Framework for the Implementation of a Big Data Ecosystem in Organizations

PH.D. THESIS

Submitted by:

SERGIO ORENGA ROGLÁ

Supervised by:

DR. RICARDO CHALMETA ROSALEÑ

Castellón de la Plana, June 2017

Acknowledgements

I wish to express my sincere appreciation to those who have contributed to this thesis and supported me in one way or the other during this doctoral journey.

First of all, I am extremely grateful to my advisor, Professor Dr. Ricardo Chalmeta, for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. And especially for relying on me even in moments when it seemed that nothing made sense. His deep insights helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D. study.

I extend my sincere gratitude to my wife Mónica, not only for her unconditional support, both emotionally and technically, that she has given me during all these years, but also for her patience because I have not been able to spend time with her, as much as I wanted. She has been a constant source of strength and motivation for me, and she has always been my support in the moments when there was no one to answer my queries. All this would not have been possible without her.

I would also like to take this opportunity to thank the companies and professionals who have collaborated with me to make this project come true. Moreover, I would like to thank my friends, who have been interested in this project and have given me moral support.

Last but not the least, I would like to thank all my family and in-laws for supporting me spiritually throughout this project and my life in general. And special thanks to my parents for all the sacrifices that they have made on my behalf.

Contents

List of Figures

List of Tables

Notations

ARDIN	Spanish acronym of Reference Architecture for INtegrated Development
BI	Business Intelligence
BPD	Business Process Diagram
BPMN	Business Process Model and Notation
CRM	Customer Relationship Management
ERP	Enterprise Resource Planning
GPS	Global Positioning System
KM	Knowledge Management
KMS	Knowledge Management Systems
KMSTD	Knowledge Management System Traditional Development
KPMT	Knowledge Project Management Team
MGI	McKinsey Global Institute
NLP	Natural Language Processing
NoSQL	Not Only Structured Query Language
ROI	Return On Investment

RSS	Really Simple Syndication
SCM	Supply Chain Management
SME	Small and Medium-sized Enterprise
SNA	Social Network Analysis
Social CRM	Social Customer Relationship Management
SQL	Structured Query Language
SWOT	Strengths, Weaknesses, Opportunities and Threats
UML	Unified Modeling Language
UNECE	United Nations Economic Commission for Europe
W2KM	Web 2.0 Knowledge Management

Chapter 1. Introduction

This chapter presents the research topic discussed in this Ph.D. thesis. Initially, it shows a brief description of the problem statement of the research carried out, and the main objectives established. Then, the research methodology used is indicated. Finally, it explains how the thesis is organized, and provides a brief description of the chapters that compose it.

1.1. Problem Statement

Nowadays we are living in the information age, where large amounts of data are generated every second. This is due, among other things, to the huge quantity of different processes and devices that generate data, such as sensors, log files, mobile devices, transaction records, etc., and the high speed with which data are created. Because of the number and characteristics of such data (they are unstructured, complex, multi-format, multi-channel, and so on), traditional computer systems cannot commonly manage them properly (Elgendy and Elragal, 2014; Syed et al., 2013). All this, together with the recent decrease in computational and storage costs (Tekiner and Keane, 2013), have encouraged the emergence of Big Data technology. This technology is able to analyze and cross-reference those amounts and types of data generating useful knowledge (Almeida and Bernardino, 2015).

Big Data technologies allow organizations to extract value and insight from data generated both inside and outside the organization (Assunção et al., 2015). In this way, Big Data is an organizational competitiveness tool that allows them to discover business insight with the objective of increasing business performance, and thus, gaining a competitive advantage over their competitors. In addition, Big Data can be complemented by Web 2.0 technologies, which can generate a huge amount of valuable data, and also enable companies to have more effective communication, promote collaboration, and facilitate social interaction and knowledge sharing (Kirchner et al., 2009; Wu et al., 2013).

Developing and implementing a Big Data ecosystem in an organization is a very complex task that encompasses not only the technological aspects, but also the management of policies and people (Tekiner and Keane, 2013). In addition, the implementation of Big Data and Web 2.0 systems in organizations involves the coordination and collaboration of different users, as well as the execution and synchronization of many activities and tasks. Organizations could be based on frameworks in order to carry out such implementation with more guarantee of success.

A framework describes concepts, aspects, features, processes, data flows, or relationships between components, for certain domains (such as software development), with the aim of creating a better understanding (as the description of components or design aspects) or

guiding the achievement of a specific objective (Pawlowski and Bick, 2012). Frameworks are composed by several dimensions, usually interrelated. In the literature, there exist several frameworks to manage Big Data ecosystems (Das and Kumar, 2013; Demchenko et al., 2014; Ferguson, 2012; Géczy, 2015; Miller and Mork, 2013; Sun and Heller, 2012; Tekiner and Keane, 2013).

Nevertheless, in their present state there are some issues that limit the effectiveness of the existing frameworks to develop and implement a Big Data ecosystem. The main proposed frameworks are based primarily on data, i.e., in operations with them, in how they are generated, in their characteristics, in the use made of them, and in the purpose of the operations performed on it. However, they do not consider other aspects that are also very important, such as: they do not contemplate all aspects that influence in Big Data ecosystems; they do not have a methodology to guide the necessary steps to follow in the development and implementation process of Big Data ecosystems, which makes this process more difficult; they have not proved their validity, because they do not provide strong case studies in which they are evaluated; or they do not consider human, organizational and business process impact of the Big Data implementation.

In this context, an important research topic is the development of a framework that will help in the complex process of incorporating Big Data ecosystems into organizations. The work done in this thesis tries to facilitate the development and implementation of Big Data ecosystems in organizations, guiding them in detail in all the necessary steps to be carried out and considering all the aspects to be taken into account in each phase in order to obtain a satisfactory result.

1.2. Objectives

The main objective proposed in this thesis (Objective 1) is to develop a generic framework that allows organizations to use Big Data and Web 2.0 technologies, minimizing the invested time and effort, and considering all aspects related to the implementation, both technical, political, cultural, behavioral, etc.

In order to improve and validate it, the generic framework is applied in two case studies to different areas of the organization. On the one hand, it applies to the customer relationship management in an SME (Small and Medium-sized Enterprise) from the metal sector (Objective 2), and on the other hand, to knowledge management in a large oil and gas company (Objective 3).

1.3. Research Methodology

In this Ph.D. thesis different research methods have been used: exploratory research, which structures and identifies new problems (Stebbins, 2001); constructive research, which develops solutions to problems (Lukka, 2003); and empirical research, which tests the feasibility of a solution using empirical evidence (Wohlin et al., 2006).

For the development of the initial version of the framework, exploratory research was done on the existing literature related to the development and implementation frameworks of Web 2.0 and Big Data systems. As a result of this exploratory research, existing solutions were found and analyzed, and different weaknesses and deficiencies, as well as possible improvements, were detected.

With the base obtained from the review of the literature and with the previous experience of the Ph.D. student, the first version of the framework was constructed. That version took advantage of the strengths of existing solutions, improving them as much as possible and solving or minimizing the deficiencies and weaknesses found.

That first version was applied in real environments using the methodology of the case study. The work plan that was followed in order to carry out the case studies was based on Runeson and Höst (2009) and consists of five phases: Case study design and planning; Preparation for data collection; Collecting evidence; Analysis of collected data; and Validation of collected data.

The obtained results were used to debug, improve and validate the framework, as well as to create two specializations of it, one of them for the management of the relationships with the customers and the other one for the management of the business knowledge.

1.4. Structure

This thesis is organized in five chapters, which are summarized below.

In Chapter 1 an introduction to the subject of this thesis is presented. Chapters 2 to 4 are three complete and independent research articles, but they deal with related topics, covering the three objectives proposed in this thesis. Thus, Chapter 2 corresponds to Objective 1, Chapter 3 to Objective 2, and Chapter 4 to Objective 3. Each one of these three chapters has its own complete structure, and they can be read and understood independently. Finally, Chapter 5 presents the conclusions obtained from the work of this thesis.

Chapter 1: Introduction. This chapter presents a brief general description of the research topic analyzed and presented in this thesis. It indicates the problem statement, the research methodology used, the proposed objectives, and the results obtained in this work. It also explains how the document is organized.

Chapter 2: Framework for a Big Data ecosystem implementation in organizations. This chapter presents a general framework for implementing and managing Big Data ecosystems in organizations. First, a study of the most significant frameworks of this type in the literature is carried out, and their strengths and weaknesses are analyzed. Then, the proposed framework is presented, which takes advantage of the strengths and avoids the weaknesses identified in the existing frameworks studied.

The proposed framework is composed of seven interrelated dimensions, which are: Methodology, Data architecture, Organizational, Data sources, Data quality, Support tools, and Privacy/Security. The Methodology dimension is the main one, it extends over the whole life of the project, and it provides a practical guide to follow the necessary steps to implement a system with Big Data technology. The other dimensions give support to the phases, activities and tasks that make up the methodology.

Chapter 3: Social customer relationship management: taking advantage of Web 2.0 and Big Data technologies. This chapter presents a specialization of the Methodology dimension proposed in the general framework that guides the development and implementation of Social Customer Relationship Management (Social CRM) systems in organizations, taking into account organizational, human and technological aspects. It is

grounded on the methodology of implementation of Customer Relationship Management (CRM) systems proposed by Chalmeta (2006), and is supplemented, adapted and updated based on the review of the existing literature on the subject, and the experience of the Ph.D. student. The methodology is composed of phases, which are formed by activities. The phases that compose it are the following: project management and prerequisites; organizational framework; customer strategy; system for assessing customer relations; process map; human resources; computer system; implantation; and monitoring.

In the case study presented in this chapter, the initial version of the methodology was successfully applied to an SME company from the metal sector with the aim of analyzing, validating and refining it. The results and feedback obtained in this case study were used to improve both the methodology and the generic framework.

Chapter 4: Methodology for the implementation of knowledge management systems 2.0: A case study in an oil and gas company. This chapter presents a specialization of the methodology dimension proposed in the general framework to guide the implementation and development of Knowledge Management Systems (KMS) 2.0 in organizations using Big Data and Web 2.0 tools. The methodology is composed of seven phases: Draft; Planning; Analysis; Design; Development; Implementation; and Control. Each one of these phases is composed of activities, which in turn contain tasks. The methodology guides the processes of collection, generation, management, and application of the knowledge generated both internally and externally to the organization. The application of the methodology reduces the complexity of the project and increases its chances of success.

The proposed methodology was applied to a large oil and gas company in a case study, in which it was analyzed, debugged, and validated. It details all the steps that were followed, along with the users involved and the results obtained. The comments, proposals and feedback collected in the case study were used to refine and improve the methodology and the generic framework.

Chapter 5: Conclusions. This chapter explains the general conclusions and the main contributions of the thesis. Limitations of the present work and possible directions for further research are also indicated. Finally, the publications associated with this Ph.D. thesis are included.

1.5. References

Almeida P, Bernardino J (2015). A comprehensive overview of open source big data platforms and frameworks. International Journal of Big Data, 2(3):15-33.

Assunção MD, Calheiros RN, Bianchi S, Netto M, Buyya R (2015). Big data computing and clouds: trends and future directions. Journal of Parallel and Distributed Computing, 79-80(2015):3-15.

Chalmeta R (2006). Methodology for customer relationship management. The Journal of Systems and Software, 79(7):1015-1024.

Das TK, Kumar PM (2013). BIG Data Analytics: A Framework for Unstructured Data Analysis. International Journal of Engineering and Technology, 5(1):153–156.

Demchenko Y, de Laat C, Membrey P (2014). Defining Architecture Components of the Big Data Ecosystem. International Conference on Collaboration Technologies and Systems (CTS), pp.104-112.

Elgendy N, Elragal A (2014). Big Data Analytics: A Literature Review Paper. In Advances in Data Mining. Applications and Theoretical Aspects, 8557(2014):214-227.

Ferguson M (2012). Architecting a Big Data Platform for Analytics. IBM White Paper, October 2012. Available at: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IML14333USEN.

Gèczy P (2015). Big Data Management: Relational Framework. Review of Business & Finance Studies, 6(3):21-30.

Kirchner K, Razmerita L, Nabeth, T (2009). Personal and Collective Knowledge Management in the Web 2.0: Two Faces of Knowledge Management. Paper presented at the 9th International Conference on Innovative Internet Community Systems, Jena, Germany, 15-26.

Lukka K (2003). The constructive research approach. In Case study research in logistics. Publications of the Turku School of Economics and Business Administration, Series B, 1(2003):83-101.

Miller HG, Mork P (2013). From Data to Decisions: A Value Chain for Big Data. IT Professional, 15(1):57–59.

Pawlowski J, Bick M (2012). The Global Knowledge Management Framework: Towards a Theory for Knowledge Management in Globally Distributed Settings. Electronic Journal of Knowledge Management, 10(1):92-108.

Runeson P, Höst M (2009). Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering, 14(2):131–164.

Stebbins RA (2001). Exploratory Research in the Social Sciences. Sage Publications, Inc., Thousand Oaks, CA.

Sun H, Heller P (2012). Oracle Information Architecture: An Architect's Guide to Big Data. Oracle White Paper, August 2012. Available at: https://d2jt48ltdp5cjc.cloudfront.net/uploads/test1_3021.pdf.

Syed A, Gillela K, Venugopal C (2013). The Future Revolution on Big Data. International Journal of Advanced Research in Computer and Communication Engineering, 2(6):2446-2451.

Tekiner F, Keane JA (2013). Big Data Framework. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE Computer Society, 1494-1499.

Wohlin C, Höst M, Henningsson K (2006). Empirical Research Methods in Web and Software Engineering. In Web Engineering, p. 409-430. Springer, Berlin, Heidelberg.

Wu J, Sun H, Tan Y (2013). Social media research: A review. Journal of Systems Science and Systems Engineering, 22(3):257-282.

Chapter 2. Framework for the Implementation of a Big Data Ecosystem in Organizations

Chapter under review…

Chapter 3. Social Customer Relationship Management: Taking Advantage of Web 2.0 and Big Data Technologies

The emergence of Web 2.0 and Big Data technologies has allowed a new customer relationship strategy based on interactivity and collaboration called Social CRM to be created. This enhances customer engagement and satisfaction.

The implementation of Social CRM is a complex task that involves different organizational, human and technological aspects. However, there is a lack of methodologies to assist companies in these processes.

This chapter shows a novel methodology that helps companies to implement Social CRM, taking into account different aspects such as social customer strategy, the Social CRM performance measurement system, the Social CRM business processes, or the Social CRM computer system. The methodology was applied to one company in order to validate and refine it.

3.1. Introduction

The view that companies have of a customer has evolved dramatically with increasing competition, market globalization and technological advances. Prior to the 20th century companies focused only on production. During the first half of the 20th century, however, companies began competing to persuade customers to buy their products. Later, during the second half of the same century, companies realized that they did not have to sell customers what they manufactured, but had to make what was demanded in market sectors. At the beginning of the 21st century, a stage began where the prevailing business orientation focused on customer relationship management (CRM), where each customer was treated individually and uniquely, depending on their preferences (Bose, 2002).

The essence of CRM is to change the strategy of organizations, to move from a product-focused strategy to a customer-focused one. The aim is to create value for customers, understand their needs and offer value-added services (King and Burgess, 2008). This increases the value of the company and allows it to gain a strategic advantage over competitors, because customers are more satisfied and, therefore, it is easier to retain them (Nguyen et al., 2007).

CRM has been conceptualized from five different viewpoints: (1) Process, (2) Strategy, (3) Philosophy, (4) Capability, and/or (5) Technological tools (Zablah et al., 2004). Therefore, there is no agreed definition of CRM. Among the most representative definitions is that provided by Ranjit Bose (2002) "*At the core, CRM is an integration of technologies and business processes used to satisfy the needs of a customer during any given interaction. More specifically, CRM involves acquisition, analysis and use of knowledge about customers to sell more goods or services and to do it more efficiently*". In this context, the work of Michael Fayerman (2002) should also be mentioned. This author distinguishes the following three areas of CRM within the company: (a) Operational CRM, which deals with actual interactions with customers; (b) Analytical CRM, which analyses data about a company's customers and presents them in such a way as allow better and quicker business decisions to be made; and (c) Collaborative CRM, which promotes inter-departmental teamwork and communication within a company for the purpose of improving overall customer experience.

CRM is not just technology, as a proper implementation of CRM requires an integrated and balanced approach to technology, process and people (Chen and Popovich, 2003). However, it has been the evolution of information technology and communications which has allowed its implementation. CRM links the systems of *front office* (sales, marketing, and customer service) and *back office* (financial, logistics, warehousing, accounting, human resources, and so forth) through the integration of technological and functional components (Strauss and Frost, 2002; Zamil, 2011). CRM mainly relies on the use of structured data from a data warehouse, where data are extracted, transformed and loaded from operation systems such as ERP, SCM or operational CRM.

In this context, the emergence of Web 2.0 technologies has allowed the evolution of CRM, which is based on a strategy focused on customer transactions, to Social CRM, which is based on a strategy focused on creating engagement between the customer and the company (offering new points-of-contact between the two, not only with the salesperson, and building stronger customer links with the company) (Faase et al., 2011). Yet, Social CRM does not replace CRM, but complements it. Customer engagement using Web 2.0 technologies is only possible when there is already a customer management using CRM (Faase et al., 2011). What Social CRM adds are social features, functions, processes and different forms of interaction between the company and its customers, suppliers and partners (Greenberg, 2010).

The social customer is the customer that makes use of social software, which moves in a scenario characterized by permanent connectivity, mobility, being multi-channel and the progress of the Internet of Things. The publication of opinions on the Internet allows customers to share their points of view about a product or service. Companies participate in the social network of users connecting to its target group. This facilitates the opportunity to gain business-relevant insights from the data accessible from the communication among users. These electronic word-of-mouth statements are very important for organizations, because it is a way (1) to know how customers perceive their products and/or services, (2) to intensify the relationship, and (3) to align the business with consumer needs (Rosenberger, 2015). This scenario is a new model of interaction between people, which is being transferred virally to the relationship between customer and company. Unlike other technological revolutions, this change in the way they relate is not being led by the companies, but the customers and their expectation that the companies with which they relate operate a significant change in their access models and behave in accordance with this new social reality. This new model represents a business opportunity for companies in

their customer management. As a workspace, it represents a challenge for companies, because it is necessary to manage both human information, which is characterized by being complex, unstructured, ubiquitous, multi-format and multi-channel, and also the traditional information.

Moreover, Social CRM benefits from Big Data because it facilitates more accurate decision-making and a more efficient distribution of knowledge among the social customers and the company (Anshari et al., 2015). Big Data technology can be used for many purposes in Social CRM. Some of these include: (1) *Commercial recommendation*, suggesting the product or service with greater probability of success for each customer; (2) *Competitive intelligence*, showing real-time automated information customized to the situation created by the customer, thereby allowing the company to maintain a contextualized dialogue and to obtain real-time information needed to make suitable decisions; (3) *Debt recovery of customers* from public information sources; (4) *Automated categorization and routing of customer interactions* over any channel; (5) *Predictive models* of trend (purchase, abandonment, non-payment, etc.) and clustering of customers; and (6) *360° customer view*, showing the relevant customer information performed through any channel and format.

The amount of open information available online from heterogeneous sources and domains is growing very quickly, and constitutes an important body of knowledge to support Social CRM. These data sources may disclose significant business opportunities and competitive advantage to those who are able to understand and leverage their value (Torre-Bastida et al., 2015). They can infer valuable information as a support for customer-related decision-making. Therefore, Big Data and Web 2.0 technologies could change Social CRM from an unavoidable tool to keep/gain the new segment of "social customers" into a business opportunity and a competitive advantage.

However, the development and implementation of Social CRM in a company is a complex task that involves different organizational, human and technological aspects (Bebensee et al., 2011; Rosenberger, 2015). In order to assist in a process of this kind, a methodology for managing the innovation and change involved in Social CRM is needed, while also reducing both the risk of failure in the implementation, as well as the time required to obtain business benefits (Crockett and Reed, 2003; Lech, 2016; Nguyen et al., 2007).

The literature about Social CRM does not contain any specific methodologies to help in the development of Social CRM. The research that exists on the topic of Social CRM focuses primarily on the characteristics, opportunities and benefits that Social CRM offers (Faase et al., 2011; Greenberg, 2010; Mosadegh and Behboudi, 2011) but does not offer any methodological guidelines.

To help solve this problem, a methodology called SCRM-IRIS is presented in this chapter, which guides the development and implementation of Social CRM in a company. The rest of the chapter is organized as follows. Section 3.2 shows the literature review. In Section 3.3 the research method that was followed to obtain the methodology is shown. Section 3.4 presents the methodology proposed for the implementation of Social CRM. Finally, Section 3.5 presents the conclusions.

3.2. Literature Review

3.2.1. Social Software

There is no generally agreed definition of Web 2.0. One of the most widely used is that proposed by Tim O'Reilly, who defines it as "a set of economic, social, and technology trends that collectively form the basis for the next generation of the Internet – a more mature, distinctive medium characterized by user participation, openness, and network effects" (Musser et al., 2006, p. 4). Web 2.0 is not only a new generation of technologies, but also a change in the way in which users access the Internet in order to mutually interact and collectively create knowledge. Some of the most common Web 2.0 tools include: Wikis, Group chats, Social bookmarking, Mashups, Blogs, RSS, Folksonomy, Podcasts and Social Networks.

Social software (which consists of the applications created with Web 2.0 technologies for social purposes) enables the development of new communication tools that allow a competitive advantage to be created in organizations (Wirtz et al., 2013). In social media (which is the set of social software applications), users can find not only information, but are active contributors (Lai and To, 2015; Razmerita et al., 2009) and can freely express their comments, views and emotions (Feng et al., 2011). Therefore, social media encourage the creation, sharing and exchange of data. As stated above, there is a large variety of types of social software applications, such as Social Networks (which allow social capital to be managed more efficiently), Blogs (to communicate with others more effectively), Wikis and Social bookmarking (to make better use of collective intelligence), Group chats, Mashups, Multimedia Sharing, RSS, Folksonomy or Podcasts. These technologies are open and are designed to encourage collaboration as well as to facilitate social interaction and the sharing of knowledge (Dietrich et al., 2008; Kirchner et al., 2009; Ras and Rech, 2009).

Social software only provides the framework, the content is provided by people (Omerzel, 2010). Furthermore, the number of people using social software is very important. As more people use these applications, the overall value of knowledge will be significantly increased, i.e. collective intelligence will increase (Shimazu and Koike, 2007). Users provide data and services in a way that allows others to combine them again, thus creating a network of effects through the "architecture of participation" (O'Reilly, 2005). Recently, the social media have become a strategic tool for organizations, since they allow companies to meet the needs of customers as well as to provide them with new services (Go and You, 2016).

3.2.2. Social CRM

CRM can take advantage of social media, whose relational properties and characteristics are particularly well suited to customer interactions (Olbrich and Holsing, 2012). Social CRM can be defined as "*A philosophy and a business strategy, supported by a technology platform, business rules, processes and social characteristics, designed to engage the customer in a collaborative conversation in order to provide mutually beneficial value in a trusted and transparent business environment. It's the company's response to the customer's ownership of the conversation"* (Greenberg, 2009). This definition includes the central principle of *customer engagement*, which was missing in earlier CRM models, and social media technologies facilitate this customer engagement (Olbrich and Holsing, 2012).

