Supporting Competitive Intelligence With Linked Enterprise Data

Vitor Afonso Pinto

Belo Horizonte
2014
Supporting Competitive Intelligence With Linked Enterprise Data


Orientador: Prof. Dr. Fernando Silva Parreiras

Linha de Pesquisa: Tecnologia e Sistemas de Informação

Belo Horizonte
2014
To my beloved wife Carla... "You were my strength when I was weak. You were my voice when I couldn't speak. You were my eyes when I couldn't see. You saw the best there was in me. Lifted me up when I couldn’t reach. You gave me faith, cause you believed. I’m everything I am, because you loved me".
Acknowledgements

I thank God for the gift of the life. God only knows the difficulties I faced for this moment to become true. For two years, I frequently heard God’s voice saying: ‘do not fear, for I am with you; do not be dismayed, for I am your God. I will strengthen you and help you; I will uphold you with my righteous right hand’. (Isaiah 41:10).

I thank to all my family. To my wife Carla and to my son João Vitor - who is not even borned - for supporting me during research period and understanding my absences due to studies. To my parents Fernando and Edilene for all the principles and values they taught me. To my sister Fernanda, her husband Gesiel and their kids Rafael and Ana Beatriz for helping me understand how precious the family is.

I thank to all my friends, from near and from far. My friends helped me to live unforgettable moments. They are part of my history and provided so pleasurable conversations which always eliminated any haste. I also thank to VALE’s colleagues who were able to work with me even when my mood was severely affected by sleepless nights.

I thank to Prof. Dr. Fernando Parreiras for his advice, commitment, wisdom and also for his valuable contributions and recommendations which intended to properly conduct this work. I extend my gratitude to teachers, secretaries, LAIS colleagues and other professionals who brighten the Programa de Pós-Graduação em Sistemas de Informação e Gestão do Conhecimento of FUMEC University.

Lastly, I thank all of which directly or indirectly contributed to the completion of this work.
In God We Trust. All Others Must Bring Data.

(William E. Deming)
Resumo

Inteligência Competitiva é um processo que envolve a recuperação, análise e agrupamento de informações visando a oferta de um produto que responda às necessidades de inteligência de um decisor ou comunidade de decisores. Fontes de informação da Internet estão se tornando cada vez mais importantes neste processo, pois grande parte dos conteúdos disponíveis são gratuitos. Neste trabalho a seguinte questão de pesquisa foi abordada: Quais os conceitos e tecnologias relacionados a linked data que permitem a obtenção, integração e compartilhamento de informações para suportar a inteligência competitiva? Para responder a esta questão, inicialmente a literatura foi revisada possibilitando o desenho de arcabouço conceitual. Em seguida, algumas questões de competência foram definidas por meio de um grupo de foco realizado em um objeto de estudo. Por último, a ferramenta DB4Trading foi construída como um protótipo capaz de validar o arcabouço conceitual proposto. Os resultados da pesquisa apontam que a adoção de tecnologias de Web Semântica possibilitam a obtenção dos dados necessários para a análise de ambientes externos. Além disso, os resultados da pesquisa indicam que as empresas utilizam tecnologias da Web Semântica para apoiar suas operações apesar de considerarem essas tecnologias como complexas. Este trabalho contribui para o processo de tomada de decisão, especialmente no contexto da inteligência competitiva. Este trabalho também contribui para a redução dos custos para obtenção de informações além das fronteiras da organização por meio de tecnologias da Web Semântica.

Palavras-chaves: Web Semântica; Linked Data; Linked Enterprise Data; Inteligência Competitiva.
Abstract

Competitive Intelligence is a process which involves retrieving, analyzing and packaging information to offer a final product that responds to the intelligence needs of a particular decision maker or community of decision makers. Internet-based information sources are becoming increasingly important in this process because most of the contents available on the Web are available free of charge. In this work the following research question was addressed: What are the concepts and technologies related to linked data which allow gathering, integration and sharing of information to support competitive intelligence? To answer this question, firstly, the literature was reviewed in order to outline the conceptual framework. Next, some competency questions were defined through a focus group in a study object. Finally, DB4Trading tool was built as a prototype able to validate the conceptual framework. Results point out that adoption of Semantic Web technologies enable to obtain the data needed for the analysis of external environments. Besides that, results indicate that companies use Semantic Web technologies to support its operations despite consider these technologies as complex. This work adds to the decision-making process, specially in the context of competitive intelligence. This work also contributes to reducing costs to obtain information beyond organization boundaries by using Semantic Web technologies.

Keywords: Semantic Web; Linked Data; Linked Enterprise Data; Competitive Intelligence.
List of Illustrations

Figura 1 – Semantic Web Stack ................................................. 24
Figura 2 – Linked Open Data Cloud ........................................... 28
Figura 3 – SLR - Reasons for Exclusion of Studies ....................... 32
Figura 4 – SLR - Transformation of Raw Results ......................... 32
Figura 5 – SLR - Applications of Semantic Web for Corporate Environments ........................................... 35
Figura 6 – SLR - Kinds of Data Sources Interlinked ..................... 35
Figura 7 – SLR - Concerns Regarding Semantic Web Adoption ........ 36
Figura 8 – SLR - Similar Functional Layers of Frameworks .......... 38
Figura 9 – Objectives and Methodologies .................................. 41
Figura 10 – Focus Group - Participants Roles ............................... 48
Figura 11 – Focus Group - Age and Experience of Participants ......... 49
Figura 12 – Focus Group - Participants Specialties ........................ 50
Figura 13 – Focus Group - Matrix of Datasources ....................... 53
Figura 14 – Focus Group - Relevant Information .......................... 55
Figura 15 – Conceptual Framework .......................................... 56
Figura 16 – Implementation - Wrapper Layer Architecture ............. 65
Figura 17 – Implementation - Wrapper Layer Platform .................. 66
Figura 18 – Implementation - Integration Layer RDF/XML Example .... 66
Figura 19 – Implementation - Presentation Layer Main Page ........... 67
Figura 20 – Implementation - Presentation Layer Structure .............. 68
Figura 21 – Implementation - Presentation Layer Heat Map Formula .... 69
Figura 22 – Framework Validation - Semantic Repository Queries Results .... 72
Figura 23 – Framework Validation - DB4Trading Scenario 1 Parameters .... 73
Figura 24 – Framework Validation - DB4Trading Scenario 2 Parameters .... 74
Figura 25 – Framework Validation - DB4Trading Scenario 1 Results .... 75
Figura 26 – Framework Validation - DB4Trading Scenario 2 Results .... 75
Figura 27 – Appendix - Focus Group Invitation (English Version) .... 95
Figura 28 – Appendix - Focus Group Invitation (Portuguese Version) ... 95
Figura 29 – Appendix - Focus Group Consent Form (English Version) ... 97
Figura 30 – Appendix - Focus Group Consent Form (Portuguese Version) ... 98
Figura 31 – Appendix - Focus Group Commitments (English Version) .... 100
Figura 32 – Appendix - Focus Group Commitments (Portuguese Version) ... 101
List of Tables

Tabela 1 – SLR - Theoretical Foundations ........................................... 30
Tabela 2 – SLR - Conceptual Search Strings ......................................... 31
Tabela 3 – SLR - Searched Fields at Databases ..................................... 31
Tabela 4 – SLR - Exclusion Criteria ..................................................... 31
Tabela 5 – SLR - Studies by Year ......................................................... 34
Tabela 6 – SLR - Category of Concerns Regarding Semantic Web Adoption . 37
Tabela 7 – SLR - Frameworks Functional Layers .................................... 37
Tabela 8 – Related Work - Similarities and Differences ............................ 39
Tabela 9 – Focus Group - Generic Datasources ..................................... 54
Tabela 10 – Functions of Conceptual Framework .................................... 57
Tabela 11 – Implementation - Data Layer Sources .................................. 57
Tabela 12 – Implementation - Data Layer Categories ............................... 59
Tabela 13 – Implementation - Data Layer Variables ................................ 59
Tabela 14 – Implementation - Presentation Layer Website Parts ................ 67
Tabela 15 – Implementation - Presentation Layer Files and Purpose ............ 68
Tabela 16 – Implementation - Presentation Layer Heat Map Formula Parameters 69
Tabela 17 – Framework Validation - Semantic Repository Queries ............... 70
Tabela 18 – Appendix - Focus Group Questions ...................................... 94
## List of abbreviations and acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>ASP</td>
<td>Active Server Pages</td>
</tr>
<tr>
<td>B2B</td>
<td>Business To Business</td>
</tr>
<tr>
<td>CI</td>
<td>Competitive Intelligence</td>
</tr>
<tr>
<td>EAI</td>
<td>Enterprise Application Integration</td>
</tr>
<tr>
<td>ERP</td>
<td>Enterprise Resource Planning</td>
</tr>
<tr>
<td>FAO</td>
<td>Food and Agriculture Organization</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>HDI</td>
<td>Human Development Index</td>
</tr>
<tr>
<td>HR</td>
<td>Human Resource</td>
</tr>
<tr>
<td>HTML</td>
<td>HyperText Markup Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>ISO</td>
<td>International Standards Organization</td>
</tr>
<tr>
<td>IT</td>
<td>Information Technology</td>
</tr>
<tr>
<td>JSON</td>
<td>Javascript Object Notation</td>
</tr>
<tr>
<td>LD</td>
<td>Linked Data</td>
</tr>
<tr>
<td>LED</td>
<td>Linked Enterprise Data</td>
</tr>
<tr>
<td>LOD</td>
<td>Linked Open Data</td>
</tr>
<tr>
<td>MS-SQL</td>
<td>Microsoft SQL Server Database</td>
</tr>
<tr>
<td>OWL</td>
<td>Ontology Web Language</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>SLR</td>
<td>Systematic Literature Review</td>
</tr>
<tr>
<td>SPARQL</td>
<td>Protocol and RDF Query Language</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>SSL</td>
<td>Secure Socket Layer</td>
</tr>
<tr>
<td>SW</td>
<td>Semantic Web</td>
</tr>
<tr>
<td>TCP/IP</td>
<td>Transmission Control Protocol / Internet Protocol</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>UNESCO</td>
<td>United Nations Educational, Scientific and Cultural Organization</td>
</tr>
<tr>
<td>URI</td>
<td>Uniform Resource Identifier</td>
</tr>
<tr>
<td>URL</td>
<td>Uniform Resource Locator</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>WWW</td>
<td>World Wide Web</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
</tbody>
</table>
Summary

