomes obtained, Nevertheless, to complete the data cleansing we have also removed spam comments and correlated posts, because we reckon that useful information cannot extrapolate from them. For our analysis and for the type of textual data, these steps are adequate to obtain our canonical form.

4.3 Preliminary findings

In this section, we present the preliminary results of the design posts' analysis. This category is divided into two sub-categories: (i) colour; and (ii) size.

We identified 16 posts for the subcategory colour related to a gold coloured smartphone and to the comparison of a smartphone coloured in pink versus the green version. In the posts for the "gold smartphone" the visitors leave opinions, comments and "likes". All this information is merged in the same data set. For our descriptive analysis, we started with a word count to evaluate the word-frequency and "love", "beautiful" and "expensive" were the most frequent words; in a certain percentage, negative comments were also provided. This initial evaluation helped us to reduce the set of keywords to focus on.

As second step we analysed the adjective "expensive" in the context, inasmuch as there are two considerations: (i) the user is interested to buy this version of the product, but the cost can be a limit; (ii) the adjective can be the main motif for the negative comment or reinforcement to that. Nevertheless, the adjective expensive is used in positive comments in the 62%.

In the comparison colour post, users just indicate their preference: "1" for the pink coloured smartphone and "2" for the green one. The result obtained is:

- The 50% of the users prefer the smartphone in green, and the 85% of them has expressed the preference in a correct way (as indicated by the administrator), whereas the 15% indicated their preference with a word – green or two – or repeating n-times the number 2;

- The 20% of visitors is interested in both versions;

- And the 30% tends to the pink version.

In the posts identified for the subcategory "Size" is required to the users a "like", nevertheless part of the visitors left a comment – positive or negative. The most frequent words are: (i) Want; (ii) Fake. In the graph the analysis of the word counting is presented.
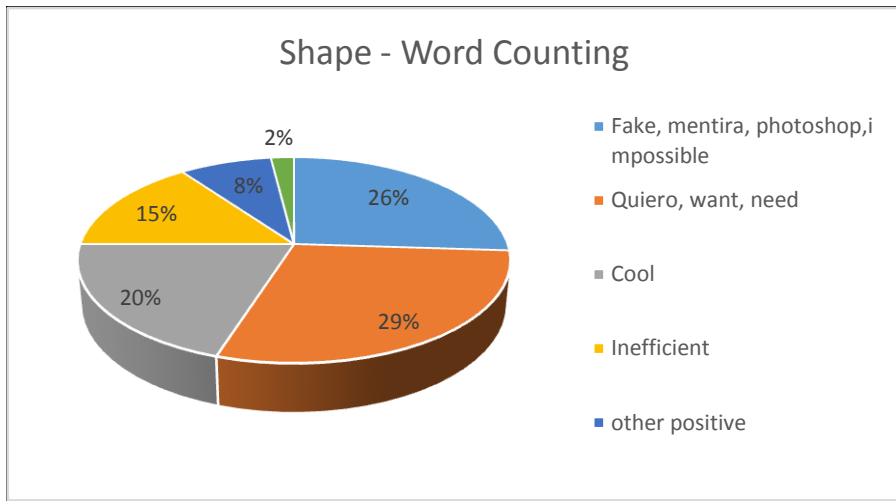
Figure 8: Distribution of word counting

The table shows, as two other words are more frequent, that are "inefficient" and "cool". These words have two different potential consumer underlying perspectives. Parts of users give more importance to the smartphone "external" aspect, the rest of users focus on the performance technical performances. The use of the "inefficient" word rather than "out-dated" stand for the users who gives these comment are probably expert. The weight of the two positions is different. In fact, the opinions about the external aspect of the smartphone change rapidly moreover, there are many *escamotage* to mask that, such as cover. Wrong technical features can be compromise the use of the smartphone and can be generate non-qualitative costs as guarantee costs, negative word of mouth and client disaffection.

5. CONCLUSIONS AND FURTHER RESEARCH

Our work can be included in the contingency theory stream and it focuses on the decision-making process. It takes into account the role of Big Data tools as support for strategic decision. In deeper we investigate if and how a variety of the information contributes to make more objective the decision process. To this purpose, we propose a theoretical framework in which the Big Data technologies can play a role in the exploration of unstructured and unusual data sources.

The framework is tested on the decision-making process for the product development in the hi-tech industry. The first results come out from the social network posts analysis that have highlighted as the Big Data tools can contribute to obtain a feedback from the main stakeholder, that is, the consumer. The consumers have indirectly contributed to the development of the new smartphone version. The information elaborated integrates the traditional information with the aim to make more objective the strategic decision-making as well as to improve the decision-maker knowledge.

Our study contributes to researcher and practitioner perspective. Regarding the former we propose a framework on which investigate for future analysis on Big Data and BPM. We also illustrated that the Big Data can be used to unveil novel trends and patterns from several data sources. The latter addresses the industry towards the customers' needs before a product can be effectively released. For example, a high-tech company can replace a survey with posts on social networks to find and understand the market requirements.