Therefore, for a Social CRM system to work, there must be an important cultural and behavioral change both in the company as well as in the customers, as they have to change the way in which they interact (Greenberg, 2009). Contribution, sharing, collaboration, dynamism and bidirectional trust between the company and customers become fundamental aspects in Social CRM (Lee and Lan, 2007). The concept of social customer thus appears, which can be defined as a new type of customer that uses social software to search for, compare and exchange views on products and services offered by a company, and who expects companies to not only be present in that social software but also to respond to questions and participate. This customer acquires knowledge about new products and services through social channels and networks, prefers a conversation with the particular brand rather than it being just a way to send messages and at the same time wait for an answer, and wants the company to listen to and solve their problems quickly.

The social customer creates a new business model, called social business, which can help companies increase their profitability because it allows a number of qualitative and quantitative benefits to be obtained. The qualitative ones include: a better understanding and interpretation of the market, by interacting with customers in real time; benefiting from word of mouth; involving and engaging the customer at all stages of development of the product or service offered by the company (design, production, testing, etc.); improving the overall customer experience and lifetime value; enhancing products and services, or building up trust (Faase et al., 2011; Mosadegh and Behboudi, 2011; Reinhold and Alt, 2011; Sarner et al., 2012). Some the most significant quantitative benefits that could be achieved with the use of Social CRM are: increased sales; decreased service costs; reduced or replaced direct costs of printing and online advertising; reduced direct staff time costs; increased direct revenue from memberships, registers and advertising, exhibitions and sponsorship (Dreyer and Grant, 2011b).

However, this new business model has a number of risks for companies (Assaad and Marx Gómez, 2011). Both good and bad news spread quickly; social software is not well controlled or censored, so anyone can publish anything good or bad about the company or its products or services; and problems regarding personal privacy and security can emerge as the user is required to share at least some personal data.

3.2.3. Big Data

Moreover, Social CRM benefits from Big Data, which is based on the current ability to have a large amount of data and draw conclusions about all sorts of company-customer processes and interactions. The digital world, mobility, and permanent connectivity have completely changed these processes and interactions over the last two decades. In addition, advances in infrastructure, storage techniques, and data-processing allow these huge volumes of structured and unstructured customer data to be analyzed in a very fast and efficient way, and with an acceptable cost for most organizations. Due to the amount and complexity of these data, it is difficult to process them using traditional tools, so the use of Big Data technology is essential in order to take advantage of this kind of data (Syed et al., 2013).

Big Data technology is able to overcome the difficulties involved in understanding and extracting relevant knowledge from different kinds of data, which include: Diversity in types of fonts, formats and languages; Unstructured information (ideas, emotions, nuances, ambiguities, polysemy, etc.) that is contextual and has complex and fuzzy relations, such as distance, overlap, correlation, similarity, opposition, etc.; Dependence on the context in

which it is emitted; Semantic problems due to the fact that language is constantly changing; and Dependence on grammar, language and the medium used.

A more subtle aspect of Big Data which is not frequently mentioned is that the analysis of massive data, which are often incomplete and even slightly inaccurate, seeks to find correlations and detect "things that are happening", largely ignoring the analysis of causality. The emphasis is on the “what”, not the “why”. However, the growing analytical arsenal and existing advanced modeling techniques applied to massive datasets by professionals with appropriate levels of creativity and expertise are currently reaching an enormous degree of success in discovering correlations in previously unknown customer knowledge.

3.2.4. Big Data in Social CRM

Big Data is a technology with a real ability to transform very significant aspects of customer relationship management, thereby providing companies with a competitive advantage over its competitors. Big Data technology allows knowledge to be extracted from customer information and converted, in an effective, secure and scalable way, into real business value. From customer information and through Big Data, a company is able to reveal hidden knowledge of the customer, turning it into opportunities to maximize the business value of each customer, to act preventively, to improve customer satisfaction, to identify new opportunities, or to predict their tendency and intention profile.

It is noteworthy that companies are harnessing the power of Big Data and analytics to apply it in customer relationship management (Marshall et al., 2015). The business value is derived from the knowledge generated, once it is transferred to the design of products or services, to the segmentation of customers and markets, to the acquisition of new customers, to the understanding of customers, to the evolution of the portfolio, to the optimization of any of the internal procedures and production processes, or to the changing way companies relate with employees, citizens, suppliers, partners or customers.

3.3. Research Methodology

Since Social CRM complements CRM, in order to obtain a Social CRM methodology, called the SCRM-IRIS methodology, that guides the development and implementation of Social CRM in a company, an initial version was first developed based on the CRM implementation methodology presented by Chalmeta (2006). This methodology was supplemented, adapted and updated based on the review of the existing literature on Web 2.0, Big Data, CRM and Social CRM, as well as on the experience of the authors. This initial version was then applied to one company with the aim of analyzing, validating and refining it. In order to carry out the application, a work plan based on the case study methodology proposed by Runeson and Höst (2009) was followed. This consists of the following stages: Design and planning of the case study; Preparation for data collection; Collecting evidence; Analysis of collected data; and Validation of collected data. Each of them is described below:

3.3.1. Design and Planning of the Case Study

The aims of the case study are: (a) to validate the SCRM-IRIS methodology by verifying and confirming its usefulness, accuracy and quality, and (b) to refine and improve the methodology developed initially from the experience acquired by the researchers, the

feedback obtained from the company involved, and the conclusions drawn in the case study.

The research work was conducted over a period of 10 months. The first task was to select the company in which the case study was to be applied. The criteria underlying the selection of this company were essentially: (1) a willingness to collaborate in the research, and (2) the fact that the management of this company was considering the idea of improving the efficiency of their customer relationship management using Web 2.0 and Big Data technologies. The selected company was an SME from the metal sector with a workforce of 250 employees. Their target customer ranges from large supermarkets to little grocery stores and individuals, from all over the world. It is important to note that this company was already using a traditional CRM application.

3.3.2. Preparation for Data Collection

To begin the research work, an introductory series of group interviews were held in the company. The presentation focused on the basic points of a Social CRM project and, at the same time, the methodology that was going to be used (initial version of the SCRM-IRIS methodology) was also explained to them.

In order to undertake all the research tasks during the application of the methodology in the company, a mixed work team was set up with members that came from both the IRIS Research Group and the Social CRM team of the company. The company Social CRM project team was made up of five area managers, representing the main areas of the company: General management, Commercial management, Financial management, Technical management and Operations management.

3.3.3. Collecting Evidence

The data collected were the results of applying the different stages of the initial version of the SCRM-IRIS methodology to the company. Qualitative data were used, which were collected through direct methods (using an assortment of questionnaires and templates) and independent methods (copies of the documents and reports used in the company).

The questionnaires were answered by IRIS researchers during individual interviews with Social CRM project team members. Once the implementation of each of the nine phases that compose the SCRM-IRIS methodology had finished, the IRIS researchers interviewed the five area managers from the Social CRM project team on an individual basis. These interviews lasted approximately 20 minutes and were open (thus allowing interviewees to give a wide range of answers) and semi-structured (the questions were planned only as a guide, not to be asked in that same order, thereby allowing both the interviewers and the interviewees to improvise). The aims of these interviews were: to analyze the execution of the phase, to obtain feedback about the experience and the observations of each manager in each phase, to detect errors, and to collect proposals for improvement to the SCRM-IRIS methodology from each of them. There was a different questionnaire for each phase, and those questionnaires were common to all interviewees. Table 3.1 presents an example of the questionnaire followed by the IRIS researchers to conduct the interviews after the process map phase.

Table 3.1. Interview questionnaire for the process map phase.

Interview questionnaire **Phase: Process map**	
1	What new business processes have been created in your department at this phase?
2	What existing business processes have been improved in your department at this phase?
3	Are there any business processes that have not been considered at this phase?
4	What information does Social CRM offer that you did not have before?
5	How do you think the company benefits from new or modified business processes?
6	Has the company assigned the necessary resources for a successful implementation of this phase?
7	Has the researcher provided the necessary means for a successful implementation of this phase?
8	Have you missed the collaboration of someone or something in the implementation of this phase?
9	What problems have you noticed in the implementation of this phase?
10	What would you change or improve in the implementation of this phase? How would you do it?
11	Is there any information that has not been considered at this phase and that you think should have been taken into account?
12	Is it worth the effort invested in the implementation of this phase in view of the expected result?
13	Have the desired results been achieved in the estimated time?
14	What is your general opinion about the implementation of this phase?

Most comments obtained were positive, indicating that the SCRM-IRIS methodology guided them perfectly throughout the implementation of all the phases and made them consider things that had not been proposed so far, such as for example a strategic focus on social customer engagement, the social customer profile that they should lead, and they had to take into account the average age of users. They were also very surprised by the amount of information that could be obtained about social customers. Moreover, negative comments were taken into consideration to improve the methodology, such as the lack of a company social media policy and guidance on how the employees had to use social software, besides training them in legal issues, and the need for different levels of segmentation based on communities and sub-communities. In addition, some negative comments said that once Social CRM was implemented, all possible tasks related to its use should be carried out by low-level staff, as their labor costs were lower, but they must be properly trained and high-level staff must support them when they needed it.

Once the project had finished, meetings were held with the Social CRM project team in the company in order to enrich the initial SCRM-IRIS methodology by modifying/incorporating/removing phases, tasks, tools, and so on. After this process of revising the initial methodology, it was enhanced by incorporating all the contributions detected and then validated with the general agreement of the Social CRM project team.

3.3.4. Analysis of Data Collected

The persons responsible for the application of the SCRM-IRIS methodology in the company indicated that the use of this methodology has allowed them to have an excellent view of the needs, scope, consequences and opportunities of the project, as well as allowing them to implement Social CRM quickly and without any significant problems. They also indicated that this has enabled them to have greater control over the project, because all the steps to be performed in each stage of the implementation are clearly defined.

On the other hand, the following benefits have also been highlighted by the Social CRM team of the company as being the most important provided by Social CRM:

- Centralization of knowledge relating to the company's customers in an accessible (for both internal users and users outside the company) and easy-to-use system, allowing a constant flow of that knowledge.
- Quick compilation and dissemination of information relating to customers.
- Allows an exchange of customer portfolios between salespeople that is quick, easy and reliable, because the system centralizes all the knowledge about customers, including the historic features, preferences, movements, etc.
- It records all the history of queries and problems from customers with the solutions that were adopted. This history is available to answer queries or similar problems (for that client or others) more efficiently and requiring less time.
- Users can access the information they need at the time and place where they need it (even in real time). Furthermore, such information is always up to date.
- Decrease in the use of other communication channels (e.g. e-mail, phone, etc.), as the Social CRM enables more effective communication.
- Decrease in the work undertaken by the company's employees due to: (1) the simplicity, speed, centralization, efficiency and control provided by Social CRM, and (2) customers can manage different tasks on their own, for example, can track their orders, can make or change their orders, etc.
- It allows potential social customers to be found quickly, as well as the tracking of current social customers.
- It allows the company to know in real time what people think about the products and/or services offered by the company, or by their competitors.

After a year using Social CRM, a comparison of the value of some indicators with the value obtained a year before the implementation of Social CRM was performed. In this comparison, some significant increases can be observed due to the introduction of Social CRM in the company. The most significant increases are: New supermarket customers (3%); New grocery store customers (18%); New individual customers (27%); Customer loyalty (11%); Customer satisfaction (24%); Amount of sales (13%); Amount invoiced (10%); and Presence in new countries (33%).

3.3.5. Validation of the Data Collected

As the data collected was qualitative, it was analyzed using qualitative data methods of analysis. In this case, the analysis was inductive and was carried out parallel to the data collection, as it was performed after each of the stages that make up the SCRM-IRIS methodology had finished. The purpose of this was to be able to react quickly to the assessments encountered during the analysis of each stage and thus rectify each one of them and take advantage of these improvements in the following stages.

Any threats to the validity of the case study were reduced by using the Lincoln and Guba model (Robson, 2002), in which five strategies are proposed for use in data collection to tackle three types of threats to validity. The three types of threats considered were *reactivity* (the researcher's presence can affect the setup of the study), *researcher bias* (the researcher's preconceived ideas can affect the way the researcher asks questions or interprets answers) and *respondent bias* (the researcher's influence on the attitude of the people being studied) (Karlström and Runeson, 2006).

With regard to the five possible strategies, in the present case study they were considered in the following way in order to make the results valid: (1) *Prolonged involvement*: the researcher is familiar with the environment being studied (in this case study, the researchers and the company had already been collaborating on previous projects). (2) *Triangulation*: the application of several methods in the study of a single object. In this case study, four types were considered: *Spatial triangulation of data* (three sources of data were considered: observation, interviews and documentation); *Personal triangulation of data* (all the members of the company Social CRM project team were interviewed in order to obtain information from each of them); *Investigator triangulation* (the interviews were conducted by a researcher and reviewed by another researcher); and *Theoretical triangulation* (the different points of view of the members of the Social CRM project team were taken into account). (3) *Member checking*: obtaining feedback from the people who are interviewed (in the case study, after each interview, a report containing the relevant information from the interview was checked by each interviewee). (4) *Negative case analysis*: attempting to find another explanation that differs from the one initially assumed for the observed phenomenon (here, the two researchers were working separately (investigator triangulation)). (5) *Audit trail*: keeping a record of all the documentation of the project so as to be available in the future.

3.4. SCRM-IRIS Methodology

The methodology for the implementation of Social CRM presented in this chapter is based on the proposal by Chalmeta (2006) for CRM. This methodology, called CRM-IRIS, is organized in nine phases and helps during the process of developing and implementing a CRM System. It considers and integrates various aspects, such as defining a customer strategy, re-engineering customer-oriented business processes, human resources management, computer system, management of change or continuous improvement.

The SCRM-IRIS methodology does not replace the CRM-IRIS methodology, but instead complements it, in order to adapt it to the features of Web 2.0 and Big Data technologies. It has not been necessary to add or remove phases, modify their sequence of application or delete the previous activities inside each phase. However, new activities inside each phase and modifications to some of the previous activities of the CRM-IRIS methodology have been added. Figure 3.1 shows these additions and modifications.

3.4.1. Project Management and Pre-requisites

There are no substantial differences in the project management between a Social CRM Project and a CRM Project, both of which must be managed as an engineering project. However, in the Social CRM Project the following basic prerequisites for success should be taken into account:

- *Viability of the project.* Before starting the project there should be an analysis of whether Social CRM can be viable in the company, considering different aspects such as whether Social CRM is appropriate for the sector in which the company operates, the technological level of the company and its capacity to increase it, the ability of users to use social software applications, resistance to sharing knowledge and resistance to technological change especially in older employees and customers, etc. (Assaad and Marx Gómez, 2011), as Social CRM will change the way of working (Dreyer and Grant, 2011a).

- *Profit Estimation.* An estimate of both the quantitative and qualitative benefits expected to be achieved with the implementation of Social CRM must be carried out. When estimating benefits it is very important to think of not only the social objectives such as the number of fan pages and weekly tweets but to correlate Social CRM with the contribution of top business objectives. It is necessary to estimate return on investment (ROI), business value and budget justification for social projects before developing it.

- *Risk management.* A risk assessment of the Social CRM must be carried out during this stage of the project in order to determine what issues need attention. First, the risks are identified, such as: possible misuse of the social software, posting of negative viral messages, privacy management, security of information, publication of private or confidential information, publication of misleading or false information, posting of negative comments, etc. Afterwards, every risk is assessed in relation to the frequency with which they may occur compared to the potential negative effect (financial, security, image, privacy, etc.) if they happen. Finally, priorities are established to mitigate the risks, addressing first the most severe and frequent events (Dreyer and Grant, 2010). Once the risks have been understood, the methods to be used to manage them must be defined, such as for example (Dreyer and Grant, 2010): Defining policies for users; Monitoring the social website to find out what is being said; Educating users on legal issues such as copyright and anti-trust; Educating users on the principles of social media; or Updating insurance policies to cover social media work.

- *Assignment of Roles.* The role of Community Manager must be assigned to an employee (or several employees) of the company (the marketing manager is recommended). His/her duty is to manage, build and moderate existing communities in the company by committing to social customers, making them feel part of the company and motivating them to take action, both on their own platforms, as well as using other public social software. Depending on the company's structure, there may be several Community Managers located in different departments (usually in membership, communications or government relations) who collaborate with each other (Dreyer and Grant, 2011a). It is also necessary to consider whether other roles need to be created (such as content creator, data analyst, and so on) to cover certain duties and to work with the Community Manager on social media projects (Dreyer and Grant, 2011b).

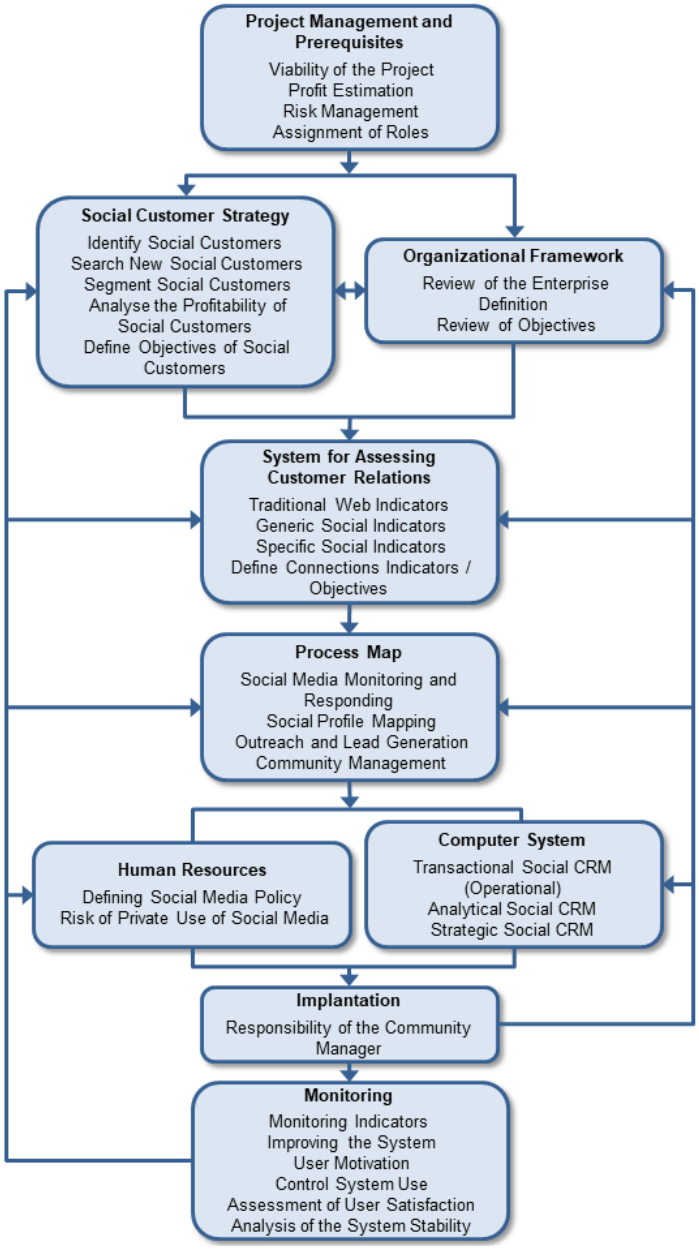

Figure 3.1. Additions and modifications to the CRM-IRIS methodology in order to consider Social CRM.

3.4.2. Organizational Framework

In this phase, the analysis of the company's objectives and culture should be performed taking into account the different characteristics of Social CRM:

- *Review of the enterprise definition.* The aim, vision, strategy, policy and values of the company should be reformulated, taking into account how the company sees and manages its customers to enhance the benefit of both parties. The following key aspects of Social CRM should be taken into account: (1) Corporate strategy has to consider social customer (Greenberg, 2010). (2) The relationship between business and the social customer focuses on a collaborative effort, and on social customer engagement and commitment, not social customer management (Greenberg, 2010; Mossadegh and Behboudi, 2011). (3) Customers are immune to

the complexity of the departments of the company, so that all departments should manage customers. (4) The customer is the one who sets the schedule, because if the company is not responding to him when he needs it, other companies will respond to him. (5) The customer already coexists with the multi-channel, and he expects the same from the company, that his information, the state of his products and services, or the processes in progress, are visible and can be interacted with from any channel. (6) The company must change its focus, from business transactions to managing customers and connected social communities with which they can maximize their business transactions and opportunities.

- *Review of objectives.* It is very important to define the general objectives that can benefit from the use of social software, as well as the strategy to be followed in order to achieve them, and the role that primary users (staff and key volunteers) should adopt (Dreyer and Grant, 2010). An example of an objective is to increase customer loyalty by making use of social software.

3.4.3. Customer Strategy

Social CRM is focused on social customer management, as well as on identifying, attracting and achieving new social customers (Greenberg, 2010). Social customers are those customers of the company that are managed using Social CRM. For them it is necessary to define the Social Customer Strategy, which is a part of the overall Customer Strategy.

Social customer strategy is defined by social customer engagement, not social customer management as in traditional CRM, which implies that there is a mutual benefit planned from the beginning (Mosadegh and Behboudi, 2011). In order to define the Social CRM strategy, the company should take the following steps:

3.4.3.1. Identify Social Customers

First, the social customers must be identified from the current customers with whom collaboration and relationships of commitment by Social CRM can be set up. It is important to consider that building relationships with inappropriate customers is one of the main reasons for failure in projects of systems related to the customer (Hu et al., 2013; Lambert, 2010). The identification of social customers can be accomplished by Big Data techniques of advanced analytics, since they help to discover trends, patterns and other insights, applied to historical information from past interactions with customers and public information on social networks.

3.4.3.2. Finding New Social Customers

One of the strengths of Social CRM is that it facilitates the search for new potential social customers that can contribute in the future to the enterprise in terms of branding, development or improvement of products or services, etc. In order to carry out this search it is very useful to perform social profile mapping, which is the process of collecting social data about people and their relationships, to know more about them and to analyze whether they or their contacts are potential possible customers for the company (Dreyer and Grant, 2011a). Big Data tools of advanced analytics can be used to listen to and learn from the social media activity and apply the insights to identify possible customers.

3.4.3.3. Segmenting the Social Customers

The traditional methods of customer segmentation are based on characteristics such as age, gender, interests and consumer habits. However, Social CRM allows another form of segmentation, since it has a lot of information that can be gathered online and is growing quickly. To manage and analyze that vast amount of information in real time, with the aim of segmenting social customers, it is necessary to use Big Data tools (such as Apache Mahout for databases, and R as a programming language), with data mining techniques (Petz et al., 2014): clustering, classification, association, regression and visualization. Regarding the types of segmentation, the most frequently used are: attitudinal, behavioral, demographic, loyalty and value-based (Fotaki et al., 2014).

Social customers can be organized in communities (social groups), which are groups of users linked by some kind of feature, relationship or common interest (Karrer et al., 2008; Wu et al., 2009). To do this, Social Network Analysis (SNA) tools are used, which provide mathematical and statistical routines that are applied to analyze the social networks, the results of which are represented in a social network diagram. The SNA employs community detection algorithms using the social contacts of individuals (Mossadegh and Behboudi, 2011).