1 INTRODUCTION .................................................. 14
1.1 Research Problem ........................................... 15
1.2 Motivation .................................................. 15
1.3 Objectives .................................................. 15
1.3.1 Main Objective ........................................... 15
1.3.2 Specific Objectives ...................................... 15
1.4 Adherence to FUMEC’s Research Program .............. 16
1.5 Thesis Structure ............................................ 17
1.6 Communications of this Thesis ............................ 17

2 SYSTEMATIC LITERATURE REVIEW ...................... 18
2.1 Introduction ............................................... 18
2.2 Theoretical Background ................................... 19
2.2.1 Competitive Intelligence ............................... 19
2.2.2 Semantic Web ........................................... 21
2.2.2.1 Semantic Web Technologies ....................... 23
2.2.2.2 Linked Data ........................................ 26
2.2.2.3 Linked Open Data .................................. 27
2.2.2.4 Linked Enterprise Data ............................ 28
2.3 Design ...................................................... 30
2.3.1 Planning ................................................. 30
2.3.2 Execution ............................................... 32
2.4 Results ...................................................... 33
2.4.1 Studies Overview ....................................... 33
2.4.2 Research Questions Evaluation ....................... 34
2.4.2.1 RQ1 What are the applications of Semantic Web for Corporate Environments? .... 34
2.4.2.2 RQ1.1 Which kind of data source enterprises who adopted Semantic Web often interlink? 34
2.4.2.3 RQ1.2 What are the enterprises’s concerns regarding Semantic Web adoption? .......... 36
2.4.2.4 RQ1.3 What are the similarities between the frameworks used to implement Semantic Web on Enterprises? .................................................. 36
2.5 Threats to Validity ......................................... 36
2.6 Conclusion .................................................. 37

3 RELATED WORK ............................................... 39

4 METHODOLOGY ............................................... 40
4.1 Research Phases .................................................. 41
4.1.1 Phase 1 - Systematic Literature Review .................... 41
4.1.2 Phase 2 - Focus Group ....................................... 42
4.1.2.1 Focus Group Definition .................................. 42
4.1.2.2 Focus Group Phases ..................................... 43
4.1.2.3 Focus Group Limitations ................................. 43
4.1.3 Phase 3 - Experimentation .................................. 43
4.2 Research Object ................................................. 44
4.2.1 Portfolio Management Enabling Strategy .................... 44

5 COMPETENCY QUESTIONS ............................................. 46
5.1 Focus Group Design .............................................. 46
5.1.1 Participants ................................................... 47
5.1.2 Written Invitation ............................................ 47
5.2 Focus Group Execution .......................................... 47
5.2.1 Data Collection ............................................... 49
5.2.2 Ethics ......................................................... 50
5.2.3 Interview ..................................................... 50
5.2.4 Limitations ................................................... 51
5.3 Focus Group Results ............................................. 51
5.3.1 Datasources .................................................. 51
5.3.2 Information .................................................. 51
5.3.3 Difficulties ................................................... 52

6 A FRAMEWORK FOR LINKED ENTERPRISE DATA ............... 56
6.1 Conceptual Framework ........................................... 56
6.2 Implementation .................................................. 57
6.2.1 Data Layer ................................................... 57
6.2.2 Wrapper Layer ............................................... 64
6.2.3 Integration Layer ............................................ 64
6.2.4 Presentation Layer ......................................... 67
6.2.4.1 Site Structure ........................................... 68
6.2.4.2 Heat Map ................................................ 68

7 FRAMEWORK VALIDATION .......................................... 70
7.1 Semantic Repository ............................................. 70
7.2 Visualization Application ....................................... 72

8 CONCLUSION .......................................................... 76
1 Introduction

The increasing global competition and the rapid advances in information systems have led organizations to search for more effective and efficient ways to manage their business (ZANDI; TAVANA, 2012). Executives must be timely and comprehensively informed so that they can make decisions to counteract the threats to business revenue. Everyone needs to have ready and immediate access to information/data that enable her to carry out the allocated tasks (HU; SVENSSON, 2010). Organizations have to be more agile than ever before, and this agility requires a new way of working: a way that allows involvement of a wide variety of information workers without a long time to develop a new system whenever new information comes to light (ALLEMANG, 2010).

In order to obtain the sufficient knowledge to understand both the internal and external environment as a whole, organizations perform coordinated actions to seek, treat, distribute and protect information (TARAPANOFF, 2006). The set of activities performed by organizations to gather information about competitors, products and markets, is called Competitive Intelligence (MORESI, 2006). It is also seen as seeking any information which improves the organization positioning (TARAPANOFF, 2006).

Although competitive intelligence is compared to industry spying, 90% of information required for understanding competitors and markets are public and available for everyone (QUEYRAS; QUONIAM, 2006). According to Porter (2001), Internet technology provides better opportunities for companies to establish distinctive strategic positionings than did previous generations of information technology. The final purpose is to provide right information, at the right time, to the right person, aiming to create conditions for better decisions (QUEYRAS; QUONIAM, 2006).

Organizations refer to external data sets for essential information that is not available from within the corporate boundaries. For example, volcanical ash disturbance may result in changes of project execution, project management decisions, and customer relationship management (HU; SVENSSON, 2010). Natural disasters can lead to major changes in supply chain management. The political situation of a region or the existence of airports in the proximity are also examples of external information that may lead to strategic changes. To gain access to information like that, organizations usually pay for press clipping services.

The importance of such external data will not be fully demonstrated if they are not combined with internal enterprise data and consumed in realtime business decision making (HU; SVENSSON, 2010). But, specially in large organizations, enterprise data is available through a highly complex enterprise-wide IT system, as several hundreds of interconnected
systems may be employed (LINDSTRôM et al., 2006). And the size of each single system
may vary extensively from enterprise resource planning systems to custommade niche
products, making the system interconnections numerous and heterogeneous.

Semantic Web works to create an environment where people work together sharing
knowledge and having tools for information management and analysis. (PAN, 2009). Hu e
Svensson (2010) state that Linked Data initiative was proposed to remove the barriers to
data access and sharing. Linked Data enables data from different sources to be connected
and queried. Furthermore, LD makes possible large scale integration of, and reasoning on,
data on the Web. These characteristics could benefit competitive intelligence professionals.

1.1 Research Problem

This research intends to answer the following research question: What are the
concepts and technologies related to linked data which allow gathering, integration and
sharing of information to support competitive intelligence?

1.2 Motivation

This research theme has been choosen firstly because Linked Data subject is novel:
a large majority of existing frameworks, related works and references were published in
the last ten years. Another reason for choosing this theme is the expected academic
contribution. The possibility to apply the Linked Data concepts and techniques in a large
company is the third reason for choosing this theme. Finally, the results of this research
can contribute for reducing costs to obtain information beyond organization boundaries.

1.3 Objectives

1.3.1 Main Objective

The main objective of this research is to analyse the application of linked data
technologies into gathering and integrating data generated both inside and outside an
organization for supporting the competitive intelligence process, in the context of portfolio
management.

1.3.2 Specific Objectives

The specific objectives are:

1. OBJ1 - Outline the conceptual framework to support this research;
2. OBJ2 - Identify competency questions to be solved by the proposed framework;

3. OBJ3 - Build prototype for integrated data visualization, based on proposed framework;

4. OBJ4 - Test and validate conceptual framework, using competency questions;

1.4 Adherence to FUMEC’s Research Program

According to Bechhofer et al. (2013), studies continue to show that research in all fields is increasingly collaborative. Most scientific and engineering domains would benefit from being able to borrow strength from the outputs of other research. For Tarapanoff (2006), the information science interdisciplinarity is a recurring theme in area’s literature. The use of information science in other areas leads to the creation of subdisciplines which merge to pre-existing disciplines: physics information, molecular computation, biomedical information, artificial intelligence and digital social networks.