Like any investigation, our work has some limitations. The first limitation refers to the use of text mining tools only. As for the data set, the limited number of posts depends on the page data foundation – 2 months only.

Nevertheless, the several posts were fitted with the purpose to make more rational the answer and avoid problems, such as different languages, the length of comments – we have managed some criticisms that can have affected the dataset - some aspects as emoticons do not take into account. Finally, the evidences are at the preliminary stages as well as the analysis on data.

Further empirical researches will be addressed to test the potentiality of the Big Data tools as strategic decision-making support. Multiple case studies will be conduct to assess the possible benefits come out from the implementation of the framework into the real context.

6. REFERENCES

AH-HWEE, T. (1999): Text Mining: The state of the art and the challenges,

ANTHONY R.N (1965): "*Planning and Control Systems: a Framwork for Analysis*", Harvard University Graduate School of Business Administration, Boston.

BEACH L, MITCHELL T.R. (1978): "A Contingency Model for the Selection of Decision Strategies", *Academy of Management*, vol.3, n.3: 439-449.

BERTEI M.; MARCHI L. (2014): "A new perspective for Accounting-based Business Planning Simulation: an integration with Agent-based Simulation", *Information System, Management, Organization and Control: Smart Practices and Effects*, Springer, vol. 6: 273-287. http://dx.doi.org/10.1007/978-3-319-07905-9_20

BLACKWELL. D. (1953): "*Equivalent Comparison of Experiments*" Ann. Math. Statist.

BOLLEN J.; MAO H.; ZENG X. (2011): "Twitter mood predicts the stock market", *Journal of Computational Science*, vol.2: 1-8. http://dx.doi.org/10.1016/j.jocs.2010.12.007

CHEN I.J.(2001): "Planning for ERP Systems: Analysis and Future Trend", *Business Process Management Journal*, vol. 7, n.5: 374-386. http://dx.doi.org/10.1108/14637150110406768

CHEN M.; SHIWEN M.; YUNHAO L. (2014): "Big data: A Survey", Mobile Networks and Applications, vol. 19, n.12: 171-209 DOI: http://dx.doi.org/10.1007/s11036-013-0489-0

CHOO C.W. (1993): "Environmental Scanning: Acquisition and Use of Information by Managers", *Annual Review of Information Science and Technology*, vol. 28.

COHEN M.D.; MARCH J.G.; OLSEN J.P. (1972): "A Garbage Can Model of Organizational Choice", *Administrative Science Quarterly*, vol. 17, n.1: 1-25. http://dx.doi.org/10.2307/2392088

DAVENPORT T. H.; BARTH P.; AND BEAN R. (2012) "How 'Big Data' Is Different" *MIT Sloan Management Review vol.* 54, no. 1pp.

DUAN, W.; GU, B.; WHINSTON, A. B. (2008): "Do online reviews matter? — An empirical investigation of panel data", *Decision support systems*, vol. 45, n.4: 1007-1016. http://dx.doi.org/10.1016/j.dss.2008.04.001.

FORRESTER JW. (1975): "Industrial Dynamics − a Response to Ansoff and Slevin", *Management Science*, vol.14: 601-614.

FORRESTER JW. (1994): "Policies, decision and information sources for modelling", *European Journal of Operational Research*.

FROLIK M.; ARIYACHANDRA T.(2006): "Business performance Management: One Truth", *Business Intelligence*, vol. 23, n.1, 41-48. http://dx.doi.org/10.1201/1078.10580530/45769.23.1.20061201/91771.5

GALBRAITH, J.R. (1974): "Organization Design: An Information Processing View, Interfaces". http://dx.doi.org/10.1287/inte.4.3.28

GEORGE G.; HAAS M.H.; PENTLAND A.(2014): "Big Data And Management", *Academy of Management Journal*, vol. 57, n.2: 321-326. http://dx.doi.org/10.5465/amj.2014.4002

GORRY G.A.; SCOTT-MORTON M.S.(1971): "A Framework for Management for Information System", *MIT Sloan Management Review*: 21-36.

HERRING, S. C.; SCHEIDT, L. A.; BONUS, S.; WRIGHT, E. (2004): "Bridging the gap: A genre analysis of weblogs". *In System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference*. IEEE.

ISIK O.; JONES M.; SIDOROVA A. (2013): "Business intelligence success: The roles of BI capabilities and decision environments", *Information & Management*, vol.50: 13-23. http://dx.doi.org/10.1016/j.im.2012.12.001

KEEN P.G.W.; SCOTT-MORTON M.S. (1978): "*Decision Support System: An Organizational Perspective*", MA Addison-Wesley.