By being in the appropriate community, social customers can (a) attract new social customers, (b) retain other social customers, and (c) acquire a new product or service by being influenced by other social customers of the same community (McKay, 2009; Serrat, 2010). The segmentation process should be carried out as follows:

- First, organize social customer types into different segments based on the communities concept, using as criteria the enterprise products and/or services of interest to each social customer.

- Then, in the communities deemed necessary, other levels of segmentation can be considered to create sub-communities, using as criteria the characteristics that are thought to be appropriate, depending on each community.

- Finally, there is a last level of segmentation that identifies social customers who provide more value within the initial segment. Some possible criteria for segmentation at this level are: profitability, growth potential, volume, competitive positioning issues, access to market knowledge, market share goals, margin levels, level of technology, resources and capabilities, compatibility of strategies, channel of distribution, and buying behavior (Lambert, 2010).

- There must also be a periodic monitoring of the evolution of communities and sub-communities in order to better understand their life cycle and thus manage them more efficiently, and, if necessary, undertake the social customer segmentation once again (Karrer et al., 2008; Serrat, 2010).

3.4.3.4. Analyze the Profitability of Social Customers

The profitability and potential of each social customer and community that the company has is analyzed. The profitability analysis is carried out not only in economic terms but also in relation to the image, productivity or any other benefit that the company can obtain as a result of the relationship with the social customers.

3.4.3.5. Define the Objectives of Social Customers

The objectives that will be assigned to each social customer and each community, in the short, medium, and long term are defined. These objectives are established based on the profitability and potential of each social customer and community. Examples of social customer objectives might be to improve the image, collaboration in the design and development of products, generating revenue with more effective cross-marketing, cross-selling and/or up-selling, to reduce marketing costs, etc. In order to achieve this, the company must develop strategies that enable social and business customers to operate as a true community, thereby generating some feeling of belonging and loyalty from the social customers. Social interactivity with social customers must be enhanced to obtain new ideas and different points of view. This will allow the social customer to be known from another perspective, understanding their preferences and also their demands and needs in order to provide a better service and/or product. After defining the objectives for each social customer or community, the level of fulfillment of these objectives can be analyzed through Big Data advanced analysis tools.

3.4.4. System for Assessing Customer Relations

The distinguishing characteristics of Social CRM compared to traditional CRM make it necessary to define new indicators that are not in the measurement system of a traditional CRM. These indicators should help to assess not only the degree of compliance with the needs and expectations of customers, as in traditional CRM, but also new aspects such as the degree of interaction, collaboration or commitment of social customers with the company, the involvement of company staff in Social CRM, or the business performance objective for Social CRM. The use of Big Data advanced analysis tools allows the defined indicators to be assessed both quickly and accurately. The indicator system of the Social CRM will be made up of:

- Traditional indicators like web page views, number of clicks, conversion rate and page or site "stickiness" (Greenberg, 2010), although also taking into account aspects such as time spent on a page, the number of times a page has been visited by the same person and the number of page views per session.
- Generic social indicators such as (Greenberg, 2010):
  - *Volume*. The number of times a topic is mentioned compared to historical patterns.
  - *Tone*. The percentages of positive, negative and neutral opinions.
  - *Coverage*. The number of sources that are generating the conversation regarding a particular topic.
  - *Authorization*. Classification of sources by their level of authority, and observation of how many rises or falls of conversations are generated by authoritative sources.
- Specific social indicators to monitor aspects of Social CRM of interest, such as: the resolution time for queries, the number of posts, the accuracy of the answers, the number of participants, etc. (Sarner et al., 2010). Additionally, the number of times an issue has been read on each channel, when and by which social customer, as well as whether it has been shared.

- Finally, the cause-effect relationships between the above indicators and the business objectives are defined.

3.4.5. Process Map

Initially, Social CRM was mostly a concern of marketing, but it now affects every customer-oriented discipline, from marketing and sales to customer service and support, as well as other internal company processes such as design, research, innovation, etc. Each of these enterprise business processes must be analyzed, defining how they can benefit from Social CRM through Web 2.0 and Big Data technology. This will involve redesigning the processes, modifying or adding new activities within each process to consider the contributions of the social customer. These contributions can come from both direct feedback as well as in an indirect form by extracting knowledge about their emotions and behavior from the profiles and maps of experience that are stored by the technological part of Social CRM (Mossadegh and Behboudi, 2011). As an example, Table 3.2 shows the business processes of different departments which have been improved as a result of the implementation of the SCRM-IRIS methodology in the company of the case study.

Moreover, due to the characteristics of Social CRM, there are several processes that did not exist in the company and had to be designed and implemented for the first time. These processes can be grouped into four areas (Dreyer and Grant, 2011a): Social media monitoring and responding; Social mapping profile; Outreach and lead generation; and Community management.

3.4.6. Human Resources

Almost every department of the company must participate in the Social CRM. Each department will be responsible for the part that is related to their roles in the company. The communication department is generally the one that should monitor and assign the right people to respond in public social spaces (Dreyer and Grant, 2011a).

Social software is not free, because time is money, and time must be invested in company staff working on a job that involves Social CRM. So it is important that much of this work is carried out by low-level staff, as their labor costs are lower. Lower-level staff can be trained with the necessary skills to enable them to manage social software applications, supported by the highest-level staff only when needed (Dreyer and Grant, 2011b).

A social media policy must also be defined, which must be derived from the social media strategy adopted, and which aims to educate employees by providing guidance on how the company requires them to use the social software. This policy should focus less on the "don'ts" and more on the "do's" and should facilitate and make the interaction with customers more pleasant and safer, as well as improving the ability to carry out the work. The main characteristics of good policies are: built on trust, practical, designed to educate, without absolutes, in plain language, friendly, consistent, prepared for mistakes, and clear about due process (Dreyer and Grant, 2010).

One last thing to consider is the risk of employees using the social software for private use. While such use is made in the proper environment among employees and with partners and social customers, it can lead to a better business relationship. Controlling use is very difficult, so there must be trust and such control should not be undertaken (Assaad and Marx Gómez, 2011).

Table 3.2. Business processes improved as a result of the implementation of the SCRM-IRIS methodology in the case study.

Department	Business process
Strategic management	Real-time analysis of the competitive environment
	Detection of changes in the competitive environment
	Data-driven decision-making
	Strategic planning
Operations	Troubleshooting in the products/services offered
	Increasing the quality of the products/services offered
	Offering an efficient catalogue of products/services based on sales trends analysis
Research and Development	Monitoring the performance and quality of the products/services offered
	Identifying the needs of customers of new products/services
	Identification of improvements in the products/services offered
Marketing / Sales	Analysis of customer information
	Identification of potential customers
	Identification of the most valuable customers
	Analysis of competing companies
	Gathering information about customers' needs
	Research about the company image
	Service acceptance analysis
	Monitoring social networks
	Price monitoring
	Detection of new releases by competing companies
	Analysis of relations in social networks
	Predicting customer behavior
	Accurate prediction and awareness of customers' needs
	Making real-time customized offers
	Encourage participation and interaction in every channel
	Quick reaction to market opportunities
	Analysis of sales trends
Customer assistance	Identifying customers who are at risk of ceasing to be customers of the company
	Analysis of how customers use the company website
	Monitoring how customers use the products/services offered by the company to detect potential problems and/or improvements

3.4.7. Computer System

This phase considers the Big Data and social software (or Web 2.0) applications and tools to be used in Social CRM. Both public (developed by other enterprises) and private (developed by the enterprise) ones should be taken into account, and those that will make up the computer system should be decided together with how they will be related to each other.

The computer support system of Social CRM is of great importance as it is the basis on which the Social CRM is run. This computer system has three basic objectives (Reinhold and Alt, 2011): To create a tool to efficiently discover the opinions and user reviews about the company and its products or services; to establish a contact channel for two-way interaction with users of social software; and to provide the means to integrate social content from the social software to processes and systems oriented towards the social customer.

The minimum requirements and/or abilities that the computer system must have to achieve these objectives are: (a) Data storage, (b) Customer Profiles storage, (c) Social knowledge, advanced analysis and monitoring, (d) User Generated Content (Mosadegh and Behboudi, 2011), and (e) User interaction (Reinhold and Alt, 2011).

The computer system to support Social CRM will consist of a set of Big Data and social software applications, and a set of Big Data and social software tools.

The Big Data and social software applications are Wikis, Blogs, Social Networks, Hadoop, MapReduce, Cassandra, etc. The private Social CRM applications must be developed considering the following characteristics (Sarner et al., 2012):

- To make social customers feel more involved in their own decisions.
- To give social customers access to more and better information on products and services.
- To provide more control in managing the public image and reputation online as well as how to decide what personal information is to be used.
- To improve self-esteem, friendship, the level of respect and commitment of social customers.
- To encourage participation in many-to-many relationships with social customers, prospects, selling partners and employees.
- To capture and share user-generated data and content.
- To provide various levels of autonomy and commitment to cede control of the community.
- To demonstrate the existence of mutual and balanced benefits for both the company and the community.
- To understand the profile, needs and feelings of social customers about the products and/or services offered by the company.

Regarding the Big Data and social software tools, there are many types of tools that can be part of a Social CRM computer system. Below are the most important types of tools (Dreyer and Grant, 2011a):

- *Social media monitoring*. These filter the web content based on various characteristics, such as for example the mention of keywords or comments about a particular topic. The Big Data advanced analysis tools provide speed and accuracy in monitoring social media, allowing companies to have real-time information with which to make decisions.
- *Social media Management/Marketing Systems*. These manage the process of posting and responding through social media channels, facilitating and unifying such management across multiple channels.

- *Social Discovery*. These search for contacts in social media profiles and map the relationships between contacts, allowing information to be found from the contacts. In addition, they categorize, standardize and structure the big unstructured data existing in the social media, in order to harness the wisdom of crowds, without human intervention.

- *Email Systems*. These help to connect and communicate with contacts and customers by e-mail, creating and/or segmenting mailing lists using social discovery data and allowing information from social discovery data to be provided in e-mails.

- *Communities.* These help to create social links through the website, offering various extra features to the contacts.

- *Association Management Systems.* Databases where data such as contacts, interactions, transactions carried out, etc. are recorded.

- *Social Network Analysis.* These analyze the social network links using graph theory and display the results in social network diagrams. This is a nice graphical way to analyze and visualize the large amount of existing data, as well as those generated every day. It is a part of Big Data that focuses on relationships and/or interactions between users of online social networks (Alamsyah and Peranginangin, 2013).

To support Social CRM, a Social CRM Computer System Architecture is proposed. It allows advantage to be taken of Web 2.0 and Big Data technologies (Figure 3.2).

This Social CRM Computer System Architecture is composed of three levels:

- *Content.* This level allows both structured and unstructured data to be gathered from different sources. At this level the links with the data sources are made by applying the necessary filters and patterns to obtain only valuable data for the company. Data sources can be both traditional computer systems (like corporate ERP, BI, structured databases, etc.) and social software that supports and stores human information and communication. The latter represents a major technological challenge because it makes it necessary to work with complex, unstructured, ubiquitous, multi-format and multi-channel information.

- *Enrichment*. This level works with the data gathered by the Content level. The objective is to enhance the data and to identify and to extract valuable information for the company CRM from those data. For this purpose different advanced data analysis techniques are used. These techniques are organized in four groups: Inquire, Interact, Investigate and Improve. The result is valuable social data obtained from social software, links to valuable corporate data, and valuable new data (metadata) generated from the analysis of the previous data gathered at the Content level. These data are stored in a Social CRM structured database. Metadata are used to simplify and reduce the complexity and processing time at the next level (Modules level).

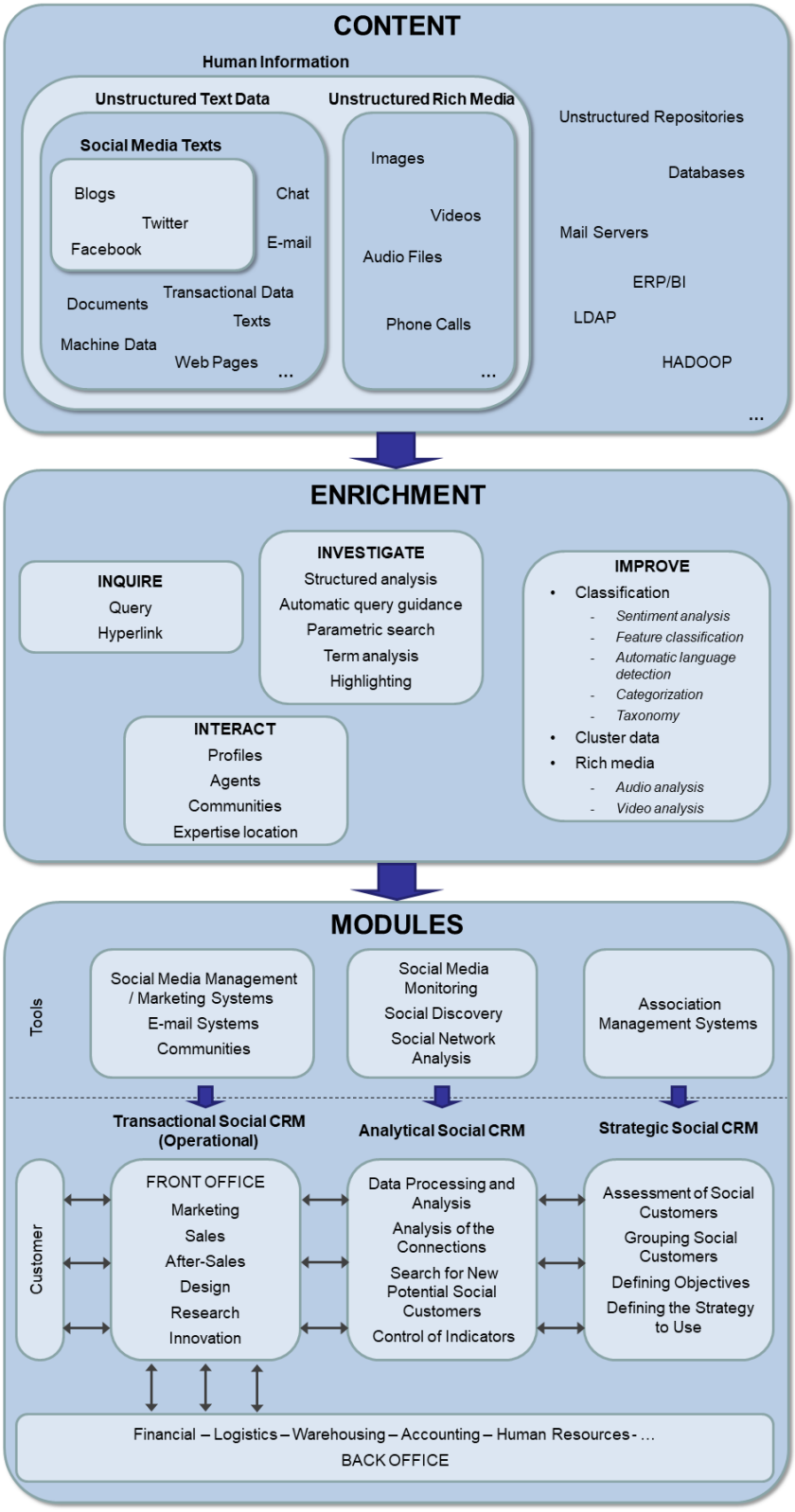

Figure 3.2. Social CRM Computer System Architecture.

- *Modules.* The enriched information extracted from the Enrichment level and stored in the Social CRM structured database is processed by the three main modules of the Social CRM software, using different techniques like data mining, predictive analysis or machine learning: (1) *Transactional Social CRM (Operational)*. This

supports the business processes that can be improved with the Social CRM: Marketing, Sales, After-sales, Design, Research, Innovation, etc.; (2) *Analytic Social CRM*. This allows data from social customers to be used to manage and improve relationships with existing social customers and achieve business objectives. Moreover, it also carries out an analysis of online social communities to find new potential social customers in these communities. Finally, it also allows measurement of the indicators of the Social CRM performance measurement system; and (3) *Strategic Social CRM*. This permits each social customer (profile, contacts, etc.) to be assessed and grouped in segments depending on their characteristics. In addition, it allows the definition of objectives for each segment in the short, medium, and long term, as well as definition of the strategy to be used in each segment to meet the proposed objectives.

3.4.8. Implantation

The person in charge of the implantation should be the Community Manager, who should always be available to solve problems or queries arising from users, both in the implementation as well as when the system is operating. It is important that the response time in resolving problems or concerns is short.

Implementing a Social CRM system is very similar to the implementation of a CRM system; therefore, the steps to be followed in implementing a Social CRM system are the same as those shown in Chalmeta (2006). The collaboration of all Social CRM users in the implementation is critical, so it is essential that they change their mentality and assume that the center of Social CRM is the social customer. On the other hand, in order to take advantage of Big Data advanced analysis tools, it is necessary to have good quality data. To ensure that the data are of good quality, organizations must maximize the following properties (Chiang and Sitaramachandran, 2015): a) *Existence*. The organization has or can get the data; b) *Validity*. The data values are within an acceptable range; c) *Consistency*. The same data has the same value regardless of where it is located; d) *Integrity*. Completeness relationships between data elements; e) *Accuracy*. The data describe the properties of the model; and f) *Relevance*. The data are appropriate to achieve the proposed objectives.

The person in charge of the creation and maintenance of the continuous improvement system is also the Community Manager, as he/she is the one who knows the entire system best and, therefore, is also better able to identify potential future improvements.

3.4.9. Monitoring

Social CRM must be monitored throughout its entire lifespan. The Community Manager is in charge of carrying out this monitoring process. The characteristics of Web 2.0 and Big Data technologies allow the monitoring to be carried out quickly and effectively, while providing great control over the system.

In order to establish an effective monitoring system, the following tasks must be carried out:

- To monitor the indicators defined in stage four of the methodology so as to be able to carry out a follow-up of the system in order to measure its performance. To do so, the following technologies (which are based on Big Data advanced analysis

tools) are used: Social Media Monitoring, Social Media Management/Marketing Systems and Association Management Systems.

- To adapt or modify the functionalities required, in order to fix errors and improve the system.
- To generate and maintain a system of periodic motivation for users, to reward users that make good use of the Social CRM system, since the participation of users is essential for Social CRM.
- To carry out a control of accesses and input from all users, as it is very important they make good use of the system.
- To carry out periodic user surveys to assess their level of satisfaction with the Social CRM, and to ascertain the level of acceptance of the system and compile their suggestions for improvement. After the results of the surveys have been obtained, they are studied and the relevant modifications are made.
- To periodically carry out an analysis of the system's stability, which checks whether the Social CRM system is working properly, if it reacts correctly to the data it manages and if users use it properly.

3.5. Conclusions

Organizations must be aware of the shift that is occurring in the use of data and must actively prepare to participate in it. Among the measures to be taken, three are absolutely critical: Treating information and data as a corporate asset at the same level as human and financial resources; The company should be capable of generating and sharing knowledge from the data; and Designing and implementing a technology infrastructure that makes it possible to address the challenges and opportunities presented by technological disruptors such as Security, Cloud, Mobility, and Big Data.

This chapter presents a methodology, called the SCRM-IRIS methodology, to help companies to obtain value from data, by developing a Social CRM system. The methodology has been applied to a company in order to refine and validate it. Those responsible for the application of the SCRM-IRIS methodology in the company have indicated that the use of the SCRM-IRIS methodology has allowed them to gain an excellent view of the needs, scope, consequences and opportunities of the project, as well as allowing them to implement Social CRM quickly and without significant problems. They have also indicated that the methodology has enabled them to have greater control over the project, because all the activities to be performed in each phase of the implementation are clearly defined.

There are various proposals for future investigations. The future of Big Data and Web 2.0 technologies is going through a general expansion for all industries to be applied to all business processes and aspects of organizations. Through Big Data analytics and Web 2.0 technologies, the company can not only quickly and reliably monitor the acceptance of its products and/or services in the marketplace, but they also allow them to understand their business environment as well as find and strengthen competitive advantages (Kwon et al., 2014). Therefore, new methodologies, similar to the SCRM-IRIS methodology, are needed to support the adoption and implementation of Big Data and Web 2.0 technologies in other

areas of the company, such as company strategy, supply chain management, product design, and so on.

Finally, some limitations of this chapter should be noted and discussed. First, the qualitative method used for the analysis of the benefits obtained by the company from the application of the methodology is not as accurate as a quantitative analysis. The method is based on the opinion of those in charge at each implantation. However, their experience and professionalism make it possible to rely on the veracity of these results. Only a basic quantitative analysis was performed. This can be future research for academics that can apply advanced quantitative methods to measure the benefits of SCRM-IRIS at department level as well as business level. Finally, the company where the SCRM-IRIS methodology was applied already had CRM technologies (without Social CRM features). Therefore, if this methodology were to be applied in a company that did not have them, the implementation process would be more expensive and complex, because the implantation of CRM strategies, culture and computer systems, and training of the employees would have to be undertaken from scratch.

3.6. References

Alamsyah A, Peranginangin Y (2013). Effective Knowledge Management using Big Data and Social Network Analysis. Learning Organization Management and Business International Journal, 1(1):17-26.

Anshari M, Alas Y, Yunus N, Sabtu NI, Hamid MH. (2015) Social customer relationship management and student empowerment in online learning systems. International Journal of Electronic Customer Relationship Management, 9(2/3):104-121.

Assaad W, Marx Gómez J (2011). Social Network in Marketing (Social Media Marketing) Opportunities and Risks. International Journal of Managing Public Sector Information and Communication Technologies, 2(1):13-22.

Bebensee T, Helms R, Spruit M (2011). Exploring Web 2.0 Applications as a Mean of Bolstering up Knowledge Management. The Electronic Journal of Knowledge Management, 9(1):1-9.

Bose R (2002). Customer relationship management: key components for IT success. Industrial Management & Data Systems, 102(2):89-97.

Chalmeta R (2006). Methodology for customer relationship management. The Journal of Systems and Software, 79(7):1015-1024.

Chen IJ, Popovich K (2003). Understanding customer relationship management (CRM): People, process and technology. Business Process Management Journal, 9(5):672-688.

Chiang F, Sitaramachandran S (2015). A Data Quality Framework for Customer Relationship Analytics. Web Information Systems Engineering, vol 9419. Springer, Part II, pp 366-378.

Crockett B, Reed K (2003). The foundation of insight: three approaches to customer-centric understanding. The Ultimate CRM Handbook, McGraw-Hill, New York, NY.

Dietrich J, Jones N, Wright J (2008). Using social networking and semantic web technology in software engineering - Use cases, patterns, and a case study. The Journal of Systems and Software, 81(12):2183-2193.

Dreyer L, Grant M (2010). Social Media, Risk, and Policies for Associations. SocialFish & Croydon Consulting white paper (Online), 14 pages, January 13, 2010. Available at: http://www.socialfish.org/whitepaper. Accessed 2 May 2015.

Dreyer L, Grant M (2011a). Social CRM for Associations. SocialFish & Avectra white paper (Online), 14 pages, March 29, 2011. Available at: http://www.socialfish.org/whitepaper. Accessed 18 Mar 2015.