Queyras e Quoniam (2006) define information as a complex object, created by man in a cycle which encompasses cognitive process. The information grouping generates useful knowledge, which in turn enables intelligence creation. But, according to Tarapanoff (2006), it is not possible create intelligence only by passive access to information. Intelligence must be created and it is throughout this creation process that is possible to build useful systems to organizations.

Knowledge is defined a group of factors with potential to influence the human thinking and behaviour and which sometimes enables the explanation, prediction and control of physical phenomenon (OLIVEIRA, 2011). Knowledge management is a set of processes aiming to create, storage, distribute and use the knowledge, considering internal and external knowledge sources (MITJA, 2011). Organizational knowledge includes the experiences of human resources, lessons learned, manuals, etc (MORESI, 2006).

The strategic administration is supported by information Management, knowledge management and intelligence (TARAPANOFF, 2006). Software engineering is a cross-disciplinary subject. It stretches from technical issues such as databases and operating systems, through language issues, for example, syntax and semantics, to social issues and psychology (WOHLIN et al., 2012). This research is classified as cross-disciplinary because its contents belong to several disciplines, such as: Strategic Administration, Portfolio Management, Knowledge Management and Information and Systems Technology.
1.5 Thesis Structure

This work is structured in eight chapters. Chapter 1 provides information for understanding the problem and presents the research question along with the objectives to be achieved. Chapter 2 presents a systematic literature review which is used to conceptually support this research. Firstly, theoretical background is presented highlighting two main constructs: Competitive Intelligence and Semantic Web. Secondly, results obtained from literature are presented.

Chapter 3 presents other studies related to this thesis. Chapter 4 presents both the methods used in this study and the research object. Chapter 5 presents the development and the results of focus group conducted with the objective of identifying information relevant for external scan analysis. Chapter 6 presents a framework for application of Linked Data in corporate environments. Firstly the semantic repository is described. Secondly, a Web application named DB4Trading is itemized.

Chapter 7 presents the validation of Linked Enterprise Data framework. Lastly, Chapter 8 presents the conclusion of this thesis.

1.6 Communications of this Thesis

We have communicated the research presented in this thesis through the conference papers listed next. All communications are related to Chapter 2.


2 Systematic Literature Review

2.1 Introduction

According to Kawase et al. (2013), Over the past years we have witnessed the Web becoming an established channel for learning. Nowadays, hundreds of repositories are freely available on the Web aiming at sharing and reusing learning objects, but lacking in interoperability. The explosion in the number of resources available on the global computer network is a challenge for indexing and searching through a continually changing and growing *database*. (LI, 2002). The existence of huge amounts of data on the Web has raised an increasing challenge to locate right information at right time, as well as to integrating information effectively to provide a comprehensive source of useful information. (COSTILLA et al., 2004).

For Li (2009), the Semantic Web is an evolving extension of the World Wide Web in which contents can be expressed not only in natural language, but also in a format that can be read and used by software agents, thus permitting them to find, share and integrate information more easily. The goal of the semantic Web initiative is to provide an open infrastructure for intelligent agents and Web services. This infrastructure is based on formal domains that are linked to each other on the Web. (HA; LEE, 2006).

For Thuraisingham (2003), data has become a critical resource in many organizations, and therefore, efficient access to data, sharing the data, extracting information from the data, and making use of the information has become an urgent need. As a result, there have been many efforts on not only integrating the various data sources scattered across several sites, but extracting information from these databases in the form of patterns and trends. The same technologies aimed for organizing the World Wide Web can deliver value to closed environments of large enterprises that suffer similar problems of information overflow and disorganization, creating the Linked Enterprise Data. (WESTERSKI; IGLESIAS, 2011).

In this chapter, it is presented a comprehensive literature review on the state-of-the-art in the research field of Linked Enterprise Data. More precisely, this Systematic Literature Review intends to answer the following research question: What are the applications of Linked Data for Corporate Environments?

This chapter is organized as follows. Section 2.2 describes the theoretical background. Section 2.3 describes the research methodology underlying this systematic review. Section 2.4 presents the results obtained. Section 2.5 presents the threats to validity of this SLR. Finally, section 2.6 Concludes the chapter.
2.2 Theoretical Background

2.2.1 Competitive Intelligence

Organizations today seek to thrive in turbulent times. As environments become increasingly volatile, organizations are turning their gaze to the external horizon, watching and grappling with a confusion of signals, messages, and cues. Sensing and making sense of the environment have become the *sine qua non* for organizational growth and survival. (BOUTHILLIER; SHEARER, 2003). However, a great deal of data is available through direct and open channels. These channels include: 1) Published sources; 2) Governmental data; 3) Other public documents. (GELB et al., 1991).

Intelligence gathering goes on everyday, without necessarily being called by its rightful name. More specifically, it implies legal research efforts by business studying their competitor’s products, organizations and related matters. As this gathered intelligence is used to help business positioning, it is called *Competitive Intelligence (CI)*. CI is defined as the use of public sources to develop information on competition, competitors, and the market environment, including economic, regulatory, political, demographic influences, etc. (CRONIN et al., 1994).

For McCrohan (1998), Competitive Intelligence runs the full gamut from identification of corporate goals, strengths, weaknesses, personnel, product and market entry plans and strategies, etc. In short, how competitors have behaved, how they are behaving, and in a given environment how they are likely to behave. CI also includes information collected on many actors and situations relevant to a competitive landscape, such as customers, suppliers, and relevant technologies. (MARIADOSS et al., 2014). CI aims to monitor an organization’s external environment for information relevant to its decision-making process. Many major companies have formal and well-organized CI units that enable managers to make informed decisions about critical business matters such as investment, marketing, and strategic planning. (CHEN; CHAU; ZENG, 2002).

According to Bouthillier e Shearer (2003), the more recent approaches to strategy involve the positioning of a business to maximize the value of its capabilities in order to distinguish it from its competitors. Competitive Intelligence has a strategic role, and corporate strategic plans cannot be fully developed without a good grasp of the trends in a given industry and the activities of one’s competitors. CI is considered an essential part of any corporate system aimed for providing strategic information. Strategic planning requires detailed information and assessment of both the internal and external environments of an organization.

Competitive Intelligence is, at the same time, both a **Process** and a **Product**. It is a Process because it is a systematic and ethical program for gathering, analyzing, and managing external information that can affect the company’s plans, decisions, and
operations. It is also a Product because it is essential to have an application that can handle various formats such as text, tables, graphs, pie charts, photographs, drawings, or even multimedia files with sound, in order to convey intelligence in the many ways CI requires. (BOUTHILLIER; SHEARER, 2003).

To support CI as a process, there must be corporate mechanisms for following up on such information that trigger the initiation of a structured search process to complete and update information from individuals. (DROTT, 2001). In this perspective, CI Professionals transform raw data and information into intelligence by collecting and organizing information resources, a fundamental step, but also by extracting information and adding value by assessing the relevance of the information in a particular context. (BOUTHILLIER; SHEARER, 2003).

To support CI as a product, organizations need to develop effective intelligence systems for the collection and analysis of information. (EWUSI-MENSAH, 1989). For Bouthillier e Shearer (2003), information services and systems are the mechanisms used for delivering, storing, organizing, and packaging information. They are an integral part of an organization’s intelligence system and should be concerned with the acquisition, analysis and communication of information within the organization. (EWUSI-MENSAH, 1989). Any corporate information is useless until it is distributed - the right information at the right time in the right place. (DROTT, 2001).

According to Assaf et al. (2012), companies have traditionally performed business analysis based on transactional data stored in legacy relational databases. However social media feeds, weblogs, sensor data, or data published by governments or international organizations are nowadays becoming increasingly available. The Internet has the potential to become a major strategic information tool for commercial enterprises: Companies of all shapes and sizes are finding that the Internet provides new opportunities for competitive advantage. (CRONIN et al., 1994). Although the Internet represents significant CI opportunities, it has also brought about many technical, cognitive, and organizational challenges. As the amount of information available on the Internet is overwhelming, CI Professionals are constantly facing the problem of information overload. (CHEN; CHAU; ZENG, 2002).

In order to perform Competitive Intelligence, Companies need to integrate two kinds of information: 1) Data from internal systems - which are not always integrated; 2) Data generated outside its boundaries - which are not always structured or standardized. In the next section, Semantic Web is presented as a solution for increase Competitive Intelligence through interlinking heterogeneous data sources.
2.2.2 Semantic Web

Semantic Web has two main motivators: the first is data integration, which is a significant bottleneck in many IT applications; the second is more intelligent support for end users. (HARMELEN, 2004). Nowadays, the form of Web content is designed for humanity but not machine to understand. While current Web works well for posting and rendering content of all sorts, it could provide only very limited support for processing Web contents. Users have to spend large amounts of time in searching, accessing, extracting, interpreting, and processing information on the Web in order to find out how the documents are related to each other. (YALAN; JINLONG; MI, 2006).