KLAUS H.; ROSEMANN M.; GABLE G.G.(2000), "What is ERP?", *Information System Frontiers*, vol.2, n.2:141-162. http://dx.doi.org/10.1023/A:102

6543906354

KWON O.; LEE O.; SHIN B. (2014): "Data quality management, data usage experience and acquisition intention of big data analytics", *International Journal of Information Management*, vol. 34, n. 3: 387-394, http://dx.doi.org/10.1016/j.ij infomgt.2014.02.002

LIN Y.; COLE C.; DALKIR K. (2014): "The relationship between perceived value and information source use during KM strategic decision-making: a study of 17 Chinese business managers", *Information Processing and Management*, vol. 50:156-174. http://dx.doi.org/10.1016/j.ipm.2013.07.006

LINDBLOM C.E (1959): "The Science of "Muddling through"", *Public Administration Review*, vol. 19: 79-88. http://dx.doi.org/10.2307/973677

LUNA-REYES L.F.; ANDERSEN D.L.(2003): "Collecting and analysing qualitative data for system dynamics: methods and models", *System Dynamics Review*, vol.19, n.4:271-296. http://dx.doi.org/10.1002/sdr.280

MADDEN S.(2012), "From Databases to Big Data", *IEE Internet Computing*, vol. 16, n.3: 4-6. http://dx.doi.org/10.1109/MIC.2012.50

MAYER-SCHÖNBERGER V.; CUKIER K. (2013): "*Big Data: A Revolution That Will Transform How We Live, Work, and Think*", Houghton Mifflin Harcourt, New York, NY

MCAFEE A.; BRYNJOLFSSON E. (2012): "Big Data: The Management Revolution", *Harvard Business Review*, vol.10: 1-9.

MILLER G.H.; MORK P.(2013): "From Data to Decisions: a value chain for Big Data", *IT Professional*, vol.15, n.1:57-59. http://dx.doi.org/10.1109/MITP.2013.11

MISHNE G.A. (2007), *Applied Text Analytics for Blogs.*

MUNRON D.(1977): "Determining Management Information Needs: a Comparison of Methods", *MIS Quarterly*, vol.1, n.2: 55-67. http://dx.doi.org/10.2307/249168

NEGASH S.(2004): "Business Intelligence", *Communications of the Association for Information Systems*, vol.13: 177-195.

PROVOST F.; FAWCETT T. (2013), "Data Science and its Relationship to Big Data and Data-Driven Decision Making", *Big Data*, vol.1, n. 1:51-59. http://dx.doi.org/10.1089/big.2013.1508

RIEMER, K.; OVERFELD, P.; SCIFLEET, P.; RICHTER, A. (2012): "Oh, SNEP! The Dynamics of Social Network Emergence-the case of Capgemini Yammer". http://goo.gl/UJMozt

RUSSOM P. (2011): "Big Data Analytics", *TDWI Best Practices Report*, vol. 4.

SAGIROGLU, S.; SINANC, D. (2013): "*Big data: A review*" Collaboration Technologies and Systems (CTS), 2013 International Conference: 42-47, http://dx.doi.org/10.1109/CTS.2013.6567202

SHAH S.; HORNE A.; CAPPELLÀ J.(2012): "Good Data Won't Guarantee Good Decisions", *Harvard Business Review*, vol.4:1-4.

SIMON H.A. (1960): "*The New Science of Management Decision*", Harper & Row, New York. http://dx.doi.org/10.1037/13978-000

SIMON H.A. (1969): "*The Sciences of the Artificial*", MIT press, Cambridge, Mass, 1st edition.

SIMON H.A. (1979): "Rational Decision Making in business Organisations", *The American economic review*, vol. 69, n.4, 343-371.

STERMAN JD. (2000): "*Business Dynamics: Systems Thinking and Modelling for a Complex Word*", Irvin/McGraw-Hill, Boston.

SWALES, J. (1990). Genre analysis: English in academic and research settings. Cambridge University Press.

THUSOO A.; SHAO Z.; ANTHONY S. *et al.* (2010): "*Data Warehousing and Analytics Infrastructure at Facebook*", SIGMOD'10, June 6–10, Indianapolis, Indiana, USA. http://dx.doi.org/10.1145/1807167.1807278

VROOM V. H.; JAGO A. G.(1974), "Leadership and Decision Making", *Decision Science*, vol.5: 734-755.

WATSON H.J.; WIXOM B.H.(2007): "The current state of Business Intelligence", *Computer*, vol.40,n. 9:96-99. http://dx.doi.org/10.1109/MC.2007.331

WLADAWSKY-BERGER I.(2013): "Data-Driven Decision Making Promises and Limits", *CIO Journal*.

WU X.; ZHU X.; WU G.Q.; DING W. (2014): "Data Mining with Big Data", *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, n.1: 97-107. http://dx.doi.org/10.1109/TKDE.2013.109

WALLER, M. A. and FAWCETT, S. E. (2013): "Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management". *Journal of Business Logistics*, vol.34:77–84. http://dx.doi.org/10.1111/jbl.1201

XIANG, Z.; SCHWARTZ, Z.; GERDES, J. H.; UYSAL, M. (2015): "What can big data and text analytics tell us about hotel guest experience and satisfaction?", *International Journal of Hospitality Management*, vol. 44: 120-130. http://dx.doi.org/10.1016/j.ijhm.2014.10.013