Dreyer L, Grant M (2011b). ROI and the Impact of Social CRM. SocialFish & Avectra white paper (Online), 16 pages, August 17, 2011. Available at: http://www.socialfish.org/whitepaper. Accessed 19 Feb 2015.

Faase REM, Helms RW, Spruit MR (2011). Web 2.0 in the CRM domain: defining Social CRM. International Journal of Electronic Customer Relationship Management, 5(1):1-22.

Fayerman M (2002). Customer Relationship Management. New Directions For Institutional Research, 2002(113):57-68.

Feng S, Wang D, Yu G, Gao W, Wong KF (2011). Extracting Common Emotions from Blogs Based on Fine-grained Sentiment Clustering. Knowledge and Information Systems, 27(2):281-302.

Fotaki G, Spruit M, Brinkkemper S, Meijer D (2014). Exploring big data opportunities for online customer segmentation. International Journal of Business Intelligence Research, 5(3):58-73.

Go E, You KH (2016). But not all social media are the same: Analyzing organizations' social media usage patterns. Telematics and Informatics, 33(1):176-186.

Greenberg P (2009). Social CRM Comes of Age. ORACLE. Available at: http://hosteddocs.ittoolbox.com/social-crm-comes-of-age_paul-greenberg.pdf. Accessed 15 Sep 2015.

Greenberg P (2010). The impact of CRM 2.0 on customer insight. Journal of Business & Industrial Marketing, 25(6):410-419.

Hu Y-H, Huang T, Kao Y-H (2013). Knowledge discovery of weighted RFM sequential patterns from customer sequence databases. The Journal of Systems and Software, 86(3):779-788.

Karlström D, Runeson P (2006). Integrating agile software development into stage-gate managed product development. Empirical Software Engineering, 11(2):203-225.

Karrer B, Levina E, Newman MEJ (2008). Robustness of community structure in networks. Physical Review E., 77(4):046119.

King SF, Burgess TF (2008). Understanding success and failure in customer relationship management. Industrial Marketing Management, 37(4):421-431.

Kirchner K, Razmerita L, Nabeth T (2009). Personal and Collective Knowledge Management in the Web 2.0: Two Faces of Knowledge Management. Paper presented at the 9th International Conference on Innovative Internet Community Systems, Jena, Germany, pp.15-26.

Kwon O, Lee N, Shin B (2014). Data quality management, data usage experience and acquisition intention of big data analytics. International Journal of Information Management, 34(3):387-394.

Lai LSL, To WM (2015). Content analysis of social media: a grounded theory approach. Journal of electronic commerce, 16(2):138-152.

Lambert DM (2010). Customer relationship management as a business process. Journal of Business and Industrial Marketing, 25(1):4-17.

Lech P (2016). Causes and remedies for the dominant risk factors in Enterprise System implementation projects: the consultants' perspective. SpringerPlus 5:238 DOI: 10.1186/s40064-016-1862-9.

Lee M, Lan Y (2007). From Web 2.0 to Conversational Knowledge Management: Towards Collaborative Intelligence. Journal of Entrepreneurship Research, 2(2):47-62.

Marshall A, Mueck S, Shockley R (2015). How leading organizations use big data and analytics to innovate. Strategy & Leadership, 43(5):32-39.

McKay L (2009). Strategy and Social Media: Everything's Social (Now). Customer Relationship Management; CRM 13(6).

Mosadegh MJ, Behboudi M (2011). Using social network paradigm for developing a conceptual framework in CRM. Australian Journal of Business and Management Research, 1(4):63-71.

Musser J, O'Reilly T, the O'Reilly Radar Team (2006). Web 2.0 Principles and Best Practices. O'Reilly Radar Report, Nov. 2006.

Nguyen TH, Sherif JS, Newby M (2007). Strategies for successful CRM Implementation. Information Management & Computer Security, 15(2):102-115.

O'Reilly T (2005). What is Web 2.0 – Design Patterns and Business Models for the Next Generation of Software. Available at: http://oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1. Accessed 18 Aug 2015.

Omerzel DG (2010). The impact of knowledge management on SME growth and profitability: a structural equation modeling study. African Journal of Business Management, 4(16):3417-3432.

Olbrich R, Holsing C (2012). Modeling consumer purchasing behavior in social shopping communities with click stream data. Int. J. Electron. Commerce, 16(2):15–40.

Petz G, Karpowicz M, Fürschuß H, Auinger A, Stříteský V, Holzinger A (2014). Computational approaches for mining user's opinions on the Web 2.0. Information Processing & Management, 50(6):899-908.

Ras E, Rech J (2009). Using Wikis to support the Net Generation in improving knowledge acquisition in capstone projects. The Journal of Systems and Software, 82(4):553-562.

Razmerita L, Kirchner K, Sudzina F (2009). Personal knowledge management: The role of Web 2.0 tools for managing knowledge at individual and organizational levels. Online Information Review, 33(6):1021–1039.

Reinhold O, Alt R (2011). Analytical Social CRM: Concept and Tool Support. Proceedings of the 24th Bled eConference eFuture: Creating Solutions for the Individual, Organizations and Society. Bled, Slovenia, pp.226-241.

Robson C (2002). Real world research: a resource for social scientists and practitioner-researchers, 2nd edn. Blackwell.

Rosenberger M (2015). Social Customer Relationship Management: An Architectural Exploration of the Components. Lecture Notes in Computer Science, 9373:372-385.

Runeson P, Höst M (2009). Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering, 14(2):131–164.

Sarner A, Thompson E, Dunne M, Davies J (2010). Top Use Cases and Benefits for Successful Social CRM. Gartner Research, ID Number: G00209091 (Online), 9 pages, December 3, 2010. Available at: http://www.gartner.com/id=1485114. Accessed 22 Jul 2015.

Sarner A, Thompson E, Sussin J, Nikos D, Maoz M, Davies J, Mann J (2012). Magic Quadrant for Social CRM. Gartner Research, ID Number: G00237431 (Online), 20 pages, September 27, 2012. Available at: http://www.gartner.com/id=2179417. Accessed 24 Jul 2015.

Serrat O (2010). Social network analysis. Washington, DC: Asian Development Bank.

Shimazu H, Koike S (2007). KM2.0: Business Knowledge Sharing in the Web 2.0 Age. NEC Technical Journal, 2(2):50-54.

Strauss J, Frost R (2002). Customer Relationship Management. E-Marketing, 2nd edn. Prentice Hall, New York.

Syed A, Gillela K, Venugopal C (2013). The Future Revolution on Big Data. International Journal of Advanced Research in Computer and Communication Engineering, 2(6):2446-2451.

Torre-Bastida AI, Villar-Rodriguez E, Gil-Lopez S, Del Ser J (2015). Design and Implementation of an Extended Corporate CRM Database System with Big Data Analytical Functionalities. Journal of universal computer science, 21(6):757-776.

Wirtz BW, Piehler R, Ullrich S (2013). Determinants of social media website attractiveness. Journal of electronic commerce, 14(1):11-33.

Wu B, Ye Q, Yang S, Wang B (2009). Group CRM: a new telecom CRM framework from social network perspective. Proceeding of the ACM First International Workshop on Complex Networks Meet Information & Knowledge Management, pp.3-10.

Zablah AR, Bellenger DN, Johnston WJ (2004). An Evaluation of Divergent Perspectives on Customer Relationship Management: Towards a Common Understanding of an Emerging Phenomenon. Industrial Marketing Management, 33(6):475-489.

Zamil AM (2011). Customer Relationship Management: A Strategy to Sustain the Organization's Name and Products in the Customers' Minds. European Journal of Social Sciences, 22(3):451-459.

Chapter 4. Methodology for the Implementation of Knowledge Management Systems 2.0: A Case Study in an Oil and Gas Company

Web 2.0 and Big Data tools can be used to develop knowledge management systems based on facilitating the participation and collaboration of people in order to enhance knowledge. This chapter presents a methodology that can help organizations using Web 2.0 and Big Data tools to discover, gather, manage and apply their knowledge by making the process of implementing a knowledge management system faster and simpler. First, an initial version of the methodology was developed and it was then applied to an oil and gas company in order to analyze and refine it. The results obtained show the effectiveness of the methodology, since it helped this company to carry out implementation quickly and effectively, thereby allowing the company to gain the maximum benefits from existing knowledge.

4.1. Introduction

Knowledge management (KM) helps enterprises to provide customers with better products and services, in response to their ever-increasing demands in terms of flexibility, speed and quality (Ipe, 2003). Thus, an important part of overall business administration is the management of knowledge, which comprises the systematic analysis, planning, acquisition, creation, development, storage and use of knowledge (Nakamori, 2003).

A key factor for achieving correct KM in an organization is the development and implementation of a knowledge management system (KMS) to manage the knowledge of organizations automatically (Alavi and Leidner, 2001; Day, 2001). KMS have three common applications: (a) codifying and sharing best practices, with the aim of transferring them internally; (b) creating directories of corporate knowledge by identifying, classifying and codifying existing internal abilities, since organizations possess a great deal of knowledge that remains hidden and uncoded; and (c) creating knowledge networks, which allow users to communicate in a quick and simple way (Alavi and Leidner, 2001).

Since they appeared in the mid-1990s, the majority of KMS have been based on identifying and capturing explicit and tacit knowledge related with the company and centralizing it in a widely available company platform. According to Ernst and Young (2001), the main kinds of KMS platforms are Intranets and corporate portals, data warehouses or knowledge repositories, decision support tools and document management

systems. However, the results show the expected outcomes have not been achieved (Serenko et al., 2010). Two main causes for this have been identified (Davenport, 2005). One is the difficulty in finding what users need, and the other is that the above-mentioned platforms do not allow knowledge to be captured, shared and applied easily. As a result of this, the greater part of the knowledge of company best work practices, relevant experience, tacit knowledge and outputs are invisible to most people (McAfee, 2006).

In addition to these KMS (which could be called traditional), in recent times KMS 2.0 have also appeared. KMS 2.0 are KMS that use Web 2.0 and Big Data technologies and are focused on facilitating collaboration in order to enhance knowledge (Kakizawa, 2007; McAfee, 2006; Shimazu and Koike, 2007). KMS 2.0 have generated renewed expectations for the way in which they might help organizations to improve their KM (Pawlowski et al., 2014). KMS 2.0 are based on the participation of people, who generate new knowledge and are not limited to just consuming it, i.e. users are active contributors (Razmerita et al., 2009). Therefore, in contrast to traditional KMS, which concentrate only on capturing knowledge, KMS 2.0 are focused on the practices and output of knowledge workers (McAfee, 2006). KMS 2.0 provide the framework, while the content is provided by users (Omerzel, 2010). Another difference is that traditional KMS are highly structured from the start, and users have little opportunity to influence this structure. This increases the difficulty involved in capturing highly unstructured knowledge work that has to be fitted and recorded in a database of inflexible categories (Trimi and Galanxhi, 2014). However, in KMS 2.0, companies build an initial structure and hierarchy, and users can constantly change this structure, thus creating new content, links and tags as a part of their regular daily routines. Nevertheless, both types of KMS work on similar principles, which allow them to manage knowledge within the enterprise (Levy, 2009; Paroutis and Al Saleh, 2009; Schneckenberg, 2009).

On the one hand, KMS 2.0 use Web 2.0 technologies because they offer a variety of tools that make it possible to communicate with others more effectively, encourage collaboration and facilitate social interaction and the sharing of knowledge (Kirchner et al., 2009; Wu et al., 2013). Nowadays organizations are very interested in Web 2.0 tools, since these (1) act as a harbinger of how people will behave in the future (Abramowicz et al., 2010); and (2) increase agility (new ideas, suggestions and opportunities are shared), flexibility (work elements are broken down) and productivity (they provide faster and easier communication, collaboration and content management within and across companies) (Trimi and Galanxhi, 2014). These Web 2.0 technologies can be provided by the companies, which develop and integrate them into company social software platforms to foster employees' collaboration and communication, or they can be developed by external companies and open to everybody (Kügler et al., 2015).

On the other hand, KMS 2.0 can also take advantage of Big Data, since recent technological revolutions such as Web 2.0 technologies enable data to be generated much faster than in the past (McAfee and Brynjolfsson, 2012). To describe Big Data, the Vs frameworks have emerged as a common structure. Hence, Big Data can be defined as follows: "*Big Data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making"* (Gartner IT Glossary, 2016). Three other Vs have been added, subsequently, to the previous ones: Veracity, which represents the unreliability inherent in some sources of data; Variability (and complexity), which refers to the fact that Big Data are generated through a myriad of sources; and Value, which refers to the fact that data in their original form usually have a low value relative to their volume, but a high value can

be obtained by analyzing large volumes of such data (Gandomi and Haider, 2015). Regarding the fundamental question of how big the data have to be to qualify as Big Data, there is little consensus on the issue. The limits depend upon the size, sector and location of the company and these limits evolve over time (Gandomi and Haider, 2015). In this sense, Kaisler et al. (2013) defined these limits as the amount of data just beyond the capacity of traditional data management and analysis technologies to store, manage and process it efficiently.

Big Data allows knowledge to be extracted (with few hardware resources) from large amounts of data, such as machine-generated data, log files, e-mail messages, unstructured text, video, images, audio posted on public/company social networks, and other types of information sources (Wieczorkowski and Polak, 2014). Based on the characteristics of collected data, different methods and technologies can be applied to discover knowledge (Gandomi and Haider, 2015). Organizations are viewing Big Data as a valuable asset and a source of competitive advantage in many business settings (Schermann et al., 2014; Shao and Lin, 2016), and they are making important efforts on the development and optimal utilization of Big Data technologies in order to take the appropriate decisions (Zhao et al., 2015). Big Data technologies allow monitoring of key factors for strategic decisions, such as customer opinions about a product, service or company, by mining social media data (Tan et al., 2013).

The development and implementation of a KMS 2.0 in an organization is a complex task that requires the participation of users (who need to acquire skills in selecting, reflecting and redistributing knowledge online while ensuring its quality) and the organization (which needs to acquire the organizational capacity to react adequately to the content generated by users) (Schneckenberg, 2009). Significant technological and cultural changes also need to be carried out within the organization, since this is not just a technological improvement but involves a new interpretation of knowledge management based mainly on the contribution made by users (Bebensee, 2011). Therefore, Pawlowski et al. (2014) identify four categories of challenges related to KMS 2.0 development: social and cultural dimensions, organizational dimensions, technical dimensions, and knowledge protection and legal dimensions.

Like any IT project, organizations also need a methodology that guides them on how to deal with the innovation and change involved in implementing advanced software, so as to shorten the time needed to obtain business benefits and reduce the risk of failure in the implementation (Fichman and Moses, 1999). This KMS 2.0 development methodology can be defined as a framework for applying KMS development practices, and should establish the phases of system development along with the proper sequence for applying them, the human roles in each phase, the products of each phase, and guidelines and metrics for progress monitoring and quality assurance (Razieh and Raman, 2015).

The literature, however, does not contain any specific methodologies to help in the development of KMS 2.0 (Mariscal et al., 2010). The literature that does exist on the topic of KMS 2.0 technology focuses primarily on the characteristics, opportunities and benefits they offer (McAfee, 2006; Musser et al., 2006; O'Reilly, 2005; Schneckenberg, 2009) but does not offer any methodological guidelines regarding phases, human roles, products or metrics of the KMS 2.0 development process.

On the other hand, the methodologies that are oriented towards traditional KMS Development (KMSTD) could be considered. The most comprehensive methodologies

which exist in this context are those presented by Amine and Ahmed-Nacer (2011), Chalmeta and Grangel (2008), Iglesias and Garijo (2008), Moteleb et al. (2009), Rubenstein-Montano et al. (2001), Sarnikar and Deokar (2010), and Smuts et al. (2009). However, currently existing KMSTD have the following weaknesses that make it difficult to develop a proper KMS 2.0 (Razieh and Raman, 2015): they do not fully cover the basic phases of KMS development (requirements engineering, analysis, design, implementation, test, deployment and maintenance); planning activities have been neglected; lack of feasibility-study activities; lack of a clear and accurate specification of the activities of each phase and their execution sequence, prescribing an activity without suggesting specific techniques for performing it; some of these methodologies have not been used in practice; some of these methodologies are of a more theoretical nature rather than practice-based; poor user involvement, restricted to validation; failure to specify appropriate technologies and tools, not taking into account the possibilities of Web 2.0 and Big Data technologies to support KM; lack of attention to distinguishing tacit KM from explicit KM; absence of periodical validation; the enterprise model 2.0 is not considered; no mechanisms are established for promoting the cultural change needed in order to foster the sharing of knowledge; failure to determine managerial responsibilities and their assignment to the right individuals; failure to manage the financial resources properly; lack of attention to user requirements at different organizational levels; and they do not allow the business processes and jobs to be redesigned so that they can use the knowledge that resides in the KMS 2.0 and generate new knowledge.

Consequently, there are a number of problems concerning the methodologies for managing KMS 2.0 development and implementation projects that remain unsolved. Hence, there is still room for significant improvement as regards both their theoretical aspects and their practical applicability (Šajeva, 2007).

To help solve this problem, this chapter proposes a step-by-step methodology, called Web 2.0 Knowledge Management (W2KM) methodology, which can guide the entire process of developing and implementing a KMS 2.0, reducing the complexity of this process. This W2KM methodological guide is composed of phases, which contain different activities and these activities are in turn made up of several tasks. In order to improve and debug the W2KM methodology, it was applied in a large oil and gas company.

This chapter is organized as follows: Section 4.2 presents a review of the literature related with Web 2.0 tools, Big Data tools and knowledge management 2.0. Section 4.3 outlines the W2KM methodology proposed here for the implementation of a KMS 2.0, which is applied in a case study that is described in Section 4.4. Finally, the main conclusions and the limitations of this work are analyzed and discussed in Section 4.5.

4.2. Literature Review

The term Web 2.0 was first used by O'Reilly Media and Media Live International in 2004 as the name of a series of conferences given by them (Antonova et al., 2009; Lee and Lan, 2007; Levy, 2009). There is no generally agreed definition of Web 2.0. One of the most widely used is the one proposed by Tim O'Reilly, who defines it as "*a set of economic, social, and technology trends that collectively form the basis for the next generation of the Internet – a more mature, distinctive medium characterized by user participation, openness, and network effects*" (Musser et al., 2006, p. 4). Web 2.0 applications are constantly updated and improved as more and more people use them, thereby consuming

and mixing information from multiple sources. Users provide data and services in a way that allows others to blend them again, thus creating a network of effects through the "architecture of participation" (O'Reilly, 2005). Web 2.0 is not only a new generation of technologies, but also a change in the way in which users access the Internet in order to mutually interact and collectively create knowledge. The characteristics of the knowledge managed by means of Web 2.0 tools are as follows (Lee and Lan, 2007):

- **Contribution**. Each user has the opportunity to freely provide his or her knowledge.
- **Sharing**. Knowledge contents are freely available for others (through security mechanisms).
- **Collaboration**. Knowledge contents are created and maintained by means of collaboration among the suppliers of knowledge.
- **Dynamism**. Knowledge contents are constantly updated to reflect changes in the environment and the situation.
- **Trust**. The contribution of knowledge must be based on trust among the suppliers of knowledge.

Web 2.0 consists of a set of emerging tools that provide the basis for a more mature Internet, in which users collaborate, share information and create networks in large communities (McAfee, 2006; Musser et al., 2006; O'Reilly, 2005). Some of the most common Web 2.0 tools include: Wikis, Group chat, Social bookmarking, Mashups, Blogs, RSS, Folksonomy, Podcasts, and Social Networks.

From the users' point of view, to be able to manage knowledge in an organization successfully, the Web 2.0 tools must possess certain fundamental features (Dai et al., 2007), including:

- **System functionality**. Users must find the system "friendly", easy-to-use and under control.
- **Quality of the content**. The core of all online information systems is the content, which must therefore be reliable, relevant, timely and appropriate.
- **Exchange and accessibility of the content.** The system must motivate the user to exchange useful information and to share knowledge.
- **Sociability**. The system must possess a high level of social integration, since this is crucial for the success of any online community.

Normally, emerging technologies appear in enterprises and are then passed on to consumers. But in the case of the Web 2.0 the flow has been inverted, since they appeared first among consumers and were later transferred to enterprises (Kakizawa, 2007). Thus, Web 2.0 tools have already successfully proved their capacity to manage knowledge related with people's leisure. To validate the claim that Web 2.0 tools are appropriate for managing the knowledge in an enterprise, Levy (2009) compared Knowledge Management and Web 2.0 on the basis of four aspects: (a) conceptual, (b) principles, (c) functional skills of tools and applications, and (d) organizational culture. The conclusion that was reached

was that Web 2.0 tools are perfectly well suited for managing the knowledge of any enterprise, but it must be borne in mind that the Web 2.0 is focused on people, while knowledge management is centered on the organization. Hence, to take advantage of the characteristics of the Web 2.0, enterprises need to change the approach they had with the traditional KMS.

Therefore, during the development of a KMS 2.0 for an enterprise it is essential to take into account the common elements that characterize all the Web 2.0 technologies used in the enterprise. These elements were first identified by McAfee through the abbreviation SLATES, which stands for: Search (providing search query capabilities that allow content to be located easily, quickly and automatically); Links (guiding the user towards what is really important and also structuring the online content); Authorship (allowing any user or any group of people to create contents); Tags (offering a new collaborative way of categorizing contents by means of folksonomy); Extensions (using suggestions and recommendations to speed up searches); and Signals (receiving alerts when a site that is of interest to the user is modified) (McAfee, 2006). One year later, Dion Hinchcliffe proposed another mnemonic to represent the elements that characterize Web 2.0 technologies in the enterprise. The abbreviation in this case is FLATNESSES, which consists of the same elements as SLATES plus four new elements: Freeform (the system must be capable of evolving freely, so as to become what users want it to be); Network-oriented (the content of the applications must be Web-oriented, as well as addressable and reusable); Social (allowing users to share their social information); and Emergence (something complex can arise from relatively simple interactions) (Hinchcliffe, 2007).

Traditional KMS are closed systems that store answers to issues that may possibly arise in the course of a job, supposing that workers are carrying out tasks that have previously been anticipated and described. Such an assumption creates a barrier hindering innovation because it prevents workers from sharing their new ideas with their colleagues, so that they can be discussed, debated or generated. Closed systems do not allow communities to have control over their own knowledge – instead they separate creation from integration. Innovations therefore take place outside the systems and the systems contain information that is passed on chronologically, which reflects a point of view from outside the work itself (Brown and Duguid, 2000; Fischer and Ostwald, 2001).

The Web 2.0 has reinvented the concept of knowledge management by basing itself on the idea of facilitating the interaction, cooperation and exchange of knowledge among individuals, groups and communities. In the Web 2.0, there is no distinct differentiation between individual and collective knowledge. The Web 2.0 focuses on the exchange of knowledge and collaboration among employees, who are the knowledge workers in the organization. The aim of such an approach is to take advantage of collective intelligence and speed up the flow of knowledge among people through formal and informal communication, collaboration and social networking. Web 2.0 tools cover the different facets of knowledge management well (Kirchner et al., 2009).