The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines - not just for display purposes, but for using it in various applications. (XU et al., 2004). It is a web that includes documents, or portions of documents, describing explicit relationships between things and containing semantic information intended for automated processing by machines. (LI, 2002). It also carries the promise to make web machine-understandable by enriching available information with logic-based semantics and provide us with a new paradigm for knowledge interchange and sharing. (CHEN; WU, 2003).

According to Niles e Jeremijenko (2001), the fundamental concept of the Semantic Web is to make the information, available on the web, more meaningful by making it accessible to automated tools that can augment our experience. It provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. (CHEN et al., 2012). It also refers to a range of standards, modelling languages and tool development initiatives aimed at annotating Web pages with well-defined metadata, so that the semantics associated with web contexts can be effectively interpreted. (ABRAHAMS; MCGRATH; DAI, 2004).

For Chen et al. (2005), Semantic Web has been proposed as the next generation of the World Wide Web and it adds the machine-understandable ability onto the current Web environment, and also enables the data to be automatically shared, processed, and reused across various applications. It is an important gateway for computers to understand human semantics. Semantic Web also offers a new approach to managing information and applications, where creation and use of semantic metadata is fundamental. This envisages the ability of creating and capturing knowledge in machine understandable format, developing agent grids that can integrate that knowledge, reason about it, and forward results both to humans and other agents. So, Semantic Web can be seen as a cluster of technologies, techniques, protocols and processes. (CHARATSIS et al., 2007). It can also be defined as a collection of technologies for managing linked data on the World-Wide Web. Its vision is to transform the web into one global database, using URIs and common data models. (EISENBERG; KANZA, 2011).
According to Thuraisingham (2003), Semantic Web can be thought of as a web that is highly intelligent and sophisticated and one needs little or no human intervention to carry out tasks such as scheduling appointments, coordinating activities, searching for complex documents as well as integrating disparate databases and information systems. It is the "web of data or things" which knows the context of text and also is capable of making a logical relation between such things. It enables computers to understand the text written on them. (KHAN; HUSSAIN, 2009).

Another perspective is deal with Semantic Web as a graph in which the nodes are anything that can be named and the labeled edges are meaningful properties that describe the relationships between the nodes. (WILKINSON; VANDERVALK; MCCARTHY, 2009). It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners and it is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming. (NIKRAVESH, 2006). Since the data is described in a clear way with universally unique identifier, there will be no data ambiguity and machines can easily understand the meaning of data presented on the web. Such data can be dynamically linked with the help of ontologies. (VASILIU; SAKPOTA; KIM, 2006).

The initiatives underlying the semantic web establish a set of semantic technologies that promise to improve the categorization and association of data thereby enhancing the ability to create relationships and to generate inferences among diverse systems and data. (IZZA; VINCENT; BURLAT, 2008). Its techniques provide an answer to the request for the evolution from a chaotic state generated by the production of great volumes of data coming in online activities to a state characterized by information sharing, cooperation and collaboration. (CAPUANO et al., 2010). Because of this, Semantic Web is a well-established branch of computer science and software engineering with its own standards, languages, technologies and applications. It is also a foundation for what is termed "Web Science", where the Web itself is the object of a dedicated science of its own when it is deployed in a wide range of domains. (BRESLIN et al., 2010).

According to Bussler (2008), the greatest promises of the Semantic Web are:

1. Seamless interactions among agents (people and services) based on reliable communications and uniform data and process semantics;
2. A solution to the heterogeneity and interoperability problem in data and processes via dynamic and automatic discovery and integration;
3. Semantic correctness and dependability (including trust and explanation of reasoning results);

For Chen et al. (2007), the emerging semantic web standards provide the web
services infrastructure with the semantic interoperability that integration needs. Generally, different data sources may have different structures, they may vary in syntax and semantics and may differ as to the query language features, organizations, availability and may contain data with different schemata. In the Semantic Web domain, the problem of integration of heterogeneous data sources may be considered as the problem of integrating data obtained from various sources on the Web. (MESTROVIC; CUBRILO, 2006). The Semantic Web is a reality for the development of personal computing through the exposition and sharing of information. The idea is that information is globally ubiquitous, linkable and interoperable. (OLIVER et al., 2009).

2.2.2.1 Semantic Web Technologies

According to Harmelen (2004), to achieve its goals, Semantic Web must meet the challenges of providing:

1. A syntax for representing metadata;
2. Vocabularies for expressing the metadata;
3. Metadata for lots of Web pages;

According to Willer e Dunsire (2013), Semantic Web adds a layer of information with well-defined meaning, or semantics, on top of the syntactical infrastructure of HTTP and HTML for the machine-processing of Web documents. These authors also say that the Semantic Web is based on Resource Description Framework (RDF), a set of specifications published by the World-Wide Web Consortium (W3C) which maintains standards for the Web.

Figure 1 shows the Semantic Web Stack, as proposed by W3C Consortium. At the lowest level, Semantic Web has the protocols for communication including TCP/IP (Transmission Control Protocol/Internet protocol), HTTP (Hypertext Transfer Protocol) and SSL (Secure Socket Layer). The next level is the XML (eXtensible Markup Language) layer that also includes XML schemas. The next level is the RDF (Resource Description Framework) layer. Next come the Ontologies and Interoperability layer. Finally at the highest-level Semantic Web has the Trust Management layer. (THURAISINGHAM, 2003).

The Web is an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URIs). URIs are a cornerstone of the Web architecture, providing identification that is common across the Web, and are tied to a unique definition that everyone can find on the Web. The resource is identified by the URI, the information is presented by HTML, and HTTP typically is used for transmitting the messages. (YING-HUI; GUAN-YU; XING,
2008). A URI is a string of characters, numbers, and punctuation marks, and it functions as a global, unique, machine-readable identifier and is not intended to convey any intrinsic meaning. (WILLER; DUNSIRE, 2013).

According to Abrahams, McGrath e Dai (2004), a number of Semantic Web markup languages have been developed in recent years. These languages provide a means to implement Web site annotations. HTML and XML are presented at the bottom of the pile. At the next level, the Resource Description Framework (RDF) is presented. RDF is a directed, labeled graph data format for representing information on the Semantic Web. It is often used to represent metadata about digital artifacts, as well as to provide a means of integration over disparate sources of information. (CHEN et al., 2012). RDF model and syntax is based on a triplet representation consisting of subject, verb, and object. The 'subject' and 'object' can be any resource. Resources are identified by using the URI specification. The 'verb' can be interpreted as a predicate, a property, or a relation. Thus, a verb can impose a relation between resources as well as a restriction of the relation. (KULVATUNYOU; IVEZIC, 2002). The RDF data model is not tied to a particular syntax. It can be expressed in any syntactic representation; it can also be extracted from non-RDF data sources. The XML serialization syntax of RDF is hard to understand, but RDF application programming interfaces (APIs) are supposed to shield developers from the details of the serialization syntax and to let them handle RDF data as graphs. (DECKER; MITRA; MELNIK, 2000).

Ontology is defined as 'explicit specification of a conceptualization'. It is (meta)data schema, providing a controlled vocabulary of concepts, each with an explicitly defined and
machine-processable semantics. (LI, 2009). Ontology technologies organize system knowledge in conceptual domains according to its meaning. In the semantic web field, ontologies provide shared domain conceptualizations representing knowledge by a vocabulary and, typically, logical definitions to model the problem domain as well as the solution domain. (PARREIRAS, 2012). **OWL** is a Web ontology language which is used to publish and share explicit and common descriptions of domain knowledge and provide support for efficient knowledge management. (LI, 2009).

According to Breslin et al. (2010), **SPARQL** (Protocol And RDF Query Language) aims to query published RDF metadata, and provides both a query language and a protocol for accessing RDF data. SPARQL can be thought of as the SQL of the Semantic Web, and offers a powerful means to query RDF triples and graphs. SPARQL is a graph-querying language, which means that the approach is different than SQL where people deal with tables and rows. Since query patterns are based on the RDF graph model, advanced queries are made possible such as "find every person who knows someone who knows someone else interested in Semantic Web technologies". SPARQL provides a graph pattern matching based paradigm for flexible RDF data graphs. (LI, 2009).

Since data sources can have completely different structures, it is important to include a module between the sources, which would enable data translation, called **Wrapper**. Along with translating data into the source language, the wrapper should upgrade query features of the source if these do not support requested queries. (MESTROVIC; CUBRILO, 2006). Typical information integration systems have adopted a wrapper-mediator architecture. In this architecture, **Mediators** provide a uniform user interface to query integrated views of heterogeneous information sources. Wrappers provide local views of data sources in a global data model. (GARDARIN; DANG-NGOC, 2004).

All technologies aim integrate, share and reuse data. On the Semantic Web, **Data** is modelled as a set of relationships between resources, and with inference, new relationships based on the data and some additional information will be generated automatically with vocabularies and a set of rules. (CHEN et al., 2012). **Metadata** is "data about data" of any sort. Metadata provides context for data, and is used to facilitate the understanding, characteristics and management of data. In data processing, metadata is definitional data that provides information about the data managed within an application or environment. (CARSTOIU et al., 2010). The Semantic Web represents facts through the use of metadata that is associated with Web resources, and behind this metadata there are specific vocabularies that describe what are the semantics (or meaning) of this metadata and how it is all related to each other. (BRESLIN et al., 2010). The **Crawler** component is responsible for locating and fetching data on the Web. This component systematically browses the World Wide Web, typically for the purpose of Web indexing. It also can extract data from all visited pages for later processing. (HOGAN; HARTH;
BRESLIN, 2005).

Different infrastructure systems for semantic web have emerged with the popularity of Semantic Web. It’s obvious that more and more infrastructure systems will appear with the progress of semantic web. (XU et al., 2004). Semantic Web technologies have been maturing to make e-commerce interactions more flexible, reliable and automated. Based on the underlying ontology, the semantic web provides automated processing and integration on the information available on the web. (YUEH et al., 2007).

2.2.2.2 Linked Data

According to Willer e Dunsire (2013), for the purposes of the Semantic Web, linked data is usually represented in RDF. But the authors recognize that, in general, linked data is just data that is interlinked in some way to make it more useful. Normalized relational databases are a form of linked data, and there is a variety of methods for transmitting such structured data over the Internet, including the open standard Javascript Object Notation (JSON) and eXtensible Markup Language (XML).

Linked Data (LD) is a term describing a set of best practices to facilitate the publishing, accessing and interlinking of the data of the Semantic Web. Thus, Linked Data is an open framework for the loose integration of data in the Internet, where data sources can easily cross-link. (FERIDUN; TANNER, 2010). LD is the data exposed, shared, and connected via URIs on the Web. It uses URLs to identify things as resources to facilitate people to dereference them. It also provides useful information about these resources, as well as links to other related resources which may improve information discovery. (MI et al., 2009).

According to Thoma, Sperner e Braun (2012), the Linked Data is based on the idea of using the Web to create typed links between data from different sources. It is described in a machine-readable way with an explicitly defined meaning. LD can be linked to and linked from other external data. Linked Data is a promising concept for providing and retrieving widely distributed structured data. It uses the principles and technologies of the Semantic Web to publish, interlink and query data. Using a subset of the Semantic Web stack, Linked Data can provide a common framework for inter-organizational collaboration and data integration. Based on formal ontologies, different information spaces can be integrated in a large data cloud using semantic relations. The cloud can then be accessed using standard methods and web technologies. (ZIEGLER; GRAUBE; URBAS, 2012).

Linked data technology uses web standards in conjunction with four basic principles for exposing, sharing and connecting data. (FERIDUN; TANNER, 2010; THOMA; SPERNER; BRAUN, 2012; CURRY et al., 2013). These principles are:
1. Use URIs as names for things;

2. Entity URIs should be dereferenceable via HTTP;

3. For a given entity URI, the Linked Data provider should respond with 'useful information' about the entity in the form of RDF triples;

4. The provider should include links to additional, related data URIs in the response in order to maximize the interlinking of the web of data;

There are roughly two approaches to fulfill the LD vision, namely community-driven and data-driven. **Community-driven** tries to fulfill the data request of a community, e.g. movie fans, gene researchers, etc. **Data-driven** starts with a set of core data and tries to establish connections with as many relevant data sets as possible to emerge patterns not possible to individual data sets alone. (HU; SVENSSON, 2010). The Linked Data paradigm involves practices to publish, share, and connect data on the Web, and offers a new way of data integration and interoperability. The basic principles of the Linked Data paradigm are: (a) use the RDF data model to publish structured data on the Web, and (b) use RDF links to interlink data from different data sources. (GALIOTOU; FRAGKOU, 2013)

### 2.2.2.3 Linked Open Data

A pragmatic vision of the Semantic Web has emerged via the Linking Open Data project (LOD), focusing on translating various datasets available on the Web into RDF and interlinking them, following the Linked Data principles. Lots of different datasets have been provided via this LOD initiative, such as DBpedia (the RDF export of Wikipedia) or Geonames (a large geolocation database). All together, they form a complete Web-scaled graph of interlinked knowledge, commonly known as the Linked Open Data Cloud. (PASSANT et al., 2009). Figure 2 presents the LOD.

For Willer e Dunsire (2013), Open data is data that is freely available for anyone to use and re-publish. An open data environment is essential if the Semantic Web is to provide a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. Open linked data, usually referred to as Linked Open Data (LOD), is just linked data that is open data. Linking Open Data (LOD) initiative is a community project focusing on providing interlinked RDF data from existing open sources, leading to the availability of billion of resources and triples on the Web, and based on the Linked Data principles. (BRESLIN et al., 2010).
2.2.2.4 Linked Enterprise Data

According to Eisenberg e Kanza (2011), another goal of Semantic Web is to use the Semantic Web technologies for Enterprise Information Integration, by considering all the data of an enterprise, or of a closed community, as one global database, creating a Corporate Semantic Web. Today’s business is based on huge amount of information and extracting right information at right time is a difficult and tedious task. By applying semantics within structured (ERP, Billing, Financial, HR systems) and unstructured data (email, fax, office documents) we can take business decision on the basis of overall organization knowledge base. (KHAN; HUSSAIN, 2009).

Corporate Semantic Web uses a corporate ontology to drive the semantic annotation of organizational data and thus facilitate data retrieval, integration, and processing. (D’AQUIN et al., 2008). Linked Data (LD) initiative has demonstrated its value through a variety of projects aiming at improving data accessibility for primarily public and academic users. Because of this, cautious attempts were made to experiment the LD principles and to evaluate the benefits, leading to the so-called "Linked Enterprise Data" paradigm, the counterpart of LD in the business domain. (HU; SVENSSON, 2010)

Linked Enterprise Data (also called as "Linked Data Enterprise") is an organization in which the act of information creation is intimately coupled with the act of information sharing. In a linked data enterprise, individuals and groups continue to produce and con-

Figure 2 – Linked Open Data Cloud
Source: Adapted from W3C
sume information in ways that are specific to their own business needs, but they produce it in a way that can be connected to other aspects of the enterprise. (ALLEMANG, 2010).

For Breslin et al. (2010), when data is represented using RDF and can be accessed with SPARQL, queries can be created that are relevant to a particular organization, this can bring organizations into a 'Linked Enterprise Data' framework. LED aims for both expose and link enterprise data, while shows that there are benefits in terms of solutions that can be made available immediately. According to Halb et al. (2010), LED participating entities are persons, enterprises, associations, and research institutes owning at least one of the following roles:

1. **Linked Data Providers** - provide any kind of data in a Linked Data format. They consume raw data provided by Raw Data Providers and transform it into Linked Data;

2. **Linked Data Application Providers** - provide Linked Data Applications. They consume Linked Data provided by Linked Data Providers, process it within their applications and transform it into Human-Readable-Data;

3. **End users** are humans who (like to) consume Human-Readable Data, which is a human-readable presentation of Linked Data provided by Linked Data Application Providers;

Linked Data also offers a big potential for enterprises in the media and knowledge management area. More and more datasets are published following Linked Data principles and thus give the opportunity to link enterprise content with background information. Linked Data can also offer a comparably simple solution for enterprise information integration inside companies. (SCHAFFERT et al., 2012). Linked data technology can be accommodated with minimal disruption to existing information infrastructure, as a complimentary technology for data sharing, and should not be seen as a replacement for current IT infrastructure (relational databases, data warehouses, etc.). The objective is to expose the data within existing systems, but only link the data when it needs to be shared. (CURRY et al., 2013).

Linked Data technology will enable enterprises to act as Linked Data Providers, providing Linked Data for third parties, and to act as Linked Data Application providers, consuming data from third parties and providing more valuable Human-Readable-Data on top of it for the human enduser. (HALB et al., 2010).
2.3 Design

This Systematic Literature Review aims to analyze publications related to the applications of Linked Data for Corporate Environments, by executing a systematic literature review. According to Kitchenham (2004), a SLR is a means of evaluating and interpreting available research relevant to a particular research question, topic area, or phenomenon of interest. This kind of study comprises three consecutive phases: planning, execution and results. This Systematic Literature Review is based on guidelines proposed by Kitchenham (2004) in order to ensure its validity.

2.3.1 Planning

The main purpose of this phase is to deliver a protocol which drives the research efforts. There are four stages associated with planning phase: (1) Research Background, (2) Research Questions, (3) Research Strategy and (4) Data Extraction Strategy.