Knowledge is one of the most valuable resources for an organization and the most important type of knowledge is located inside people's heads: it is *embrained* (Blackler, 1995). In this regard, KMS 2.0 makes it necessary to change the way knowledge is managed, since management must now be person-based. Furthermore, the use of Web 2.0 tools to manage knowledge enables organizations to obtain important benefits at a lower cost than by using traditional KMS (Razmerita et al., 2009). When it comes to promoting products and services, Web 2.0 tools enable organizations to reach a high communication

visibility more economically, as well as providing them with valuable feedback (Kirchner et al., 2008): They enable to capture the "wisdom of the crowd" (Surowiecki, 2004).

Moreover, KMS 2.0 can also be complemented by Big Data tools, since these tools allow enterprises to extract and generate new knowledge from large amounts of structured and unstructured data (Syed et al., 2013). In recent years there has been a decrease in the cost of data storage and data processing, and an increase in data sources (social networks, mobile devices, machine-generated data, etc.), which has caused the exponential growth, availability, and use of information (Jeong and Shin, 2015). Big Data refers to data sets that are too large and complex to be processed using traditional means of storage like relational database technologies and analysis technologies (Debortoli et al., 2014). Big Data tools summarize technological developments and techniques in the area of data storage and data processing that allow the handling of exponential increases of data in terms of volume, variety, velocity, value and veracity (Schermann et al., 2014). Big Data could be seen as an evolution of business intelligence, which focuses on obtain reports, mainly indicators to measure past business performance, from structured internal company databases. Thus, Big Data focuses on extracting value from semi-structured and unstructured data originated in data sources like the Web, mobile devices or sensor networks. Another difference is the types of questions they answer, which, in the case of Big Data, are related with exploration, discovery and prediction (Debortoli et al., 2014).

Companies have far more data available to them, and they want to take advantage of such amount of data. Big Data is able to generate knowledge from these data (Erickson and Rothberg, 2014) and can do it with speed and accuracy, which can be very relevant and valuable for the performance of the enterprise in various dimensions, as well as to support decision-making (Dutta, 2015; Song et al., 2015). Not only is Big Data able to extract knowledge from data generated by the enterprise itself (e-mail messages, machine-generated data, log files, transaction records, sensor data, internal Web 2.0 tools, and so on), but it can also extract knowledge from data generated by external applications (messages posted on public Social Networks, data in public repositories, data published on websites, GPS signals, and so forth) (Wieczorkowski and Polak, 2014). This knowledge will enable the organizations to achieve competitive advantage over their competitors, develop new products and/or services, make strategic and operational decisions, identify what has happened, and predict what will happen in the immediate future.

Nevertheless, despite their benefits, the chances of failure in the implementation of KMS 2.0 in organizations are high. According to Šajeva (2007) there are five types of barriers that restrict knowledge management in organizations, and they also appear when Web 2.0 technologies are used:

- Barrier 1: *Individual barriers* are the barriers related to users. The main types are: Fear, for example, of losing authority and power, or of becoming replaceable; Lack of motivation, for example, lack of commitment or refusing to do intrusive and extra work; and Personal characteristics, for example, poor communication and interpersonal skills, or lack of awareness of KM strategies and tools.

- Barrier 2: *Organizational context related barriers* refer to behavioral and organizational aspects. The main ones are: Cultural, for example, closed corporate culture or resistance to change; Structural, for example, rigid hierarchies, or lack of formal and informal tools to collaborate, reflect and generate knowledge; Management related, for example, lack of motivational and reward systems, or lack

of management commitment; and Strategic management related, for example, lack of a proper KM strategy or lack of specific business objectives.

- Barrier 3: *Technological barriers* are related to technology and tools. In order to use the tools properly and take advantage of all the benefits offered by technology, it is necessary that users have easy access to the tools and feel comfortable using them. Examples of this kind of barriers are: the use of tools is cumbersome or complicated, or a lack of training, familiarity and experience with the tools.

- Barrier 4: *Project management related barriers* are those affecting the proper development of the project. Examples are: lack commitment and involvement of users in the project, lack of suitability of training and reward systems, lack of time and resources for KM activities, or lack of staff with the required technical and business expertise.

- Barrier 5: *Knowledge nature related barriers* refer to each type of knowledge has different features and different management difficulties. For instance, explicit knowledge is easy to find and recognize, and therefore, it is easy to share. However, tacit knowledge is hard to express, and difficult to share. It is this latter type of knowledge that offers more complexity and difficulties to manage it. Examples of such barriers are difficulties in identifying and extracting knowledge, or difficulties in knowledge evaluation.

4.3. W2KM Methodology

The aim of this chapter is to present a methodology that is capable of guiding throughout the process of developing and implementing a KMS 2.0 using Web 2.0 and Big Data tools, while at the same time ensuring the success of the project and reducing its complexity. The W2KM methodology consists of phases that can be broken down into activities, which in turn are made up of tasks. The W2KM methodology uses the traditional phases of an information system project. The difference is in the tasks to carry out in each phase, because these tasks cover all the steps concerning organization, analysis, design, development, control, modification and updating that are needed to carry out a KMS 2.0 project.

The W2KM methodology can help to collect, generate, manage and apply the knowledge generated both inside the organization and from the external relationships of the organization, and then transfer it to the right people easily and quickly. During the development of the W2KM methodology the five barriers to knowledge management defined by Šajeva (2007) were taken into account. Furthermore, it can be applied both by members of staff who work in the organization and by an external consulting service. It is also valid regardless of the number of users and the number of branches the organization has.

The procedure used to develop the W2KM methodology is as follows. Initially a preliminary version of the W2KM methodology was developed based on the authors' previous experience in knowledge management projects and knowledge management 2.0, as well as on the review of the existing literature. A final version was later produced using the case study method. This case study was carried out by applying the preliminary version

of the W2KM methodology in a large oil and gas company, with the aim of using the conclusions from the study to improve it.

The final version of the W2KM methodology is divided into seven phases (similar to those of a classical software development methodology), which can be seen in **¡Error! La autoreferencia al marcador no es válida.**. We have simplified the description of phases and activities in the waterfall model. But, in addition, the processing stages may not be executed sequentially, because the KMS 2.0 development can be split into knowledge blocks. Therefore, a company can decide to follow an iterative model, and carry out series of mini-waterfalls from the analysis phase. Then, all the phases of the waterfall are completed for a knowledge block, since inside a knowledge block development each phase depends on the results obtained in earlier phases. In the following, each phase of the W2KM methodology is described in greater detail.

4.3.1. Phase 1: Draft

The aim of this phase is to study the feasibility of the project for the organization, that is to say, whether it would be in the organization's interests to undertake it and if it is going to be cost effective. It also makes those running the organization aware of the benefits that can be obtained from using Web 2.0 and Big Data tools to manage knowledge, as well as their limitations.

4.3.2. Phase 2: Planning

Several tasks are done in this phase: to achieve the commitment of Management, obtaining a proactive attitude towards the project at the Management level; to set up the project management team, whose members will make the decisions throughout the project, and also the project coordinator and the Community Manager who will have to manage, construct and moderate between the communities that exist within the organization; to establish specific aims by using SWOT analysis and mechanisms of control; to determine the material and human resources that will be needed to carry out the project, that is, the technical human resources that will be in charge of carrying out the different activities and tasks in the project, and the future users who are going to participate in the identification, extraction and codification of the knowledge that the organization wishes to manage; to define the internal communication policy of the KMS 2.0 project, trying the communication flow in both directions rather than just the traditional "from top to bottom"; and a work schedule containing all the tasks that are needed to implement the project is drafted, including the people in charge of each task and the dates on which these have to be carried out. Together with the timeline, it is also necessary to establish the quality control mechanisms and draw up the plan for change.

4.3.3. Phase 3: Analysis

The first thing to be undertaken in this phase is to identify the target knowledge, that is, all the knowledge that the company wants to know about because it is useful to the organization and will provide an added value when utilized. This target knowledge will be processed, generated, stored and distributed by the KMS 2.0. To make it easier to identify it in an organized fashion, it is better to begin by identifying the conceptual blocks of knowledge, which are the basic entities of the organization or of its environment that contain a particular type of target knowledge (Chalmeta and Grangel, 2008). These conceptual blocks of knowledge are different for each type of organization, since such blocks can only be defined by taking into account the strategic objectives of the

organization and its core activities. Examples of them include owners, suppliers and customers, employees, administration and trade unions, organization, products or services, processes, and resources, etc.

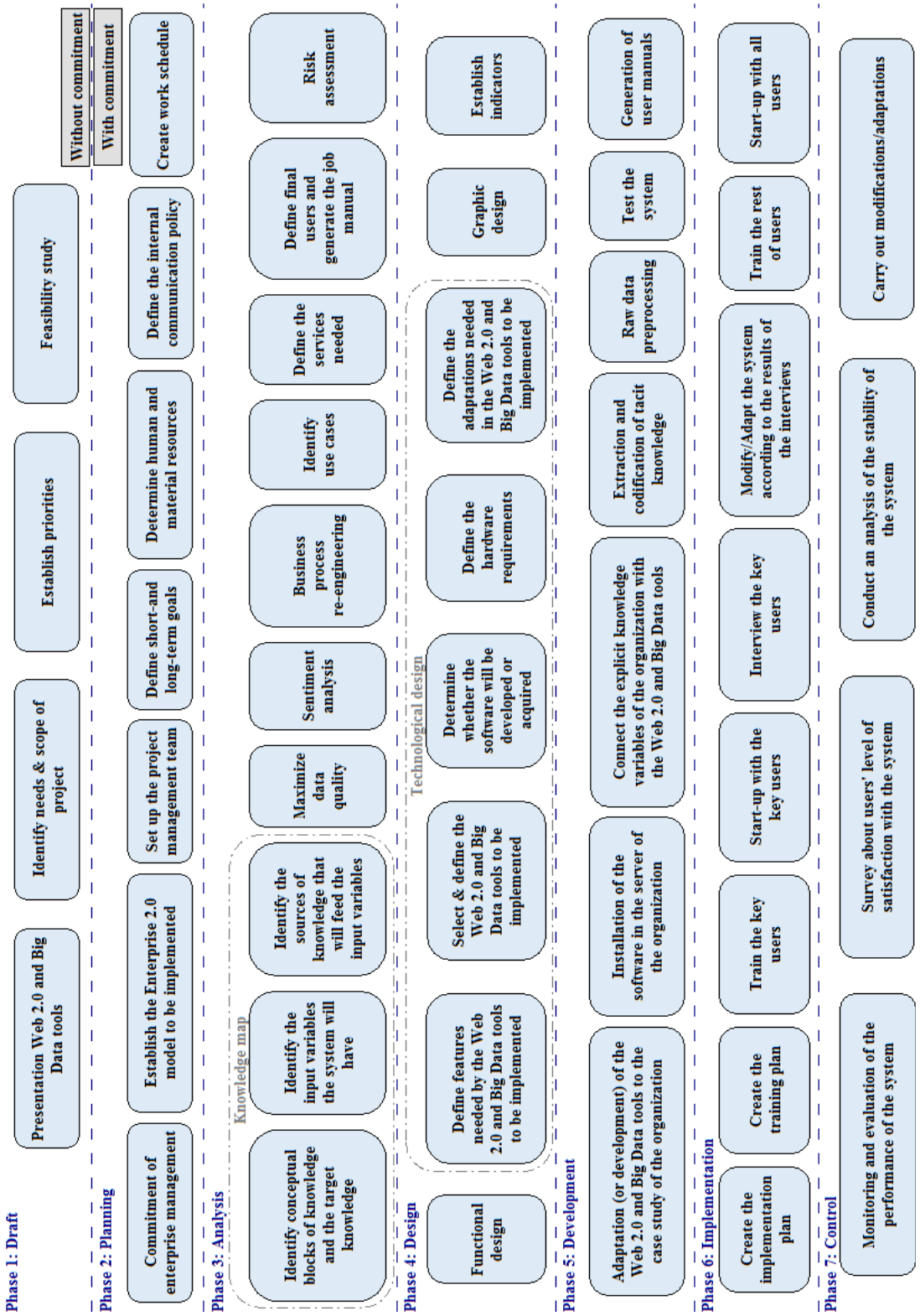

Figure 4.1. Phases and activities in the W2KM methodology.

The next stage is to identify the input variables that will make it possible to obtain the target knowledge inside each knowledge block. These input variables may be data, documents, video, audio, text, posts, etc. and information or knowledge held by people related to the organization. Furthermore, it is also necessary to identify the sources of knowledge (both internal and external to the organization) that will supply the input variables. Sources of knowledge can be tacit, like employees, customers, etc., or explicit, like social media, archived documents, records of stakeholders correspondence, company applications, public Web, machine log data, sensor data, etc. For example, a company launches a new product and is interested in knowing its customers' opinion about the price. In this case, the knowledge block is the product, the knowledge target is the product price, the knowledge source is the customers and the input variables can be, for example, the customers' tweets about the price of the new company product.

In order to take advantage of Big Data tools, it is necessary to ensure the good quality of the data (input variables in this case). This can be achieved by maximizing the following properties (Chiang and Sitaramachandran, 2015): 1) *Existence*, the organization has or can get the data; 2) *Validity*, the data values are within an acceptable range; 3) *Consistency*, the same data have the same values regardless of where they are located; 4) *Integrity*, completeness relationships between data elements; 5) *Accuracy*, the data describe the properties of the model; and 6) *Relevance*, the data are appropriate to achieve the proposed objectives.

The publication of opinions on the Web allows customers to share their point of view about a product or service. These electronic word-of-mouth statements are very important for the organizations, because it is a way to know how customers perceive their products and/or services. Therefore, in the following stage it is necessary to perform sentiment analysis (several techniques can be used, such as Natural Language Processing, Information Retrieval, and structured and unstructured Data Mining), in order to extract and analyze the public's mood and views (Ravi and Ravi, 2015).

The next step is to re-engineer the business processes that need it. This is accomplished by redesigning the work processes taking advantage of the possibilities that the KMS 2.0 offers to optimize them. The understanding of the processes of the organization that is generated as a result of the re-engineering of processes may modify the knowledge map of the organization.

In the following, the use cases, which are actions (access/generate knowledge) that the users will be able to carry out in the KMS 2.0, are identified for each activity in the business processes that is modified as a consequence of the implementation of the future KMS 2.0. Several different use cases can occur in one activity. The services that are needed, which are the capabilities that will be included in the KMS 2.0, are then defined. Then each of the final users that will interact with the system must be identified. To end this phase, an evaluation is performed of the possible risks that can arise and which may prevent the goals of the project from being fulfilled, so that if they do occur, the organization is ready to react.

4.3.4. Phase 4: Design

This is the phase in which the functional, technological and graphic design of the KMS 2.0 is carried out. First of all, the functional design of the Web 2.0 and Big Data tools is defined. For each Web 2.0 and Big Data tool that is going to be used in the KMS 2.0 it is

necessary to specify the way in which the input variables are going to be managed in order to obtain the target knowledge of each conceptual block of knowledge. This includes the procedure of extracting and calculating each variable, language, format of the variables (templates, types of documents, images, etc.), periodicity, norms of conduct, standards of development, and so forth. Furthermore, the format of all the different types of electronic documents and data that each Web 2.0 and Big Data tool will work with must also be defined. It is important to consider that Big Data tools also need to manage human information, which is characterized by being complex, unstructured, ubiquitous, multi-format, and multi-channel.

The technological design is then carried out. To do so, first the characteristics of the Web 2.0 and Big Data tools that are going to be implemented must be defined. After that, a decision must be made as to whether the software will be custom built or if the (commercial or free-distribution) application will be acquired and later tailored to meet specific needs. The next step is to define the hardware requirements as regards both the server where the Web 2.0 and Big Data tools are installed and the terminals to be made available to users and for communications. Lastly, an analysis is conducted of the modifications that must be made to the Web 2.0 and Big Data tools so that they cover the organization's needs. The application interface, which is the link between the capabilities of the application and the user, is then designed. The graphic design must be ergonomic, intuitive and in line with the message that the organization wishes to transmit. Finally, indicators are established for each Web 2.0 and Big Data tool with the aim of managing them in an efficient way.

4.3.5. Phase 5: Development

In this phase, the Web 2.0 and Big Data tools are installed, developed/customized and tested, and the corresponding user manuals are produced. First, the Web 2.0 and Big Data tools are adapted (if they are acquired either as commercial or open-source applications) or developed (if they are custom-built). The final graphic appearance has to be effective and both allow and foster interaction and collaboration among users. It also has to comply with the fundamental characteristics of Web 2.0 tools discussed in the literature review section as regards functionality, quality, accessibility and sociability.

Once the KMS 2.0 has been developed and installed in the organization's server, it has to be filled with some initial contents that will later be expanded with the new knowledge provided by the final users during the course of their day-to-day work. Therefore the initial structure of the KMS 2.0 will be evolving dynamically due to the users' interactions. The new knowledge is obtained by processing the input variables. To do so, on the one hand, the explicit variables of the knowledge of the organization must be linked automatically with the corresponding Web 2.0 and Big Data tools. On the other hand, all the input variables within tacit sources (Nie et al., 2010) must be connected, extracted, codified, and parameterized. The new knowledge will be distributed through the Web 2.0 tools or the company/external computer systems.

Sometimes raw data from certain sources needs to be preprocessed previously, in order to be analyzed properly. There are several preprocesses that can be performed on the raw data, such as parts of speech tagging, tokenization, stemming, stop-word removal, and feature extraction and representation (Ravi and Ravi, 2015).

Figure 4.2 shows the computer framework proposed to support the KMS 2.0. The framework is composed of four modules: Content, Transfer, Enrichment, and Decision Support. In this framework the flow of information is cyclic because the producer of knowledge (the knowledge source) can be also consumer of processed knowledge (for example, social networks users). The role of each module is explained below.

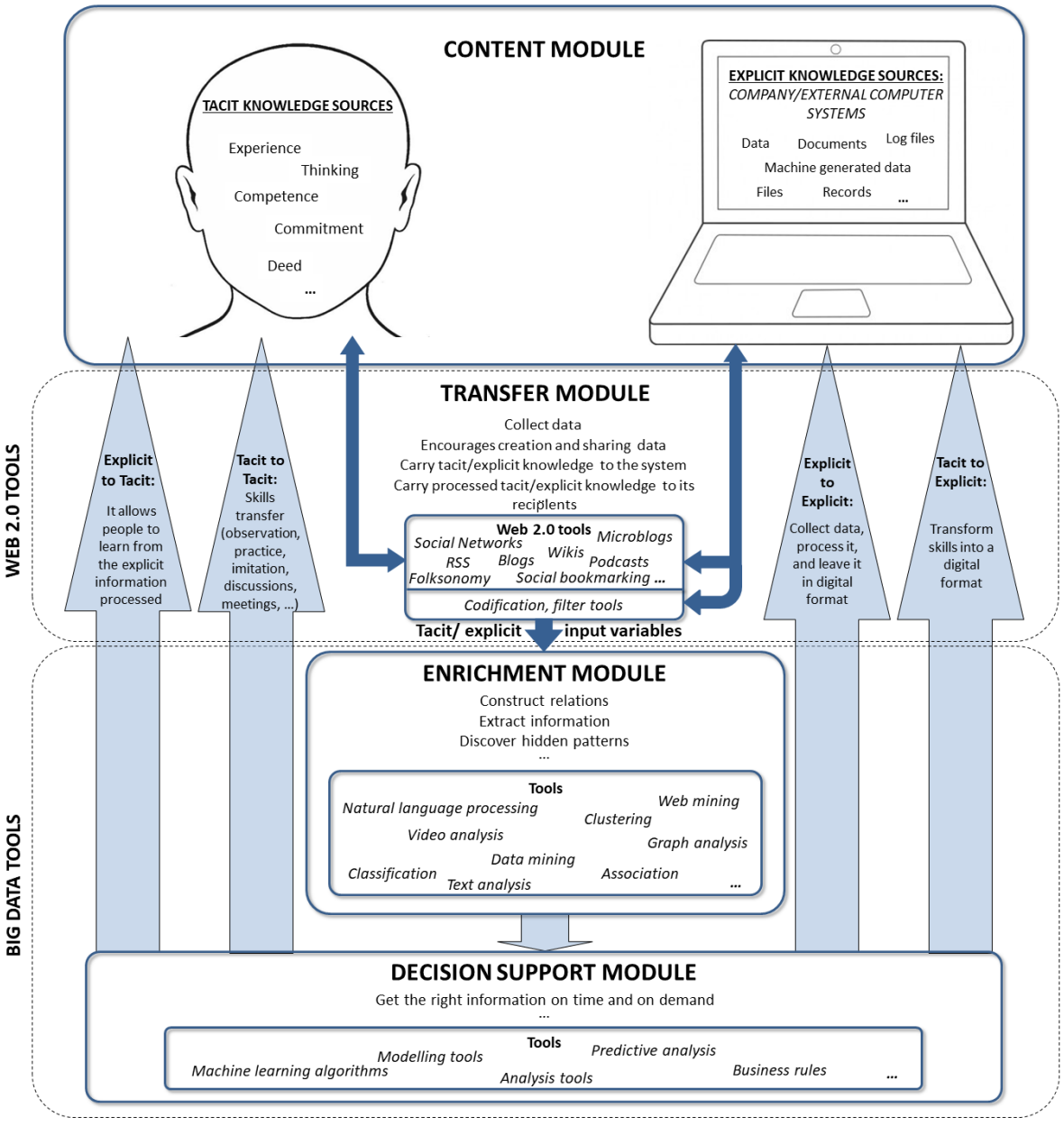

Figure 4.2. KMS 2.0 Computer Framework proposed.

Content Module includes the different types of knowledge sources, both tacit and explicit. These knowledge sources can also be consumers of the knowledge generated by the KMS 2.0.

Transfer Module is based on company/external Web 2.0 tools, and it works as a collector of the *raw material*. This *raw material* can be experiences, feelings, opinions, etc. that the different tacit knowledge sources of the Content Module introduce in the Web 2.0 tools (in

KMS 2.0, the codification of tacit knowledge is performed by the sources themselves), as well as data, documents, logs, etc. stored in company/external computer systems of the Content Module that the company links automatically to the Web 2.0 tools or directly to the codification and filter tools of the Transfer Module. Then, that collected *raw material* is filtered in order to identify only the tacit and explicit input variables needed to generate the target knowledge. Finally, the tacit and explicit input variables are inserted in the Enrichment Module to be processed and stored. Therefore, as these input variables are stored, they become permanent, even if they are removed from their original sources.

Enrichment Module is responsible for the processing of the input variables supplied by the Transfer Module and the generation and storage of new knowledge. Input variables are organized and analyzed using Big Data tools and machine learning techniques.

As data are acquired and generated from different sources and formats (video, text document, audio, image, etc.), they are analyzed separately based on their format using their corresponding machine learning algorithms. For instance, data coming from social media channels may be analyzed by using text mining, sentiment analysis, natural language processing (NLP), and so on, to manage and categorize human information. The use of machine learning techniques in this module allows the system to discover hidden patterns, extract meanings and relevant information, categorize or classify information from each individual source.