Research Background. In this stage the existing literature related to Semantic Web its application on corporate environments was analyzed. Table 1 shows the theoretical foundations upon which we based this study on.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic Web</td>
<td>(WILLER; DUNSIER, 2013)</td>
</tr>
<tr>
<td></td>
<td>(PARREIRAS, 2012)</td>
</tr>
<tr>
<td></td>
<td>(BERNERS-LEE; HENDLER; LASSILA, 2001)</td>
</tr>
<tr>
<td>Linked Data</td>
<td>(GALIOTOU; FRAGKOU, 2013)</td>
</tr>
<tr>
<td></td>
<td>(HEATH; BIZER, 2011)</td>
</tr>
<tr>
<td></td>
<td>(BECHHOFER et al., 2013)</td>
</tr>
<tr>
<td>Linked Enterprise Data</td>
<td>(ALLEMANG, 2010)</td>
</tr>
<tr>
<td></td>
<td>(HU; SVENSSON, 2010)</td>
</tr>
<tr>
<td></td>
<td>(ALLEMANG, 2010)</td>
</tr>
</tbody>
</table>

Table 1 – SLR - Theoretical Foundations

Research Questions. In this stage, the research questions addressed by this study were defined. The main research question addressed by this study is: RQ1. What are the applications of Semantic Web for Corporate Environments? With respect to linked enterprise data, the following issues were considered: RQ1.1 Which kind of data source enterprises who adopted Semantic Web often interlink? RQ1.2 What are the enterprises’s concerns regarding Semantic Web adoption? RQ1.3 What are the similarities between the frameworks used to implement Semantic Web on Enterprises?

Research Strategy. Following guidelines proposed by Kitchenham (2004), in this stage, it was defined the strategy used to search for primary studies, including search terms and resources to be searched. Searches for primary studies were conducted only in
the following electronic databases: Association for Computing Machinery (ACM), Elsevier peer-reviewed full-text articles (ScienceDirect), Institute of Electrical and Electronics Engineers (IEEE), Web of Science (ISIWeb) and SpringerLink. Considering the research questions, three kinds of conceptual search strings were defined. In order to improve this initial structure, some synonyms, related terms and alternative spellings were determined. The final conceptual strings were used as basis to construct the search strings which were converted to each electronic database syntax. Table 2 presents the conceptual strings.

<table>
<thead>
<tr>
<th>Conceptual Search Strings</th>
</tr>
</thead>
<tbody>
<tr>
<td>(&quot;Linked Data&quot; OR &quot;Semantic Web&quot;) AND (&quot;Enterprise&quot;)</td>
</tr>
<tr>
<td>(&quot;Linked Data&quot; OR &quot;Semantic Web&quot;) AND ((&quot;Portfolio&quot; OR &quot;Strategic&quot;) AND &quot;Management&quot;)</td>
</tr>
<tr>
<td>(&quot;Linked Data&quot; OR &quot;Semantic Web&quot;) AND (&quot;Enterprise Application Integration&quot; OR &quot;EAI&quot;)</td>
</tr>
</tbody>
</table>

Table 2 – SLR - Conceptual Search Strings

The advanced search tool of each electronic database was used to convert the conceptual string to the database syntax. Because of the characteristics of each electronic database, it was required to search the terms in different fields. Table 3 presents the specific fields where those terms were searched at each electronic database.

<table>
<thead>
<tr>
<th>Electronic Database</th>
<th>Searched Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM</td>
<td>Title and Abstract</td>
</tr>
<tr>
<td>Science Direct</td>
<td>Title and Abstract</td>
</tr>
<tr>
<td>IEEE</td>
<td>Title and Abstract</td>
</tr>
<tr>
<td>ISIWeb</td>
<td>Title</td>
</tr>
<tr>
<td>Springer Link</td>
<td>Title</td>
</tr>
</tbody>
</table>

Table 3 – SLR - Searched Fields at Databases

Data Extraction Strategy. In this stage the exclusion criteria and the strategy to extract data from selected papers were defined. It was decided to include all studies returned by electronic databases, except those classified in the exclusion criteria presented in table 4.

<table>
<thead>
<tr>
<th>Exclusion Criteria #1</th>
<th>The study is written in other language than English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exclusion Criteria #2</td>
<td>The study is not applying Semantic Web</td>
</tr>
<tr>
<td>Exclusion Criteria #3</td>
<td>The study has already been selected</td>
</tr>
<tr>
<td>Exclusion Criteria #4</td>
<td>The study is not an article or paper</td>
</tr>
<tr>
<td>Exclusion Criteria #5</td>
<td>The study is out of the scope</td>
</tr>
</tbody>
</table>

Table 4 – SLR - Exclusion Criteria
2.3.2 Execution

The main purpose of this phase is to perform activities which extract and synthesize data from papers. There are three stages associated with execution phase: 1) Select Primary Studies, 2) Execute Data Extraction and 3) Synthesize Extracted Data.

Select Primary Studies. The automatic search resulted 479 studies potentially being relevant for this research. SpringerLink delivered 58% of the results, followed by IEEE (27%), ISIWeb (7%), ScienceDirect (6%) and the ACM (2%). An initial manual filtering was then executed as a first step to refine the results, according to the defined data extraction strategy. The exclusion criteria were applied based on the abstract, introduction and conclusion of each study. From the 479 studies, 214 (about 45% from total) were retained for further detailed analysis. In this second step the exclusion criteria were applied based on the full text. After this detailed analysis, 167 primary studies were included in this review (about 35% from total). Figure 3 presents the reasons for the exclusion of studies from this Review intending to clearly describe how initial filtering and detailed analysis were done. Figure 4 summarizes the transformation of raw results in the final primary studies list.

![Figure 3 – SLR - Reasons for Exclusion of Studies](source: Author)

![Figure 4 – SLR - Transformation of Raw Results](source: Author)
### Execute Data Extraction

In this stage, the information needed to address the questions of this review were collected. After the search and selection procedures, the primary studies were examined. The papers were analysed considering the information required by each research question.

### Synthesize Extracted Data

In this stage, the results of the included primary studies were collated and summarized. The data extracted and synthesized in this stage is presented in the results section, which can be read next.

## 2.4 Results

### 2.4.1 Studies Overview

Table 5 shows studies included in this Systematic Literature Review categorized by publication year.

<table>
<thead>
<tr>
<th>Year</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>(DECKER; MITRA; MELNIK, 2000)</td>
</tr>
<tr>
<td>2001</td>
<td>(NILES; JEREMIJENKO, 2001)</td>
</tr>
<tr>
<td>2002</td>
<td>(LI, 2002); (KULVATUNYOU; IVEZIC, 2002); (DING et al., 2002)</td>
</tr>
<tr>
<td>2003</td>
<td>(TRASTOUR; PREIST; COLEMAN, 2003); (THURAISINGHAM, 2003); (CHEN; WU, 2003); (BUSSLER, 2003);</td>
</tr>
<tr>
<td>2004</td>
<td>(HARMELEN, 2004); (GARDARIN; DANG-NGOC, 2004); (XIANG; MADEY, 2004); (COSTILLA et al., 2004); (MULARZ; LYELL, 2004); (TORRES et al., 2004); (XU et al., 2004); (ABRAHAMS; MCGRATH; DAI, 2004);</td>
</tr>
<tr>
<td>2005</td>
<td>(CHEN et al., 2005); (LEE; MOON, 2005); (WILLMOTT et al., 2005); (HOGAN; HARTH; BRESLIN, 2005); (BRY; MARCHIORI, 2005); (ZAPF; FERNÁNDEZ-GARCÍA; SÁNCHEZ-FERNÁNDEZ, 2005); (AZIZ et al., 2005); (AGARWAL et al., 2005);</td>
</tr>
<tr>
<td>2006</td>
<td>(GOTH, 2006); (YANG; ZHANG; LOW, 2006); (KINGAS; MTSKIN, 2006); (VASILIU; SAKPOTA; KIM, 2006); (DOU et al., 2006); (ZHUO; WANG, 2006); (MESTROVIC; CUBRIOLO, 2006); (HA; LEE, 2006); (LOSADA et al., 2006); (YAN; JINLONG; MI, 2006); (DONG; TONG; FENG, 2006); (GIMÉNEZ et al., 2008); (NIKRAVESH, 2006); (BURAGA; RUSU, 2006); (LEE; KIM, 2007); (LIU; SHI, 2008); (SHARMA; GUPTA; WICKRAMASINGHE, 2006);</td>
</tr>
<tr>
<td>2007</td>
<td>(CHEN et al., 2007); (KANG; LAU, 2007); (YUEH et al., 2007); (DU; LI; DONG, 2007); (OBERHAUSER; SCHMIDT, 2007); (OREN et al., 2007); (CHARATSIS et al., 2007); (BACHLECHNER, 2008); (LIN; HARDING, 2007); (ZHANG; TANNIRU; ZHANG, 2007);</td>
</tr>
<tr>
<td>2008</td>
<td>(WEI; KANG; ZHOU, 2008); (QU; REN, 2008); (ZHANG et al., 2008); (BUSSLER, 2008); (JUNG et al., 2008a); (D’AQUIN et al., 2008); (ALANI et al., 2008); (BARHAMGI; BENSILMANE; OUKSEL, 2008); (DJAMEL; DJAMAL; MIMOUN, 2008); (MAHMoud; GOMEZ, 2008); (JUNG et al., 2008); (LEE et al., 2008); (ZHANG; XU; LI, 2007); (MOREAU; MALENFANT, 2008); (YIN; HUI; GUAN-YU; XING, 2008); (KHAN; HUSSAIN, 2009); (KAEMARIN; ARCH-INT; ARCH-INT, 2008); (JAGDEV et al., 2008); (TANG; LIU, 2008); (RAGI; ROSA; CARDOSO, 2008); (IZZA; VINCENT; BURLAT, 2008); (HU-ANG; DIAO, 2008); (BUSSLER, 2008); (BOYLE et al., 2008);</td>
</tr>
<tr>
<td>2009</td>
<td>(SAMARAS et al., 2009); (PASSANT et al., 2009); (NAYAK et al., 2009); (LOI; FENSA; SENATORE, 2009); (LI, 2009); (TAN et al., 2009); (MI et al., 2009); (LEMEN; XIAOMI; LIUQI, 2009); (MARJIT et al., 2009); (MOUTSELAKIS; KARAKOS, 2009); (FENG; JIA; YANG, 2009); (SUN et al., 2009); (CIU-MEI, 2009); (OLIVER et al., 2009); (XUANG et al., 2009); (WILKINSON; VANDERVALK; MCCARTHY, 2009); (RAO; SARDINHA; SADEH, 2009);</td>
</tr>
</tbody>
</table>
Table 5 – SLR - Studies by Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>(HALB et al., 2010); (CAPUANO et al., 2010); (CARSTOIU et al., 2010); (WEN-YUE; HAI-CHENG; HONG, 2010); (FERIDUN; TANNER, 2010); (ZHANG; MA; WANG, 2010); (QIU; SONG; MA, 2010); (HEIDARY et al., 2010); (PAGE et al., 2010); (BRESLIN et al., 2010); (BILETSKY; RANGANATHAN, 2010); (LI; YAN; JIAN, 2010); (BARJIS; 0004; SHARDA, 2011); (HUANG et al., 2010); (TANG et al., 2010); (NAJAFI; SARTIPI, 2011); (PAHLKE; BECK; WOLF, 2010); (BINSARDI; GREEN; JACKSON, 2013); (GHANI et al., 2010);</td>
</tr>
<tr>
<td>2011</td>
<td>(WESTERSKI; IGLESIAS, 2011); (DABROWSKI; GRIFFIN; PASSANT, 2011); (EISENBERG; KANZA, 2011); (MORARU; FORTUNA; MLADENIC, 2011); (FRANKE et al., 2011); (HEINO; TRAMP; AUER, 2011); (WALSH et al., 2011); (RAAFAT; CECELJA, 2011); (KASIMATI; ZAMANI, 2011); (SLEIMAN; RIVERO; CORCHUELO, 2011); (DANYARO; JIAAFAR; LIEW, 2011); (KODESWARAN et al., 2011); (CHI, 2010); (WILLIAMS et al., 2012); (DONG et al., 2008); (NAJAFI; SARTIPI, 2011);</td>
</tr>
<tr>
<td>2012</td>
<td>(ASSAF et al., 2012); (SCHAFFERT et al., 2012); (CHEN et al., 2012); (BENAISSA; BENABDELHAFID, 2012); (THOMA; SPERNER; BRAUN, 2012); (QUASTHOFF; MEINEL, 2012); (PIEDRA et al., 2012); (ASSAF; SENART, 2012); (ERDUR et al., 2012); (SCHADE; SMITS, 2012); (ROBAK; FRANCZYK; ROBAK, 2012); (ZIEGLER; GRAUBE; URBAS, 2012); (XIN; HUI, 2012); (JELENKOVIC; TOSIC, 2012); (CARBONE et al., 2012); (WANG et al., 2012); (SZYDLO; ZIELINSKI, 2012); (LI; WEI, 2012); (VAISHNAV; KUECHLER JR., 2005); (YANG et al., 2012); (HSIEH; LIN, 2012); (MUELLER, 2012);</td>
</tr>
<tr>
<td>2013</td>
<td>(RUSITSCHKA et al., 2013); (KAWASE et al., 2013); (ZADEH; REFORMAT, 2013); (BIANCHINI; ANTONELLI; MELCHIORI, 2013); (FREITAS et al., 2013); (SOWMYA; ANANTHANARAYANA, 2013); (BACCAR; ROUACHED; ABID, 2013); (RUTA et al., 2013); (ALOR-HERNANDEZ et al., 2013); (ELGAYAR; DEOKAR, 2013); (CURRY et al., 2013); (ZDRAVKOVIC et al., 2013); (WU; ZHU; ZHOU, 2013); (SALMEN et al., 2013);</td>
</tr>
<tr>
<td>2014</td>
<td>(COUTINHO et al., 2014)</td>
</tr>
</tbody>
</table>

2.4.2 Research Questions Evaluation

2.4.2.1 RQ1 What are the applications of Semantic Web for Corporate Environments?

Studies showed that enterprises seem to be focused on using Semantic Web to support Enterprise Application Integration (EAI). A large majority of studies (65%) intended to use Semantic Web technologies to solve systems interoperability issues, such as: Web Services, Systems Integration, B2B Collaboration, etc. Part of studies (12%) intended to explain Semantic Web concepts or suggest any kind of optimization on existing concepts. A portion of studies (10%) focused on integration of things, such as: mobiles, tablets, RFID devices, automation devices, among others. Another portion of studies (12%) intended to create or optimize Knowledge Management Systems through Semantic Web. A small portion (5%) of studies intended to support corporate media integration, such as: streaming, music, among others. Figure 5 summarizes the applications of Semantic Web for Corporate Environments.

2.4.2.2 RQ1.1 Which kind of data source enterprises who adopted Semantic Web often interlink?

Studies showed that enterprises not always use Semantic Web infrastructure to combine data from internal and external data sources. Considering the existence of two kinds of data sources - internal and external - most part of studies (about 51%) pointed
out that enterprises use Semantic Web technologies to interlink data generated both internal and external to their boundaries. However, some enterprises (about 31%) use technologies to interlink internal data sources only. Another part of the studies (about 18%) mentioned the use of Semantic Web technologies to interlink external data only. Figure 6 shows this distribution.

Figure 5 – SLR - Applications of Semantic Web for Corporate Environments
Source: Author

Figure 6 – SLR - Kinds of Data Sources Interlinked
Source: Author
2.4.2.3 RQ1.2 What are the enterprises’s concerns regarding Semantic Web adoption?

Studies showed that enterprises have concerns regarding Semantic Web adoption in corporate environments. According to studies, the major concern (about 34%) is related to the complexity of technologies. In second place (about 23%), the requirement of rigid terminology seems to inhibit the adoption of Semantic Web. The enterprises are also concerned regarding the value added by Semantic Web (about 11%). Figure 7 shows the enterprises’s concerns regarding Semantic Web adoption. All the concerns were categorized based on works of Alani et al. (2008) and Bechhofer et al. (2013). Table 6 explains these categories.

![Figure 7 – SLR - Concerns Regarding Semantic Web Adoption](source: Author)

2.4.2.4 RQ1.3 What are the similarities between the frameworks used to implement Semantic Web on Enterprises?

Studies mentioned frameworks whose layers perform similar functions to those presented in the table 7. Figure 8 presents an arrangement of layers. The layers were categorized based on the work of Oren et al. (2007).

2.5 Threats to Validity

In this Section, the limitations of this Systematic Literature Review are addressed. First, only papers written in English had been considered. Additionally, during data extraction stage, it was necessary to interpret the subjective information provided by studies. This happened because many studies did not present objective details regarding the


<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirement of Rigid Terminology</td>
<td>Enterprises are concerned if Semantic Web adoption will require high standardization on existing terminologies to enable data and information sharing</td>
</tr>
<tr>
<td>Complexity of Technologies</td>
<td>Enterprises are concerned if Semantic Web adoption will demand highly specialized teams to handle technologies</td>
</tr>
<tr>
<td>Cost of Build and Maintenance</td>
<td>Enterprises are concerned if Semantic Web adoption will demand extra costs of build and maintenance</td>
</tr>
<tr>
<td>Low Value-added</td>
<td>Enterprises are concerned if Semantic Web adoption will benefit consumers and competitors but offer no quick wins for themselves, as data providers</td>
</tr>
<tr>
<td>Security issues</td>
<td>Enterprises are concerned if Semantic Web adoption will bring security issues, as unauthorized access, privacy, among others</td>
</tr>
<tr>
<td>Information Provenance Assurance</td>
<td>Enterprises are concerned if Semantic Web adoption will bring novel security issues, such as unauthorized access, privacy, among others</td>
</tr>
<tr>
<td>Information Versioning</td>
<td>Enterprises are concerned if Semantic Web adoption will bring issues related to information versioning and life cycle</td>
</tr>
<tr>
<td>Lack of Documentation</td>
<td>Enterprises are concerned if the existing documentation (ex.: schema mappings between datasets, technologies, etc) is sufficient to support Semantic Web adoption</td>
</tr>
<tr>
<td>Legal Issues</td>
<td>Enterprises are concerned if Semantic Web adoption will bring novel legal issues</td>
</tr>
</tbody>
</table>

Table 6 – SLR - Category of Concerns Regarding Semantic Web Adoption

<table>
<thead>
<tr>
<th>Layer</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Sources</td>
<td>This layer refers to data sources available for integration</td>
</tr>
<tr>
<td>Adapters</td>
<td>This layer provides access to specific types of data</td>
</tr>
<tr>
<td>Wrapper</td>
<td>This layer distributes queries over data sources and aggregates results</td>
</tr>
<tr>
<td>Query</td>
<td>This layer provides an API that is independent of a specific data source and query language</td>
</tr>
<tr>
<td>Integration</td>
<td>This layer provides all the mapping functionality (semantics)</td>
</tr>
<tr>
<td>Presentation</td>
<td>This layer enables an interface to access data</td>
</tr>
</tbody>
</table>

Table 7 – SLR - Frameworks Functional Layers

topics investigated. Another potential threat to validity is the natural limitation of search engines, which may have caused the loss of relevant papers.