After data from the different sources have been analyzed and categorized, they are aggregated and integrated to create the new enriched metadata sets. That is, the newly discovered knowledge is used to add value to the original data. As a result, the enriched metadata will contain information about several topics, opinions, likes, reviews, features, etc. Although data from each source can provide useful insights by themselves, the combination of data from the different sources (enriched metadata) may help to significantly improve the performance of the prediction models in the next stage (Decision Support Module).

Decision Support Module is also based on a Big Data platform and contains machine learning algorithms. The main objective of this module is to support decisions, by generating new knowledge from the information obtained in the Enrichment Module using supervised and unsupervised prediction models, such as decision trees, logistic regression, artificial neural networks, clustering, etc. The suggestions of the prediction models are combined with business rules to support/generate the final decisions.

The use of these machine learning algorithms makes it possible to discover new trends and insights on data, examine new business opportunities, find inefficiencies in order to improve or innovate in services or products, etc. based, for example, on user preferences, wishes, actions, behavior, etc.

The new knowledge generated in this module is distributed through the Web 2.0 tools of the Transfer Module and other company/external systems (for example, to act directly on advertising banners, personalize advertising, add special offers, etc.), in order to provide the knowledge that each consumer needs when they need it.

In the case that the producer (knowledge source) and consumer of the knowledge are the same, four possibilities can happen, according to Nonaka's theory of organizational knowledge creation (Nonaka, 1994):

- *Produce Tacit Knowledge and consume Explicit Knowledge*. People learn (acquires tacit knowledge) from the appropriate explicit knowledge processed by the system.
- *Produce Tacit Knowledge and consume Tacit Knowledge.* System allows that Tacit Knowledge like skills are transferred and learned, by observation, through practice, by imitation, etc.
- *Produce Explicit Knowledge and consume Explicit Knowledge*. Explicit knowledge is collected and processed by the system, and then the processed knowledge is also expressed explicitly.
- *Produce Explicit Knowledge and consume Tacit Knowledge*. Tacit Knowledge is transformed explicitly into a digital format.

4.3.6. Phase 6: Implementation

In the implementation phase, the system is put into operation. At first the system is only used by a subset of the final users (called "key users") of each Web 2.0 and Big Data tool. The purpose is to take advantage of their own experiences or impressions to debug and refine them. The key users of a tool are the people in charge of implementing, customizing and debugging it. They are also responsible for solving all the basic issues or doubts about the tool that any user may have (more important problems are solved by the Community Manager). It is then put into operation with all the final users.

The first step is to create an implementation plan that identifies all the users involved and indicates, for each of the Web 2.0 and Big Data tools to be implemented, the dates on which each of them will be put into operation, with both the key users and all the other users. Then the training plan is created.

4.3.7. Phase 7: Control

The control phase spans the whole useful life of the system. In this phase the system is monitored and, if necessary, capabilities are adapted or modified to solve errors and improve the system or adapt it to new ways of working in the organization.

4.4. Case Study

A case study was carried out by applying the methodology in a large oil and gas company, in which the qualitative data that were collected were submitted to an inductive analysis. The work plan that was followed in order to carry out the case study was based on Runeson and Höst (2009) and consists of five phases: *Case study design and planning*, *Preparation for data collection*, *Collecting evidence*, *Analysis of collected data*, and *Validation of collected data*. The results obtained in each phase are outlined in the following section.

What brought the enterprise to undertake such an implementation was mainly the need to gather and organize the knowledge of its customers and employees, as well as to take advantage of the large amount of data, both internal and public, that are of interest to the enterprise. They also aim to optimize the search for knowledge for their employees, so that, besides placing the knowledge they need at their disposal, they can also access it quickly and easily, thereby minimizing the time invested in getting it. Moreover, regarding

the knowledge from external sources, they intend to keep it stored in a database located in a local server, in order to be able to access the knowledge faster and when they wish, regardless of the availability of such information at the source of origin at the time it is needed. In order to achieve this, it is necessary to gather and organize all the available information (both public and private) to allow the company to take the right decisions in the shortest possible time.

4.4.1. Case Study Design and Planning

The time span of the case study was set to 12 months. The goals of this case study were: (a) to test the methodology developed to guide the implementation of Web 2.0 and Big Data tools for managing knowledge, while also verifying and confirming its usefulness, rigor and quality; (b) to analyze the result in order to determine the improvement offered by the methodology; and (c) to refine and improve the initial methodology with the aid of the experience gained and the conclusions drawn from the case study.

Two research questions, which will be examined while the case study is being carried out, were formulated: (a) Are Web 2.0 and Big Data tools suitable for managing the knowledge in this organization? and (b) Does this methodology facilitate the development and implementation of a KMS 2.0 in this organization?

4.4.2. Preparation for Data Collection

To apply the W2KM methodology, a mixed work team was set up with members coming from both the IRIS Research Group and the company Knowledge Project Management Team (KPMT). This KPMT was made up of three members of the Knowledge Management Department, who were a manager, a technician and a freelance consultant that had worked for the company for many years (who would also be the Project Coordinator), and three other representatives of the company, one from each of the following areas: Internal Communication Area, Intelligence and Investigation Area, and Marketing Area. The purpose of the KPMT was to be responsible for making decisions about all issues related to the work.

The data used to conduct this case study are qualitative and were collected by both direct and indirect methods. The direct method that was used consisted in interviews, where data are collected in real time and, additionally, the interviewer (one member of the IRIS Research Group) is in direct contact with the interviewees. The indirect method involved the analysis of different documents used in the enterprise. Moreover, the data obtained from the interviewer's observations were also taken into account.

4.4.3. Collecting Evidence

Data were collected through interviews after the execution of each phase of the methodology, using an assortment of questionnaires and templates, as well as copies of the documents and reports utilized in the enterprises. After each interview was completed, the answers were reviewed by another researcher from the IRIS Research Group, which provided another point of view. The reason for conducting the interviews after the execution of each phase was to be able to solve any problems and/or apply the improvements identified before starting the execution of the next phase. The interviews were carried out with each member of the KPMT, they lasted about 30 minutes, and they were individual, open (allowing interviewees to give any response) and semi-structured (the questions were used as a guide, not to be asked in that same order, and both the

interviewer and the interviewee were allowed to improvise). The objectives of the interviews at each stage were: to analyze the execution of the phase, to detect errors and problems encountered, to obtain feedback from the experience of the interviewees, and to collect proposals for improving the W2KM methodology. The questions asked in the interviews were the same for each interviewee, but were different for each phase, as they were adapted to the specific characteristics of each phase.

Finally, when the methodology had been fully applied and the KMS 2.0 was implemented and working properly, a questionnaire was distributed among the KPMT members and the key users in order to analyze their impressions about the features of the KMS 2.0.

The following subsections outline some of the more significant results obtained from applying the W2KM methodology to the enterprise in which the case study was conducted.

4.4.3.1. Phase 1: Draft

The Web 2.0 and Big Data tools were presented to four representatives from the enterprise, with special emphasis on the features, advantages and disadvantages of each one. The Web 2.0 tools that were proposed to cover the enterprise's needs were the Wiki, the Blog and the Social Network. Regarding the proposal for Big Data tools, the IDOL software and the Hadoop ecosystem were selected.

4.4.3.2. Phase 2: Planning

The enterprise representatives decided that the KPMT would be made up of three members of the Knowledge Management Department, who were a manager, a technician and a consultant plus the representatives of the Internal Communication Area, Intelligence and Investigation Area and Marketing Area. In addition, four analysts/programmers would be responsible for the required configurations and parameter settings. The cross-functional Enterprise 2.0 model was therefore followed. The enterprise representatives also defined the internal communication policy and they told everyone involved in the project to give high priority to the tasks related to it.

4.4.3.3. Phase 3: Analysis

When they were to identify the conceptual blocks of knowledge, which are the basic entities of the organization or of its environment that contain the knowledge the organization is interested in managing (Chalmeta and Grangel, 2008), the following blocks were found within the scope of the case study project: Internal Communication Area, Intelligence and Investigation Area, and Marketing Area. The target knowledge that had to be managed by each conceptual block of knowledge was (a) Internal Communication Area: Internal documents; Employee information; Project information; Customer information; Working procedures; Information about the competences in each job; Notice board for suggestions and opinions from employees; Internal collaboration and solving doubts; Internal knowledge sharing. (b) Intelligence and Investigation Area: External Media information; External Social Media information; Security risk level classification by country; Security investigation reports. (c) Marketing Area: Information about competitors; Comparison of the prices of products and services; Customers' characteristics; Evaluation of customers; Reviews on public Social Media regarding products and enterprise image; Sector innovations.

The first four columns in Table 4.1 show the target knowledge, the input variables, the sources of knowledge, and the Web 2.0 or Big Data tool chosen to manage the knowledge from the Intelligence and Investigation block. The sources for External Media information

and External Social Media information were selected by members of the Intelligence and Investigation department based on their needs. Initially 54 External Media information sources and 2 External Social Media information sources (Facebook and Twitter) with 62 different filters (accounts, hash tags, keywords, etc.) were selected. The company aims to increase the linked sources in the future. The sources for External Media information generated an average of 489 documents per day. The sources for External Social Media information generated an average of 963 documents per day. Considering the entire project, 146 data sources were linked, and they generated an average of 2254 documents per day.

Table 4.1. Analysis and design of the knowledge from the Intelligence and Investigation Area.

Target knowledge	Input variables	Sources of knowledge	Web2.0/Big Data tools	Users (Permissions)	Format of the variables
External Media information	Web documents	Subscriptions to newspapers and other external organizations	IDOL	All department users (Read)	HTML documents
	RSS documents				HTML documents
	Video files				Any video format
	IMAP messages				E-mail texts
External Social Media information	Publications on Twitter	Twitter API	IDOL	All department users (Read)	Text published on Twitter
	Publications on Facebook	Facebook API			Text published on Facebook
Security risk level classification by country	Analysis of documents ingested in IDOL	Documents ingested in IDOL	Wiki	All department users (Write); All area managers (Read)	Text with graphics published on the Wiki
Security investigation reports	Analysis of documents ingested in IDOL	Documents ingested in IDOL	Blog	All department users (Write); User requesting the investigation (Read)	Text with graphics published on the Blog

With regard to the re-engineering of business processes, the processes affected by the KMS 2.0 were analyzed and some of them were modified. The main processes modified were: Analysis of information about competitors; Analysis of the prices of products/services; Analysis of products/services and company image in public Social Media; Ingestion of external Media information; Ingestion of external Social Media information; Analysis of customers; and Internal content management.

The services that the final users could carry out in the KMS 2.0 were then defined. For example, some of the services that it had to offer included: Writing, modifying and reading an article; Attaching a document to an article; Commenting on an article; Making a comment to a user; Posting important events on a calendar; Consulting active users; Tagging a file; Printing the articles on paper or in PDF format; Consulting the values of the indicators; Controlling for correct use and vandalism; and Reviewing the proper ingestion of the content of the feeds.

After defining the services, the profile of users that could have access was established together with the corresponding permissions. Column five in Table 4.1 shows an example of the permissions that were set for each input variable of the Intelligence and Investigation knowledge block.

4.4.3.4. Phase 4: Design

The functional design also included defining the format of the variables with which the KMS 2.0 would be working. Column six of Table 4.1 shows this format for the block of Intelligence and Investigation knowledge. In addition, it was decided that the KMS 2.0 would run in an open environment, so that the employees could access it both from within the facilities of the enterprise and from anywhere else in the world.

The features of the Web 2.0 and Big Data tools to be implemented were defined in the technological design. For example, in the case of the IDOL system, it was agreed that it should have: the standard capabilities of IDOL; Sentiment analysis; Social Media connector; IMAP connector; HTTP connector; RSS connector; Eduction; Conceptual search; and Automatic categorization.

With regard to the software to be implemented, the decision was made to acquire IDOL software, and also to acquire other free-distribution software and customize it to meet the needs of the enterprise.

Both the indicators and the frequency of the KMS 2.0 were established. Some of the indicators that were defined in order to control the use of the KMS 2.0 were: Number of accesses (daily); Number of new articles published (daily); Percentages of accesses for reading, modification and creation (weekly); Number of comments made (weekly); Number of doubts published (weekly); Percentage of doubts settled (weekly); Number of files tagged (weekly); Users' satisfaction (quarterly); and Number of IDOL queries (weekly). The indicators per user were: Number of accesses (daily); Number of new articles published (daily); Number of files tagged (weekly); Percentages of accesses for reading, modification and creation (weekly); Number of comments made (weekly); Number of doubts or problems proposed to other users (weekly); Number of doubts or problems from other users that have been settled (weekly); and Number of investigation reports generated (weekly). Once all the indicators had been defined, a reference value was established for each of them so as to allow their results to be evaluated.

4.4.3.5. Phase 5: Development

The Web 2.0 and Big Data tools were installed and customized. The explicit variables were then connected and the tacit variables were extracted, codified, parameterized and connected, and the system was tested. After that, the final integration trials were conducted by inserting fictitious data over the backup that was created for the tests, with the aim of ensuring that everything would work properly. Lastly, the KMS 2.0 user and administrator manuals were produced.

4.4.3.6. Phase 6: Implementation

In this phase both the implementation plan and the training plan were created; the latter told users who were going to have specific training, when they would receive it, and the syllabus that would be taught. In this case, one training course was given for the key users, who were 10 users in all: 5 from the Internal Communication Department, 3 from the Marketing Department, and 2 from the Intelligence and Investigation Department. The

course lasted 4 hours and was given in two 2-hour sessions. For the remaining users, manuals were created to guide them in the use of the system.

After the key users had used the system, they realized the possibilities that Web 2.0 and Big Data technology could provide to their company. In the interviews that were carried out, these users identified different business processes of several departments that could also benefit from this technology, and which are listed in Table 4.2. Some of those business processes were modified and/or adapted following the instructions of these users, and as a result, a high degree of optimization was obtained. The remaining business processes that were identified will be improved in future projects, as they were not included within the project scope of the case study.

4.4.3.7. Phase 7: Control

This phase is carried out while the KMS 2.0 is actually working. It was established that the Community Manager has to monitor the system every week to ascertain the performance of both the KMS 2.0 and its users. This is accomplished by comparing the value obtained in the indicators with that of the reference criterion of each of them. Once the indicators have been evaluated, a report is drafted and submitted to the appropriate managers of the enterprise, who then make suitable decisions based on that information.

Every quarter, the Community Manager gives all the users a survey about their level of satisfaction as regards the KMS 2.0, in order to determine the degree of acceptance as well as to gather proposals for improvement. The survey is anonymous and voluntary for users. The information thus obtained is studied and, if necessary, appropriate modifications are made. The Community Manager also conducts an analysis of the stability of the system every quarter, which involves reviewing the KMS 2.0 and carrying out tests to ensure that it is working properly and that it is also being used properly. A system of rewards was also set up for users who participated on an active basis. This reward system is modified every three months.

All the users of the KMS 2.0 were told that as soon as they detected an error or identified a proposal for improvement they should inform their superiors, who would notify the Community Manager so that he or she could study them and carry out appropriate modifications if needed.

4.4.4. Analysis of Collected Data

After each interview with the members of the KPMT, the answers were compiled and analyzed. Most of the comments were positive, noting that the W2KM methodology guided them in all the steps required for each phase, and that the implantation was faster and more comfortable than other implantations in computer system projects in which they had previously participated. They also indicated that the methodology allowed them to better identify the needs, consequences, scope and opportunities of the project. Among the main points that they highlighted were the fact that they realized the importance of data quality in order to achieve an optimal system performance, and the ease with which they could re-engineer existing business processes. They were also amazed with the amount of information that they could use and were not exploiting so far. Moreover, once they knew the potential offered by Web 2.0 and Big Data technologies, they thought about more ideas for future projects. Not all the comments were positive, but negative comments were considered to improve the methodology. Examples of this type of comments can be the need to identify the profiles of end users of the system, or the need to deliver a presentation

Table 4.2. Business processes identified in the case study that can be improved with the KMS 2.0.

Department	Business process
Strategic management	Real-time analysis of the competitive environment
	Detection of changes in the competitive environment
	Data-driven decision-making
	Strategic planning
Purchasing	Identification of suppliers
	Investigation of suppliers
	Gathering information about products and/or services
Operations	Troubleshooting in the services offered
	Increasing the quality of the services offered
	Offering an efficient catalogue of services based on sales trends analysis
Research and Development	Monitoring the performance and quality of the services offered
	Monitoring of scientific publications
	Detection of scientific advances for topics of interest
	Monitoring the granting of patents
	Production of technological maps and scientific publications
	Acquisition of technical knowledge applied to products and services
	Innovation and process improvement
	Identifying the needs of customers of new services
	Identification of improvements in the services offered
Marketing / Sales	Analysis of customer information
	Identification of potential customers
	Identification of the most valuable customers
	Analysis of competing companies
	Gathering information about customers' needs
	Research about the company image
	Service acceptance analysis
	Monitoring social networks
	Price monitoring
	Detection of new releases by competing companies
	Analysis of relations in social networks
	Predicting customer behavior
	Accurate prediction and awareness of customers' needs
	Making real-time customized offers
	Encourage participation and interaction in every channel
	Quick reaction to market opportunities
	Analysis of sales trends
Finance	Risk measurement
	Improvement of budgeting and forecasting
Human Resources	Conducting investigations about employees
	Conducting investigations about potential employees
	Staff selection
	Monitoring of employees at work (through their computer and mobile)
	Detection of applications that are most used by each employee
	Detection of misuse of applications by each employee
	Investigation of what time the employees are most productive
	Discovery of teamwork patterns
	Predicting when employees are undergoing periods of stress that affect their productivity
	Identification of the leaders
	Employee retention
	Analysis of the effectiveness of recruitment campaigns
	Measurement of employee morale
Customer assistance	Identifying customers who are at risk of ceasing to be customers of the company
	Analysis of how customers use the company website
	Monitoring how customers use the services offered by the company to detect potential problems and/or improvements
Security	Performing security investigations
	Improvements in intelligence and surveillance
	Forecasting and mitigating real-time cyber attacks
	Crime prediction and prevention

to the managers in the first phase of the methodology, explaining the potential offered by Web 2.0 and Big Data tools, so that they could understand the benefits that these tools can offer to the company.

Once the case study had finished, the questions posed in the case study planning and design phase could then be answered:

a) Are Web 2.0 and Big Data tools suitable for managing the knowledge in this organization?

   Yes, they are. Both Web 2.0 and Big Data tools have some very interesting features that make them excellent candidates for application to knowledge management, since they make it possible to collect the different types of existing knowledge (both tacit and explicit) while also fostering the generation of new knowledge. Due to the characteristics of the information managed by organizations, they can benefit greatly from the opportunities offered by those tools. Big Data tools can extract knowledge from a large amount of structured and unstructured data that are generated by businesses. Moreover, Web 2.0 tools for use by private individuals are viewed favorably by Internet users and, as has been seen in the case study, this implies that they can also be well accepted in organizations and users can adapt to them quickly. This is especially true in the case of the younger employees, who are more likely to use this kind of tools for their own particular purposes. Thus, the users' attitude in this case study was far more positive and collaborative than in implementations of other types of tools carried out by the same researchers.

b) Does this methodology facilitate the development and implementation of a KMS 2.0 in this organization?

   Yes, it does. The results obtained in the case study were satisfactory, as all the goals set out at the beginning were achieved, and the time span initially established was accomplished without deviations. The development and implementation of the KMS 2.0 were swift and straightforward. Additionally, the members of the KPMT indicated that the methodology allowed them to have greater control over the project implementation, since it clearly defines all the steps that need to be carried out in each phase of the project. Furthermore, those responsible for the implementation did not need to be experts in Web 2.0 and Big Data tools or in knowledge management because the methodology provided them with detailed guidance at each step in the process.

Additionally, the members of the KPMT highlighted the following benefits resulting from the use of KMS 2.0 in the company:

- Centralization of the knowledge of the enterprise in an accessible and easy-to-use system, which helps to keep it flowing steadily.
- Fast and efficient settlement of doubts among members of the enterprise with the involvement of as few employees as possible.
- Fast and efficient communication, using the knowledge network that enables users to communicate in a fast, straightforward manner.
- Less time spent on meetings.

- Record all the doubts or problems with the solutions that were adopted so that they can be consulted in the future; hence, when new problems arise, they will take less time to solve it.
- Employees access the information they need when they need it.
- Access is gained to knowledge that remained hidden and uncoded, through the directories of corporate knowledge that are generated, which identify, classify and codify existing internal skills.
- Reduction in the number of internal e-mails sent.
- Knowledge generation from external Media and Social Media information.
- Sending real-time alerts about certain information received (both internal and external).

Hence, the KMS 2.0 that was implemented covers the three basic capabilities that, according to Alavi and Leidner (2001), a KMS must have and which have been described earlier in the introduction section.

Finally, in order to check that the KMS 2.0 that results from the application of the W2KM methodology possesses the fundamental features to be able to manage knowledge (Dai et al., 2007) and that the methodology also helps to overcome the different barriers of a KMS 2.0 project that were identified by Šajeva (2007) – both cited in the literature review section – a survey was conducted among the KPMT members and the key users six months after the launch of the system implemented in the case study. This survey transformed the features and barriers into questions that made it possible to measure that the extent to which the users thought that the KMS 2.0 implemented in the company had these features and that the barriers had been overcome.

Table 4.3 (appendix) shows the average of the values obtained in the survey. The values assigned for answers were: 1 = Completely Disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; and 5 = Completely Agree. As can be seen from the results in Table 4.3 (appendix), only one question obtained an average result lower than 3, which was “The time allocated to the system in the project is adequate”. As a measure to solve this, the company undertook the commitment to ensure that more time will be allocated in the projects for the management of this system.

4.4.5. Validation of Collected Data

Because the data collected were qualitative, they were analyzed using qualitative data methods of analysis. The analysis was inductive and was carried out in parallel to data collection, as it was performed after finishing each of the phases of the W2KM methodology. The purpose of this was to be able to react quickly to the problems and improvements encountered during the analysis of each phase and thus solve each of the problems and take advantage of these improvements before starting the following phases.

Threats to the validity of the case study were reduced by using the Lincoln and Guba model (Robson, 2002). This model proposes five strategies to be used in the collection of data to deal with three types of threats to validity. The three types of threats considered were reactivity (the researcher’s presence can affect the setup of the study), researcher bias

(the researcher's preconceived ideas can affect the way the researcher asks questions or interprets answers) and respondent bias (the researcher's influence on the attitude of the people being studied) (Karlström and Runeson, 2006).

With regard to the five possible strategies, in the present case study they were considered in the following way in order to make the results valid: (1) Prolonged involvement: the researcher is familiar with the environment being studied (in this case study, the researchers and the company had already been collaborating in previous projects). (2) Triangulation: the application of several methods in the study of a single object. In this case study, four types were considered: (i) Spatial triangulation of data (three sources of data were considered: observation, interviews and documentation); (ii) Personal triangulation of data (all the members of the company KPMT were interviewed in order to obtain information from each of them); (iii) Investigator triangulation (the interviews were conducted by a researcher and reviewed by another researcher); and (iv) Theoretical triangulation (the different points of view of the members of the KPMT were taken into account). (3) Member checking: obtaining feedback from the people who are interviewed (in the case study, after each interview, a report containing the relevant information from the interview was checked by each interviewee). (4) Negative case analysis: attempting to find another explanation that differs from the one initially assumed for the observed phenomenon (here, the researchers were working separately [investigator triangulation]). (5) Audit trail: keeping a record of all the documentation of the project so as to make it available in the future.

4.5. Conclusions

In this chapter the authors have presented a methodology, based on recent achievements of theoretical references and related models, which helps to develop and implement a KMS 2.0. The methodology has been tested and debugged with a real-life case study. The findings demonstrate the effectiveness of the proposed methodology in terms of efficiency and effectiveness. Therefore, the KMS obtained from the application of the methodology possesses the fundamental features to be able to manage explicit and tacit knowledge, and the methodology also helps to overcome the different barriers of a KMS 2.0 project. Our research contribute to the science body of knowledge on KM using Big Data and Web 2.0 tools, a novel and rapidly expanding field, where in terms of methodologies there is a need for more experimental studies as well as theory-based research (El Ouirdi et al., 2015), and in terms of knowledge it is necessary investigate how tacit knowledge can be created and shared using Web 2.0 technologies (Antonius et al., 2015).

The findings are useful for practitioners, who will be able to benefit from a series of advantages that cannot be gained by using previous KM methodologies, such as better planning and management of the project, better definition of the vision and strategy of the project, choosing the most suitable Web 2.0 and Big Data tools, and an estimation of the potential benefits to be achieved, as well as a higher probability of being successful.

Big Data tools are a very powerful way to clean and process large amounts of data to generate knowledge, since there is a lot of hidden knowledge in the Big Data that could be considered tacit knowledge. Users are accustomed to using Web 2.0 tools for their own personal purposes in an unregulated way. Yet, in the business setting this philosophy must not be applied as it stands because it would have a negative effect on the performance of the Web 2.0 tools for managing knowledge, since we would not be optimizing it to the full

extent of its possibilities. The knowledge could be diffuse, difficult to find and control, and so on, yet with the methodology the knowledge of the enterprise can be structured and stored, while also allowing new knowledge to be channeled into the most appropriate Web 2.0 and Big Data tools. But at the same time it also lets users employ each tool freely within the area previously established by the enterprise.

It is important to state the limitations of the study, which are related with the qualitative research methodology of a case study. Since this is a case study in which the methodology was applied to a single organization (an oil and gas company), its validity has not been tested in other kind enterprises or sectors, like manufacturer enterprises. Moreover, although the qualitative data of the case study were complemented by quantitative data, no statistical significance could be obtained given the small sample size. Therefore, the benefits obtained by the company from applying the methodology were not measured objectively because they are achievements that are perceived by the people involved in the implementation. Nevertheless, their experience and professionalism lead us to trust in the honesty of their claims regarding those achievements. This is an important limitation because a single case study is not good for generalization purposes, due to the heterogeneity of companies. Therefore comparative studies of multiple case that maximize the variation of companies (each with potentially many observations) can increase the possibilities of validate the useful of the W2KM methodology for other kind company or other kind of Web 2.0 or Big Data tool.

As regards possible lines of work in the future, some of the challenges related with the W2KM methodology that have still to be dealt with are: adapting the W2KM methodology to the peculiarities of KM 3.0 tools. KM 3.0 tools are semantic tools that improve access to information and reuse the knowledge in semantic Wikis (Oren et al., 2006) and semantic Blogs (Cayzer, 2004). They can also use the Web as a source for knowledge acquisition (Java et al., 2007), and they are able to recycle data and transform it into explicit knowledge (Kohn et al., 2010). In addition, they also interconnect people and content in a significant way using semantic Social Networks (Breslin and Decker, 2007). Another possible line of work to be followed in the future is to solve the problem of semantic interoperability, since it is essential that both senders and receivers interpret the knowledge in the same way (Brannen and Wilson III, 1996). Including ontologies (Boissier et al., 2013) in the W2KM methodology may be a solution to this issue. Finally, the W2KM methodology could address the problems related to the protection of data and the security of information so that sensitive data about the enterprise is not disclosed.

4.6. References

Abramowicz W, Fensel D, Frank U (2010). Semantics and Web 2.0 Technologies to Support Business Process Management. Business & Information Systems Engineering, 2(1):1-2.

Alavi M, Leidner DE (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. MIS Quarterly, 25(1):107–136.

Amine M, Ahmed-Nacer M (2011). An agile methodology for implementing knowledge management systems: a case study in component-based software engineering. Software Engineering Applications, 5(4):159-170.

Antonius N, Xu J, Gao X (2015). 'Factors influencing the adoption of Enterprise Social Software in Australia, Knowledge-Based Systems, 73:32-43.

Antonova A, Gurova E, Nikolov R (2009). Extended architecture of knowledge management system with Web 2.0 technologies. Proceedings of the ECKM 2009, 10th European Conference on Knowledge Management, Italy, pp.48-55.

Bebensee T, Helms R, Spruit M (2011). Exploring Web 2.0 Applications as a Mean of Bolstering up Knowledge Management. The Electronic Journal of Knowledge Management, 9(1):1-9.

Blackler F (1995). Knowledge, knowledge work and organizations: an overview and interpretation. Organization Studies, 16(6):1021–1026.

Boissier O, Colombetti M, Luck M, Meyer JJ, Polleres A (2013). Norms, organizations, and semantics. Knowledge Engineering review, 28(1):107-116.

Brannen MY, Wilson III M (1996). Wilson recontextualization and internalization: lessons in transcultural materialism from the Walt Disney enterprise. CMES Business Review, 1(1/2):97-100.

Breslin J, Decker S (2007). The future of social networks on the internet: the need for semantics. IEEE Internet Computing Magazine, 11(6):86–90.

Brown JS, Duguid P (2000). The Social Life of Information. Boston: Harvard Business School Press.

Cayzer S (2004). Semantic blogging and decentralized knowledge management. Communications of the ACM, 47(12):47-52.

Chalmeta R, Grangel R (2008). Methodology for the implementation of knowledge management systems. Journal of the American Society for Information Science and Technology, 59(5):742-755.

Chiang F, Sitaramachandran S (2015). A Data Quality Framework for Customer Relationship Analytics. Web Information Systems Engineering, Springer, Part II, 9419:366-378.

Dai Y, Gao Q, Fan Z, Kang R (2007). User perceived quality of online social information services: From the perspective of knowledge management. 2007 IEEE International Conference on Industrial Engineering and Engineering Management, Singapore, pp.482–486.

Day RE (2001). Totality and Representation: A History of Knowledge Management Through European Documentation, Critical Modernity, and PostFordism. Journal of the American Society for Information Science and Technology, 52(9):724-735.

Davenport TH (2005). Thinking for a Living: How to Get Better Performance and Results from Knowledge Workers. Harvard Business School Press.

Debortoli S, Müller O, VomBrocke J (2014). Comparing Business Intelligence and Big Data Skills. A Text Mining Study Using Job Advertisements. Business & Information Systems Engineering, 6(5):289-300.

Dutta K (2015). Editorial Special issue on software architectures and systems for Big data. Journal of systems and software, 102:145 pages.

El Ouirdi A, El Ouirdi M, Segers J, Henderickx E (2015). Employees' use of social media technologies: a methodological and thematic review, Behaviour & Information Technology, 34(5):454-464.

Erickson S, Rothberg H (2014). Big Data and Knowledge Management: Establishing a Conceptual Foundation. The Electronic Journal of Knowledge Management, 12(2):108-116.

Ernst & Young (2001). Executive perspectives on knowledge in the organization. New York: Ernst & Young center for Business Innovation and Business Intelligence.

Fichman RG, Moses SA (1999). An incremental process for software implementation. Sloan Management Review, 40(2):39-52.

Fischer G, Ostwald J (2001). Knowledge Management: Problems, Promises, Realities, and Challenges. IEEE Intelligent Systems, 16(1):60-72.

Gartner IT Glossary (2016). Big Data. Available at: http://www.gartner.com/it-glossary/big-data/ (accessed 25 June 2016).

Gandomi A, Haider M (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2):137-144.

Hinchcliffe D (2007). The state of Enterprise 2.0. Available at: http://www.zdnet.com/article/the-state-of-enterprise-2-0/ (accessed 25 June 2016).

Iglesias CA, Garijo M (2008). The agent-oriented methodology MAS-CommonKADS, in Sugumaran, V. (Ed.), Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications, Information Science, Hershey, PA, pp. 445-468.

Ipe M (2003). Knowledge sharing in organizations: a conceptual framework. Human Resource Development Review, 2(4):337-359.

Java A, Nirenburg S, McShane M, Finin T, English J, Joshi A (2007). Using a Natural Language Understanding System to Generate Semantic Web Content. International Journal on Semantic Web and Information Systems, 3(4):50-74.

Jeong Y, Shin S (2015). AMultidata Connection Scheme for Big Data High-Dimension Using the Data Connection Coefficient. Mathematical Problems in Engineering, Vol. 2015, Article ID 931352, 6 pages.

Kaisler S, Armour F, Espinosa JA, Money W (2013). Big data: Issues and challenges moving forward. In: 2013 IEEE international conference on big data, pp.657–664.

Kakizawa Y (2007). In-house use of Web 2.0: Enterprise 2.0. NEC Technical Journal, 2(2):46-49.

Karlström D, Runeson P (2006). Integrating agile software development into stage-gate managed product development. Empirical Software Engineering, 11(2):203-225.

Kirchner K, Razmerita L, Nabeth, T (2009). Personal and Collective Knowledge Management in the Web 2.0: Two Faces of Knowledge Management. Paper presented at the 9th International Conference on Innovative Internet Community Systems, Jena, Germany, pp.15-26.

Kirchner K, Razmerita L, Sudzina F (2008). New Forms of Interaction and Knowledge Sharing on Web 2.0. In: Miltiadis Lytras ED, De Pablo PO (eds.) Web2.0: The Business Model, USA: Springer Science and Business Media, pp.21–37.

Kohn A, Bry F, Manta A (2010). Semantic Search on Unstructured Data: Explicit Knowledge through Data Recycling. International Journal on Semantic Web and Information Systems, 6(2):17-35.

Kügler M, Dittes S, Smolnik S, Richter A (2015). Connect Me! Antecedents and Impact of Social Connectedness in Enterprise Social Software. Business & Information Systems Engineering, 57(3):181-196.

Lee M, Lan Y (2007). From Web 2.0 to Conversational Knowledge Management: Towards Collaborative Intelligence. Journal of Entrepreneurship Research, 2(2):47-62.

Levy M (2009). WEB 2.0 implications on knowledge management. Journal of Knowledge Management, 13(1):120-134.

Mariscal G, Marban O, Fernandez C (2010). A survey of data mining and knowledge discovery process models and methodologies. Knowledge Engineering review, 25(2):137-166.

McAfee AP (2006). Enterprise 2.0: The Dawn of Emergent Collaboration. MIT Sloan Management Review, 47(3):21–28.

McAfee A, Brynjolfsson E (2012). Big Data: The Management Revolution. Harvard Business Review, 90(10):60-68.

Moteleb AA, Woodman M, Critten P (2009). Towards a practical guide for developing knowledge management systems in small organizations. Proceedings of the 10th European Conference on Knowledge Management, Vicenza, pp.559-570.

Musser J, O'Reilly T, the O'Reilly Radar Team (2006). Web 2.0 Principles and Best Practices. O'Reilly Radar Report, Nov. 2006.

Nakamori Y (2003). Systems Methodology and Mathematical Models for Knowledge Management. Journal of Systems Science and Systems Engineering 12(1).

Nie K, Lin S, Ma T, Nakamori Y (2010). Connecting Informal Networks to Management of Tacit Knowledge. Journal of Systems Science and Systems Engineering, 19(22):237-253.

Nonaka I (1994). A dynamic theory of organizational knowledge creation. Organization Science, 5(1):14-37.

Omerzel DG (2010). The impact of knowledge management on SME growth and profitability: a structural equation modeling study. African Journal of Business Management, 4(16):3417-3432.

O'Reilly T (2005). What is Web 2.0 – Design Patterns and Business Models for the Next Generation of Software. Retrieved August 23, 2010, from http://oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1.

Oren E, Volkel M, Breslin JG, Decker S (2006). Semantic wikis for personal knowledge management. Database and Expert Systems Applications, ser. Lecture Notes in Computer Science, 4080(50):509-518.

Paroutis S, Al Saleh A (2009). Determinants of knowledge sharing using Web 2.0 technologies. Journal of Knowledge Management, 13(4):52 – 63.

Pawlowski JM, Bick M, Martensen M, Peinl R, Thalmann S, Maier R, Kruse P, Hetmank L, Pirkkalainen H, (2014). Social Knowledge Environments. Business & Information Systems Engineering, 6(2):81-88.

Ravi K, Ravi V (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems, 89:14-46.

Razieh D, Raman R (2015). Methodologies for developing knowledge management systems: an evaluation framework. Journal of Knowledge Management, 19(4):682–710.

Razmerita L, Kirchner K, Sudzina F (2009). Personal knowledge management: The role of Web 2.0 tools for managing knowledge at individual and organizational levels. Online Information Review, 33(6):1021–1039.

Robson C (2002). Real World Research. Oxford: Blackwell Publishers.

Rubenstein-Montano B, Liebowitz J, Buchwalter J, McCaw D, Newman B, Rebeck K (2001). A systems thinking framework for knowledge management. Decision Support Systems, 31(1):5-16.

Runeson P, Höst M (2009). Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering, 14(2):131–164.

Šajeva S (2007). An Investigation of Critical Barriers to Effective Knowledge Management. Social Sciences/Socialiniai Mokslai, 58(4):20-27.

Sarnikar S and Deokar A (2010). Knowledge management systems for knowledge-intensive processes: design approach and an illustrative example. Proceedings of 43rd Hawaii International Conference on System Sciences, Honolulu, HI, pp.1-10.

Schneckenberg D (2009). Web 2.0 and the empowerment of the knowledge worker. Journal of Knowledge Management, 13(6):509 – 520.

Serenko A, Bontis N, Booker L, Sadeddin K, Hardie T. (2010). A scientometric analysis of knowledge management and intellectual capital academic literature (1994 – 2008). Journal of Knowledge Management, 14(1):3-23.

Shao BBM, Lin WT (2016). Assessing output performance of information technology service industries: Productivity, innovation and catch-up. International Journal of Production Economics, 172:43-53.

Schermann M, Hemsen H, Buchmüller Ch, Krcmar H, Markl, V, Hoeren, T (2014). Big Data: An Interdisciplinary Opportunity for Information Systems Research. Business & Information Systems Engineering, 6(5):261-266.

Shimazu H, Koike S (2007). KM2.0: Business Knowledge Sharing in the Web 2.0 Age. NEC Technical Journal, 2(2):50–54.

Smuts H, Van der Merwe A, Loock M, Kotzé P (2009). A framework and methodology for knowledge management system implementation. Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists, Vanderbijlpark Emfuleni, pp.70-79.

Song X, Wu Y, Ma Y, Cui Y, Gong G (2015). Military Simulation Big Data: Background, State of the Art, and Challenges. Mathematical Problems in Engineering Vol. 2015, Article ID 298356, 20 pages.

Surowiecki J (2004). The wisdom of the crowd. Ed. Doubleday, Anchor, United States.

Syed A, Gillela K, Venugopal C (2013). The Future Revolution on Big Data. International Journal of Advanced Research in Computer and Communication Engineering, 2(6):2446–2451.

Tan W, Blake MB, Saleh I, Dustdar S (2013). Social-Network-Sourced Big Data Analytics. IEEE Internet Computing, 17(5):62-69.

Trimi S, Galanxhi H (2014). The impact of Enterprise 2.0 in organizations. Service Business, 8(3):405-424.

Wieczorkowski J, Polak P (2014). Big data: Three-aspect approach. Online Journal of Applied Knowledge Management. International Institute for Applied Knowledge Management, 2(2):182-196.

Wu J, Sun H, Tan Y (2013). Social media research: A review. Journal of Systems Science and Systems Engineering, 22(3):257-282.

Zhao D, Zhao W, Sun L, Huang D (2015). Modeling and Querying Business Data with Artifact Lifecycle. Mathematical Problems in Engineering Vol. 2015, Article ID 506272.

4.7. Appendix

Table 4.3. *Appendix*. Survey questions to analyze KMS 2.0 fundamental features and project barriers

	Fundamental Features	Average
System functionality	The system is easy to use.	4.1
	The use of the system is similar to that of other systems you know.	3.8
	The system is flexible.	3.5
	The system is adapted to the needs of the company.	4.5
Quality of the content	The system contains accurate information.	3.9
	The system contains relevant information.	4.7
	The system contains trustworthy information.	4.2
Exchange and accessibility of the content	The system facilitates information sharing.	4.6
	The system facilitates the search for information.	3.7
	The system facilitates the retrieval of information.	3.7
Sociability	The system allows users to comment on the content.	4.8
	The system allows information that has something in common to be related.	3.2
	The system encourages social interactions among users.	4.0
	Barriers	**Average**
Individual barriers	The system enables a quick integration of the new members of the company.	3.7
	The use of the system is beneficial for the members of the company.	4.2
	The system improves collaboration among employees.	4.8
	The system helps members of the company to solve their problems.	4.5
	It is satisfying to help colleagues through the system.	3.2
	The use of the system poses no threat to the jobs of members of the company.	3.4
Organizational context-related barriers	The system strengthens ties between me and existing members of the company.	4.1
	The system expands the scope of my association with other members of the company.	4.6
	The system enables strong relationships to be created with members who have common interests in the company.	4.8
	The system creates new business opportunities for the company.	3.5
	The system improves work process in the company.	3.2
	The system helps the company to achieve its performance objectives.	3.4
	The company rewards users who make better use of the system.	4.1
	The use of the system is beneficial for the company.	4.3
	The system encourages people to suggest ideas for new opportunities.	4.3
	The system provides open communication among colleagues.	4.0
	My superiors make proper use of the system.	3.7
Technological barriers	The system prevents the same questions from being asked several times.	4.2
	The system provides a large amount of information.	4.6
	The system allows you to get information quickly.	4.1
	The system allows you to access the information you need at the time when you need it.	4.5
Project management related barriers	The project staff has the appropriated technical expertise in the system.	3.2
	The resources allocated to the system in the project are adequate.	3.1
	The time allocated to the system in the project is adequate.	2.7
Knowledge nature related barriers	The system facilitates the identification of valuable knowledge.	3.8
	The system facilitates the evaluation of valuable knowledge.	3.8
	The system facilitates the extraction of valuable knowledge.	4.2

Chapter 5. Conclusions

In recent years the way of generating and processing data has changed. Large amounts of data are continually being generated, and the knowledge gained from analyzing such data can be very valuable for organizations. Due to that, it is recommended for organizations to use Big Data systems capable of collecting, processing and analyzing such data in order to take advantage of them, thereby gaining a competitive advantage over their competitors. But in order to be able to make good use of these types of systems, they should be correctly implemented, considering all the necessary steps to achieve it. If it is not done properly, it is probable that problems will arise and will not be overcome, abandoning the implementation before its completion, or if the system is implemented, its reliability is reduced.

In this sense, the framework presented in this thesis is a valuable help for both professionals and organizations, as it offers them the necessary tools (methodologies, techniques, information support tools, etc.) that guide them with a great level of detail in the implementation of Big Data and Web 2.0 systems in different areas of the company. In addition, the proposed framework has been validated through its application in real organizations.

5.1. Contributions and Limitations

This section presents a summary of the main contributions of this thesis, as well as the limitations to be considered.

This thesis presents three research articles, whose general purpose is to develop a framework that guides organizations in the development and implementation of Web 2.0 and Big Data systems, allowing them to generate and share knowledge obtained from their data. Each of these three articles provides specific contributions, related with the proposed objectives, which are indicated below.

The main contributions of Chapter 2, which are linked to the achievement of the Objective 1, are:

- A framework for implementing Big Data ecosystems in organizations. The framework is focused on data, but not only takes into account the operations to be performed on the data, it also considers other important aspects such as human resources and materials, profit estimation, business process re-engineering, monitoring of the system, economic feasibility, definition of indicators, etc.

The main axis of the framework is the methodology dimension. It is a detailed methodology that indicates the activities and tasks to be carried out in each phase that compose it, as well as the dimension of the framework that serves as support to each one of them. The methodology extends throughout the entire life cycle of the project.

- A compilation and study of existing frameworks to manage Big Data ecosystems. It provides a description of each of them, and indicates the phases that compose them. The strengths and weaknesses of all of them are also indicated.

Chapter 3 provides the following main contribution, which allows overcoming the Objective 2:

- A methodology for implementing Social CRM systems. This methodology uses Web 2.0 and Big Data technology for the development and implementation of this type of systems in organizations. It is a methodology that gives a great level of detail in the necessary steps to follow. It is based on the methodology for implementation of CRM systems presented by Chalmeta (2006), which was supplemented, adapted and updated by reviewing the existing literature in the specific fields, and with the experience of the Ph.D. student. The initial version of the methodology was applied to a real company in a case study with the aim of debugging and validating it.

The main contribution provided by Chapter 4, and it manages to fulfill the Objective 3, is:

- A methodology for implementing knowledge management systems based on Big Data and Web 2.0 tools. It is a step-by-step methodology that guides all the necessary steps to develop and implement knowledge management systems 2.0, reducing the complexity of this process. It consists of phases, which contain activities, and these are divided into tasks. The methodology has been improved and validated through its application in a case study in a real company.

Additionally, this thesis contains some limitations to consider. The most relevant are those related to the methods used to perform the analysis and validation of the methodologies developed in the case studies. Such used methods are qualitative, which are not as precise as the quantitative ones. The qualitative methods are based on the perceptions that have had both the people involved in the implementation projects and the consultants that evaluated the proposed framework. Perceptions are subjective, but the experience and professionalism of these people make trust in the honesty and veracity of the provided information.

Another limitation to be highlighted is that each case study of the developed methodologies has been applied in a single organization. One of them was applied in an SME from the metal sector and the other one in a large oil and gas company, so its validity has not been verified in other types of organizations or sectors.

With regard to the validation of the framework, in the expert assessment phase the experts belonged to the same company. To avoid potential biases in their judgments, it could be evaluated by experts from different companies, nationalities, sectors, etc.

5.2. Future Work

This section identifies some opportunities for future lines of research, giving continuity to the work presented in this thesis and proposing improvements related to it.

- Applying the framework and methodologies presented in this thesis in different sectors and types of organizations, with the aim of improving and generalizing them. Furthermore, some quantitative validation method may be introduced in the case studies for more robust validation.

- Developing new versions of the presented methodologies to implement Big Data and Web 2.0 systems, specifically adapted to different areas of organizations, such as supply chain management, product design, marketing, etc.

- Optimizing the data processing in Big Data systems. This could be achieved by using techniques to transform the different formats of each data source of the system in a specific format before carrying out the data collection. With this unification of formats the system would execute a more efficient compilation and data processing, since it would perform the work on a single format.

- Carrying out a more precise analysis of the information published by the same person in different social networks. To achieve this, an automatic identification of the different identifiers used by the same person in various social networks can be performed.

5.3. Publications Associated with this Ph.D. Thesis

The contributions of this thesis have been published in different peer-reviewed international journals. All of them indexed in JCR (Journal Citation Reports – Thomson Reuters). Three research articles have been carried out as a result of the work developed in this Ph.D. thesis.

- Orenga-Roglá S, Chalmeta R (2016). Social customer relationship management: taking advantage of Web 2.0 and Big Data technologies. SpringerPlus (2016) 5:1462. DOI 10.1186/s40064-016-3128-y.

- Orenga-Roglá S, Chalmeta R (2017). Methodology for the implementation of knowledge management systems 2.0: A case study in an oil and gas company. Accepted for publication in: Business & Information Systems Engineering (BISE).

- Orenga-Roglá S, Chalmeta R (2017). Framework for the implementation of a Big Data ecosystem in organizations. Under review in: Journal of Systems and Software.

5.4. References

Chalmeta R (2006). Methodology for customer relationship management. The Journal of Systems and Software, 79(7):1015-1024.

Appendix A. Framework para la Implementación de un Ecosistema Big Data en Organizaciones

En este capítulo se presenta el tema de investigación sobre el que trata esta tesis. Inicialmente se realiza una breve descripción del planteamiento del problema de la investigación llevada a cabo, así como de los principales objetivos de la misma. A continuación se indica la metodología de investigación empleada. Finalmente, se explica cómo está organizada la tesis y se proporciona una breve descripción del contenido de los capítulos que la componen.

A.1. Planteamiento del Problema

Hoy en día vivimos en la era de la información, en la que en cada segundo se generan grandes cantidades de datos. Esto es debido, entre otras cosas, a la enorme cantidad existente de diferentes procesos y dispositivos que generan datos, como sensores, archivos log, dispositivos móviles, registros de transacciones, etc., y a la alta velocidad en la creación de los mismos. Debido a la cantidad y características de dichos datos (son desestructurados, complejos, multi-formato, multi-canal, etc.) los sistemas informáticos tradicionales tienen dificultades para gestionarlos (Elgendy and Elragal, 2014; Syed et al., 2013). Esto, unido al reciente descenso del coste computacional y de almacenamiento (Tekiner and Keane, 2013), ha propiciado la aparición de la tecnología Big Data. Esta tecnología es capaz de analizar dichas cantidades y tipos de datos generando conocimiento útil (Almeida and Bernardino, 2015).

La tecnología Big Data permite a las organizaciones extraer valor y conocimiento a partir de los datos generados tanto dentro como fuera de la organización (Assunção et al., 2015). De este modo, el Big Data es una herramienta de competitividad que permite a las organizaciones descubrir conocimiento empresarial con el objetivo de incrementar el rendimiento del negocio, y obtener así una ventaja competitiva con respecto a sus competidores. Además, el Big Data puede complementarse con la tecnología Web 2.0, que además de generar una gran cantidad de datos valiosos, permiten a las empresas tener una comunicación más efectiva, fomentar la colaboración, y facilitar la interacción social y la compartición de conocimiento (Kirchner et al., 2009; Wu et al., 2013).

Desarrollar e implementar un ecosistema Big Data en una organización es una tarea muy compleja que abarca no sólo los aspectos tecnológicos, sino que también la gestión de las políticas y las personas (Tekiner and Keane, 2013). Además, la implementación de sistemas Big Data y Web 2.0 en las organizaciones implica la coordinación y colaboración

de muchos usuarios, así como la ejecución y sincronización de una gran cantidad de actividades y tareas. Para realizar dicha implementación con garantías de éxito, las organizaciones pueden apoyarse en frameworks.

Un framework describe conceptos, aspectos, características, procesos, flujos de datos, o relaciones entre componentes, para determinados campos (por ejemplo el desarrollo de software), con el objetivo de generar un mejor entendimiento (como la descripción de componentes o aspectos de diseño) o de guiar en la consecución de objetivos específicos (Pawlowski and Bick, 2012). Los frameworks se componen de dimensiones, que normalmente están interrelacionadas entre sí. En la literatura se han propuesto algunos frameworks para gestionar ecosistemas Big Data (Das and Kumar, 2013; Demchenko et al., 2014; Ferguson, 2012; Géczy, 2015; Miller and Mork, 2013; Sun and Heller, 2012; Tekiner and Keane, 2013).

Sin embargo, en el estado actual de los frameworks existentes, estos poseen algunos problemas que limitan su efectividad para desarrollar e implementar ecosistemas Big Data. Dichos frameworks están principalmente basados en los datos, es decir, en las operaciones con ellos, en cómo se generan, en sus características, en el uso que se hace de ellos, y en el propósito de las operaciones realizadas sobre ellos. Sin embargo, no consideran otros aspectos que son también muy importantes como: (1) ninguno de ellos contempla todos los aspectos de los ecosistemas Big Data; (2) no poseen una metodología que guíe los pasos a seguir en el proceso de desarrollo e implementación de ecosistemas Big Data, lo que dificulta este proceso; (3) no proporcionan casos de estudio robustos en los que son evaluados, por lo tanto, su validez no ha sido comprobada; o (4) no consideran el impacto humano, organizacional y de procesos de negocios de la implementación de sistemas Big Data.

En este contexto, destaca como un tema importante de investigación el desarrollo de un framework que sirva de ayuda en el proceso complejo de incorporación de ecosistemas Big Data en las organizaciones. El trabajo realizado en esta tesis trata de facilitar a las organizaciones el desarrollo e implementación de ecosistemas Big Data, guiándoles detalladamente en todos los pasos necesarios a realizar y considerando todos los aspectos a tener en cuenta en cada fase para obtener un resultado satisfactorio.

A.2. Objetivos

En este marco, se plantea el objetivo fundamental de la presente tesis (Objetivo 1) que se sintetiza en desarrollar un framework genérico que permita a las organizaciones utilizar las tecnologías Big Data y Web 2.0, minimizando el tiempo y esfuerzo invertidos, y considerando todos los aspectos relacionados con la implementación, tanto los técnicos, como los políticos, culturales, de comportamiento, etc.

Para mejorarlo y validarlo, el framework genérico se aplica a dos casos de estudio, los cuales tienen lugar en dos ámbitos diferentes de la organización. Por una parte se aplica a la gestión de las relaciones con los clientes en una PyME (Pequeña y Mediana Empresa) del sector del metal (Objetivo 2), y por otra parte a la gestión del conocimiento en una gran empresa del sector del petróleo y gas (Objetivo 3).

A.3. Metodología de Investigación

En esta tesis doctoral se han utilizado diferentes métodos de investigación: exploratoria, que estructura e identifica nuevos problemas (Stebbins, 2001); constructiva, que desarrolla soluciones para los problemas (Lukka, 2003); y empírica, que pone a prueba la viabilidad de las soluciones utilizando pruebas empíricas (Wohlin et al., 2006).

En primer lugar, para el desarrollo de la versión inicial del framework, se realizaron investigaciones exploratorias sobre la literatura existente acerca de frameworks de desarrollo e implantación de sistemas Web 2.0 y Big Data. Como fruto de esta investigación exploratoria se encontraron y analizaron las principales soluciones existentes, en las que se detectaron diferentes debilidades y deficiencias, así como posibles mejoras.

Con la base obtenida a partir de la revisión de la literatura y con la experiencia previa del doctorando, se construyó la primera versión del framework. Dicha versión se aprovechaba de las fortalezas de las soluciones existentes, mejorándolas en la medida de lo posible y solucionando o minimizando las deficiencias y debilidades encontradas.

Esta primera versión fue aplicada en entornos reales utilizando para ello la metodología del caso de estudio. El plan de trabajo seguido para llevar a cabo los casos de estudio fue el basado en Runeson y Höst (2009), el cual está formado por cinco fases: Diseño y estudio del caso de estudio; Preparación para la recopilación de datos; Recopilación de pruebas; Análisis de los datos recopilados; y Validación de los datos recopilados.

Los resultados obtenidos se utilizaron para depurar, mejorar y validar el framework, así como para crear dos especializaciones del mismo, una para la gestión de las relaciones con los clientes y otra para la gestión del conocimiento empresarial.

A.4. Estructura

Esta tesis está organizada en cinco capítulos, los cuales se resumen a continuación.

En el Capítulo 1 se hace una introducción al tema sobre el que trata esta tesis. Los Capítulos 2 al 4 son tres artículos de investigación completos e independientes, pero que tratan sobre temas relacionados entre sí, cubriendo los tres objetivos propuestos en la tesis. Así, el Capítulo 2 corresponde al Objetivo 1, el Capítulo 3 al Objetivo 2, y el Capítulo 4 al Objetivo 3. Cada uno de estos tres capítulos tiene su propia estructura completa, y puede ser leído y comprendido de forma independiente. Finalmente, en el Capítulo 5 se presentan las conclusiones extraídas del trabajo de esta tesis.

Capítulo 1: Introducción. En este capítulo se muestra una breve descripción general del tema de investigación analizado y presentado en esta tesis. Se indica el planteamiento del problema de investigación, la metodología de trabajo empleada, los objetivos propuestos y los resultados obtenidos. También se explica cómo está organizado el documento.

Capítulo 2: Framework para la implementación de un ecosistema Big Data en organizaciones. En este capítulo se presenta un framework general para implementar y gestionar ecosistemas Big Data en las organizaciones. Primero se realiza un estudio de los principales frameworks de este tipo existentes en la literatura, analizando sus fortalezas y debilidades. Después se presenta el framework propuesto, el cual se aprovecha de las fortalezas y evita las debilidades identificadas en los frameworks analizados.

El framework propuesto está compuesto por siete dimensiones interrelacionadas, que son: Metodología; Arquitectura de datos; Organizativa; Fuentes de datos; Calidad de los datos; Herramientas de soporte; y Privacidad/Seguridad. La dimensión de la Metodología es la principal, se extiende sobre toda la vida del proyecto, y es la que sirve de guía en los pasos necesarios para implementar un sistema con la tecnología Big Data. Las otras dimensiones dan soporte a las diferentes fases, actividades y tareas que componen la metodología.

Capítulo 3: Gestión social de las relaciones con los clientes: aprovechando las tecnologías Web 2.0 y Big Data. En este capítulo se presenta una especialización de la dimensión de la Metodología propuesta en el framework general, la cual guía el desarrollo e implementación de la gestión social de las relaciones con los clientes (Social Customer Relationship Management – Social CRM) en las organizaciones, teniendo en cuenta tanto aspectos organizacionales, como humanos y tecnológicos. Está desarrollada a partir de la metodología de implementación de sistemas CRM (Customer Relationship Management – Gestión de la Relación con los Clientes) propuesta en Chalmeta (2006), y complementada, adaptada y actualizada basándose en la revisión de la literatura existente en el tema y en la experiencia del doctorando. La metodología está compuesta por fases, las cuales están formadas por actividades. Las fases que la componen son las siguientes: Gestión de proyectos y prerrequisitos; Framework organizacional; Estrategia de clientes; Sistema de evaluación de relaciones con los clientes; Mapa de procesos; Recursos humanos; Sistema informático; Implantación; y Monitorización.

En el caso de estudio que se presenta en este capítulo, la versión inicial de la metodología se aplicó exitosamente a una empresa PyME del sector del metal con el objetivo de analizarla, validarla y refinarla. Los resultados y feedback obtenidos en dicho caso de estudio, se aprovecharon para mejorar tanto la metodología, como el framework genérico.

Capítulo4: Metodología para la implementación de sistemas de gestión del conocimiento 2.0: Caso de estudio en una empresa del sector del petróleo y gas. Este capítulo presenta una especialización de la dimensión de la Metodología propuesta en el framework general para guiar la implementación y desarrollo de sistemas de gestión del conocimiento 2.0, utilizando herramientas Big Data y Web 2.0. La metodología se compone de siete fases: Anteproyecto; Planificación; Análisis; Diseño; Desarrollo; Implementación; y Control. Cada una de estas fases está compuesta de actividades, que a su vez contienen tareas. La metodología guía los procesos de recopilación, generación, gestión y aplicación del conocimiento generado tanto interna como externamente a la organización. Con la aplicación de la metodología se reduce la complejidad del proyecto y aumentan sus probabilidades de éxito.

La metodología propuesta se aplicó en un caso de estudio a una gran empresa del sector del petróleo y gas, en la cual fue analizada, depurada y validada. Se detallan todos los pasos seguidos, junto con los usuarios implicados y los resultados obtenidos. Los comentarios, propuestas y feedback recopilados en el caso de estudio se utilizaron para depurar y mejorar la metodología, así como también el framework genérico.

Capítulo 5: Conclusiones. En este capítulo se presentan las conclusiones generales de la tesis, así como las principales contribuciones aportadas. También se indican las limitaciones de este trabajo y las posibles futuras líneas de investigación a seguir. Por último se incluyen las publicaciones asociadas a esta tesis.

A.5. Referencias

Almeida P, Bernardino J (2015). A comprehensive overview of open source big data platforms and frameworks. International Journal of Big Data, 2(3):15-33.

Assunção MD, Calheiros RN, Bianchi S, Netto M, Buyya R (2015). Big data computing and clouds: trends and future directions. Journal of Parallel and Distributed Computing, 79-80(2015):3-15.

Chalmeta R (2006). Methodology for customer relationship management. The Journal of Systems and Software, 79(7):1015-1024.

Das TK, Kumar PM (2013). BIG Data Analytics: A Framework for Unstructured Data Analysis. International Journal of Engineering and Technology, 5(1):153–156.

Demchenko Y, de Laat C, Membrey P (2014). Defining Architecture Components of the Big Data Ecosystem. International Conference on Collaboration Technologies and Systems (CTS), pp.104-112.

Elgendy N, Elragal A (2014). Big Data Analytics: A Literature Review Paper. In Advances in Data Mining. Applications and Theoretical Aspects, 8557(2014):214-227.

Ferguson M (2012). Architecting a Big Data Platform for Analytics. IBM White Paper, October 2012. Available at: http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IML14333USEN.

Gèczy P (2015). Big Data Management: Relational Framework. Review of Business & Finance Studies, 6(3):21-30.

Kirchner K, Razmerita L, Nabeth, T (2009). Personal and Collective Knowledge Management in the Web 2.0: Two Faces of Knowledge Management. Paper presented at the 9th International Conference on Innovative Internet Community Systems, Jena, Germany, 15-26.

Lukka K (2003). The constructive research approach. In Case study research in logistics. Publications of the Turku School of Economics and Business Administration, Series B, 1(2003):83-101.

Miller HG, Mork P (2013). From Data to Decisions: A Value Chain for Big Data. IT Professional, 15(1):57–59.

Pawlowski J, Bick M (2012). The Global Knowledge Management Framework: Towards a Theory for Knowledge Management in Globally Distributed Settings. Electronic Journal of Knowledge Management, 10(1):92-108.

Runeson P, Höst M (2009). Guidelines for conducting and reporting case study research in software engineering. Empirical Software Engineering, 14(2):131–164.

Stebbins RA (2001). Exploratory Research in the Social Sciences. Sage Publications, Inc., Thousand Oaks, CA.

Sun H, Heller P (2012). Oracle Information Architecture: An Architect's Guide to Big Data. Oracle White Paper, August 2012. Available at: https://d2jt48ltdp5cjc.cloudfront.net/uploads/test1_3021.pdf.

Syed A, Gillela K, Venugopal C (2013). The Future Revolution on Big Data. International Journal of Advanced Research in Computer and Communication Engineering, 2(6):2446-2451.

Tekiner F, Keane JA (2013). Big Data Framework. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE Computer Society, 1494-1499.

Wohlin C, Höst M, Henningsson K (2006). Empirical Research Methods in Web and Software Engineering. In Web Engineering, p. 409-430. Springer, Berlin, Heidelberg.

Wu J, Sun H, Tan Y (2013). Social media research: A review. Journal of Systems Science and Systems Engineering, 22(3):257-282.

Appendix B. Conclusiones

En los últimos años ha cambiado la forma de generar y procesar los datos. Continuamente se están generando una gran cantidad de datos, y el conocimiento obtenido a partir del análisis de esos datos puede ser muy valioso para las organizaciones. Debido a eso, es recomendable que las organizaciones utilicen sistemas Big Data que sean capaces de recopilar, procesar y analizar esos datos para poder sacar provecho de ellos, y así obtener una ventaja competitiva con respecto a sus competidores. Pero para poder llegar a hacer un buen uso de este tipo de sistemas, se debe realizar una implantación correcta de los mismos, considerando todos los pasos necesarios. De no ser así, es probable que surjan problemas que no se lleguen a superar, abandonando la implantación antes de su finalización, o si se llega a implantar el sistema se reduce la fiabilidad del mismo.

En este sentido, el framework que se presenta en esta tesis es de gran ayuda tanto para profesionales como para organizaciones, ya que les ofrece las herramientas necesarias (metodologías, técnicas, herramientas informáticas de apoyo, etc.) que les guían con un gran nivel de detalle en la implementación de sistemas Web 2.0 y Big Data en diferentes ámbitos de la empresa. Además, el framework propuesto ha sido validado mediante su aplicación en organizaciones reales.

B.1. Contribuciones y Limitaciones

En este apartado se presenta un resumen de las principales contribuciones que aporta esta tesis, así como las limitaciones a considerar.

En esta tesis se presentan tres artículos de investigación, cuya finalidad general es la de desarrollar un framework que guíe a las organizaciones en el desarrollo e implementación de sistemas Web 2.0 y Big Data, permitiéndoles generar y compartir conocimiento a partir de los datos que poseen. Cada uno de estos tres artículos aporta unas contribuciones específicas, relacionadas con los objetivos propuestos, que son indicadas a continuación.

Las principales contribuciones aportadas en el Capítulo 2, las cuales se corresponden con la consecución del Objetivo 1, son:

- Un framework para implementar ecosistemas Big Data en organizaciones. El framework está centrado en los datos, pero no sólo tiene en cuenta las operaciones a realizar sobre los datos, sino que también considera otros aspectos importantes como los recursos humanos y materiales, estimación de beneficios, reingeniería de procesos de negocios, monitorización del sistema, viabilidad económica, definición de indicadores, etc.

El eje principal del framework es la dimensión de la metodología. Se trata de una metodología detallada que indica las actividades y tareas a realizar en cada una de las fases que la componen, así como la dimensión del framework que sirve de soporte a cada una de ellas. Esta metodología se extiende durante todo el ciclo de vida del proyecto.

- Una recopilación y estudio de los frameworks existentes para gestionar ecosistemas Big Data. Se aporta una descripción de cada uno de ellos y las fases que los componen. También se indican las fortalezas y debilidades de todos ellos.

En el Capítulo 3 se aporta la siguiente contribución principal, la cual permite superar el Objetivo 2:

- Una metodología de implementación de sistemas Social CRM. Esta metodología utiliza la tecnología Web 2.0 y Big Data para el desarrollo e implementación de este tipo de sistemas en las organizaciones. Es una metodología que aporta un gran nivel de detalle en los pasos necesarios a seguir. Está basada en la metodología de implementación de sistemas CRM presentada en Chalmeta (2006), la cual fue complementada, adaptada y actualizada mediante la revisión de la literatura existente en los campos específicos, además de con la experiencia del doctorando. La versión inicial de la metodología se aplicó a una empresa real en un caso de estudio con el objetivo de depurarla y validarla.

La contribución principal que aporta el Capítulo 4, y consigue alcanzar el Objetivo 3, es:

- Una metodología de implementación de sistemas de gestión del conocimiento basada en herramientas Web 2.0 y Big Data. Se trata de una metodología detallada que guía en todos los pasos necesarios para desarrollar e implementar un sistema de gestión del conocimiento 2.0, reduciendo la complejidad de este proceso. Se compone de fases, que contienen actividades, y éstas se dividen en tareas. Esta metodología ha sido mejorada y validada mediante su aplicación en un caso de estudio en una empresa real.

Por otra parte, esta tesis contiene algunas limitaciones a tener en cuenta. Las más destacables son las relacionadas con los métodos utilizados para realizar el análisis y validación de las metodologías desarrolladas en los casos de estudio. Dichos métodos utilizados fueron cualitativos, por lo que no son tan precisos como los cuantitativos. Los métodos cualitativos se basan en las percepciones que han tenido tanto las personas implicadas en los proyectos de implantación como los consultores que evaluaron el framework propuesto. Las percepciones son subjetivas, pero la experiencia y profesionalidad de estas personas hacen confiar en la honestidad y veracidad de la información aportada.

Otra limitación a desatacar es que los casos de estudio de las metodologías desarrolladas han sido aplicados en una sola organización cada uno. Uno se aplicó en una PyME del sector del metal y el otro en una gran empresa del sector del petróleo y gas, por lo que su validez no ha sido verificada en otros tipos de organizaciones o sectores.

En lo que respecta a la validación del framework, en la fase de la evaluación de expertos, los expertos pertenecían a una misma empresa. Para evitar potenciales influencias en sus

juicios, podría ser evaluado por expertos pertenecientes a diferentes empresas, nacionalidades, sectores, etc.

B.2. Trabajo Futuro

En este apartado se identifican algunas oportunidades para futuras líneas de investigación dando continuidad al trabajo presentado en esta tesis y proponiendo mejoras relacionadas con el mismo.

- Aplicar el framework y las metodologías presentadas en esta tesis en organizaciones de diferentes tipos y sectores, con el objetivo de mejorarlas y generalizarlas. Además, se puede introducir algún método de validación cuantitativo en los casos de estudio para realizar una validación más robusta.

- Desarrollar nuevas versiones de las metodologías presentadas para implementar sistemas Big Data y Web 2.0 adaptadas específicamente a diferentes áreas de las organizaciones, como pueden ser gestión de la cadena de suministros, diseño de productos, marketing, etc.

- Optimizar el procesamiento de datos en los sistemas Big Data. Esto se podría conseguir mediante la utilización de técnicas para transformar los diferentes formatos de cada una de las fuentes de datos del sistema en un formato específico antes de realizar la recopilación de datos. Con esta unificación de formatos el sistema realizaría una recopilación y procesamiento de datos más eficiente, ya que realizaría el trabajo sobre un único formato.

- Realizar un análisis más preciso de la información publicada por una misma persona en diferentes redes sociales. Para conseguir esto se puede efectuar una identificación automática de los diferentes identificadores utilizados por una misma persona en varias redes sociales.

B.3. Publicaciones Asociadas a esta Tesis

Las contribuciones de esta tesis han sido publicadas en diferentes revistas internacionales revisadas por expertos, todas ellas indexadas en el JCR (Journal Citation Reports – Thomson Reuters). Como resultado del trabajo desarrollado en esta tesis se han elaborado tres artículos de investigación.

- Orenga-Roglá S, Chalmeta R (2016). Social customer relationship management: taking advantage of Web 2.0 and Big Data technologies. SpringerPlus (2016) 5:1462. DOI 10.1186/s40064-016-3128-y.

- Orenga-Roglá S, Chalmeta R (2017). Methodology for the implementation of knowledge management systems 2.0: A case study in an oil and gas company. Aceptado para publicación en: Business & Information Systems Engineering (BISE).

- Orenga-Roglá S, Chalmeta R (2017). Framework for the implementation of a Big Data ecosystem in organizations. En revisión en: Journal of Systems and Software.

B.4. Referencias

Chalmeta R (2006). Methodology for customer relationship management. The Journal of Systems and Software, 79(7):1015-1024.