2.6 Conclusion

A state of the art survey on the applications of Semantic Web for Corporate Environments has been given in this section. The presented analysis shows that enterprises are using Semantic Web technologies for support their operations.
Studies point out that there is a pattern regarding the frameworks used for implementing Semantic Web in enterprises. This pattern enables interlinking of both internal and external data sources. However, most part of the studies describes the use of Semantic Web technologies to facilitate Enterprise Application Integration (EAI), handling internal data sources only. We believe that studies focusing on benefits brought by interlinking both internal and external data are interesting in this context.

According to studies, enterprises still believe that Semantic Web technologies are complex and require high specialized teams. Enterprises are also concerned if Semantic Web adoption requires changes in existing systems. Thus, we believe that studies focusing on clarify concepts and demystify some of the concerns regarding Semantic Web for corporate environments are interesting in this context.
3 Related Work

Related work were classified into two categories: 1) Regarding the technology used to interconnect different data sources; 2) regarding the source of data - internal or external.

The work of Hu e Svensson (2010) is directly related to this research because, as in this research, their work aims to integrate external and internal data sets to provide strategical information. This research is also related to the research performed by Azevedo (2013), because in this research, the linked data paradigms are used to prevent damages caused by flood. The author uses similar methodological procedures to achieve research’s objective. This research is also related to the research conducted by Silva (2013) which uses linked data paradigms to create an open market for metadata. This work also adopts similar methodological procedures to achieve the main objective.

This research differs from related work because it intends to test if it is possible to use Semantic Web Technologies to gather and distribute at least 90% of information required for external environment analysis. Table 8 presents similarities and differences to related work.

<table>
<thead>
<tr>
<th>Related Work</th>
<th>Similarities</th>
<th>Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>(SILVA, 2013)</td>
<td>- Existence of connection to multiple data sources;</td>
<td>- From corporate perspective, the proposed framework provides internal data to external environment</td>
</tr>
<tr>
<td></td>
<td>- Possibility to perform queries considering data from multiple data sources;</td>
<td></td>
</tr>
<tr>
<td>(AZEVEDO, 2013)</td>
<td>- Existence of connection to multiple data sources</td>
<td>- From corporate perspective, the proposed framework seeks data only from external environment</td>
</tr>
<tr>
<td></td>
<td>- Possibility to perform queries considering data from multiple data sources</td>
<td></td>
</tr>
<tr>
<td>(HU; SVENSSON, 2010)</td>
<td>- Existence of connection to multiple data sources</td>
<td>- Processed data are stored in a semantic repository and are consumed by a business process management system residing in the integration layer</td>
</tr>
<tr>
<td></td>
<td>- Possibility to perform queries considering data from multiple data sources</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- From corporate perspective, the proposed framework seeks data from both external and internal environment</td>
<td></td>
</tr>
</tbody>
</table>

Table 8 – Related Work - Similarities and Differences
4 Methodology

This research can be classified as APPLIED, regarding its nature; DESCRIPTIVE, regarding its main objective; EXPERIMENTATION, regarding its procedures.

For Gil (2002), researches can be classified in two main groups regarding their nature: intelectual purposes and practical purposes. The first group aims the knowledge for the knowledge, and is known as pure science. The second group aims knowledge to solve some practical problem, and is known as applied science.

It is usual to classify the researches regarding their main objectives. According to Gil (2002), it is possible classify researches in three main groups: exploratory, descriptive and explanatory.

Exploratory researches have as main purpose to provide more familiarity with the problem, with the aim of expliciting the problem or build new hipotesis. Descriptive researches have as main purpose to describe the characteristics of a specific population (or phenomenom) (GIL, 2002). This kind of research also aims to establish relationships between variables. A striking characteristic of this kind of research is the usage of standard techniques to obtaining data, such as survey and systematic observation.

Descriptive surveys can be conducted to enable assertions about some population. This could be determining the distribution of certain characteristics or attributes. The concern is not about why the observed distribution exists, but instead what that distribution is (WOHLIN et al., 2012).

Explanatory researches have as main purpose to identify the factors which determine or contribute to the phenomena occurrences. This kind of research explains the reasons of the things (GIL, 2002).

The researches can also be classified regarding their procedures. According to Gil (2002), in order to analyze the facts from the empirical point of view, it is necessary to design the research. The design refers to the research planning, including both the diagramming concerns and the prediction of data collection and analysis.

According to Wohlin et al. (2012), experiments are appropriate to investigate different aspects, including:

- Confirm theories, i.e. to test existing theories.
- Confirm conventional wisdom, i.e. to test people’s conceptions.
- Explore relationships, i.e. to test that a certain relationship holds.
• Evaluate the accuracy of models, i.e. to test that the accuracy of certain models is as expected.

• Validate measures, i.e. to ensure that a measure actually measures what it is supposed to.

4.1 Research Phases

In order to address each specific objective, this research is divided in three phases. Figure 9 shows the adherence to specific objectives.

![Figure 9 - Objectives and Methodologies](Source: Author)

4.1.1 Phase 1 - Systematic Literature Review

The main purpose of this phase is to support the specific objective OBJ1 - 'Outline the conceptual framework'. According to Kitchenham (2004), a SLR is a means of evaluating and interpreting available research relevant to a particular research question, topic area, or phenomenon of interest. This kind of study comprises three consecutive phases: planning, execution and results.
The Systematic Literature Review can be seen at "Chapter 2 - Systematic Literature Review". The conceptual framework used in this research is fully integrated to results obtained in literature review and can be seen at "Chapter 6 - Implementation".

4.1.2 Phase 2 - Focus Group

4.1.2.1 Focus Group Definition

The main purpose of this research phase is to attend the specific objective OBJ2 - 'Identify competency questions' which will be used to validate the framework. Procedures and results of the focus group performed in this research are presented at "Chapter 5 - Focus Group".

According to Blank (2010), a focus group is basically a way to reach out to your potential users for feedback and comment. Organizations generally use focus groups in planning, marketing, or evaluation, either to improve some specific product or service or, more globally, during the development of strategic plans or mission statements. Focus groups answer questions that the development cannot resolve and can lead to new ideas.

A focus group is a data collection procedure in the form of a carefully planned group discussion among about ten people plus a moderator and observer, in order to obtain diverse ideas and perceptions on a topic of interest in a relaxed, permissive environment that fosters the expression of different points of view, with no pressure for consensus. (OMNI, 1996).

Focus groups normally have between seven and ten participants. Groups with fewer than seven participants often result in a limited range of ideas and opinions being represented. Groups larger than ten may be hard to manage and record. (OMNI, 1996). The purpose of focus groups is to gain information about the topic(s) of interest from the perspective of participants. Specifically, a focus group session concentrates on:

- Gathering opinions, beliefs, and attitudes about issues of interest to your organization;
- Testing your assumptions;
- Encouraging discussion about a particular topic;
- Building excitement from spontaneous combination of participants’ comments;
- Providing an opportunity to learn more about a topic or issue

According to OMNI (1996), group dialogue tends to generate rich information, as participants’ insights tend to 'trigger' the sharing of others’ personal experiences and perspectives in a way that can more easily or readily tease out the nuances and tensions
of complex topics and subjects – a dynamic that is not present during key informant interviews.

Focus group provides information directly from individuals who are invested in the issue or hold expert knowledge about a topic of which little is known among researchers. It provides information from people who can provide insights about actual conditions and situations. Focus group provides a representation of diverse opinions and ideas. It also provides a relatively low cost and efficient way to generate a great deal of information. (OMNI, 1996).

4.1.2.2 Focus Group Phases

The activities of focus groups can be divided into three main phases: planning, conducting and reporting. In the planning phase, the purpose of the focus group is defined, the participants are selected and invited, the questions are generated and, finally, the script to support the meeting is elaborated. (BLANK, 2010). In the conducting phase, the participants should be introduced to the main topic and the questions should be asked. The reporting phase is the final phase. Individuals are being invited to participate in focus groups because they are viewed as possessing important knowledge about particular experiences, needs, or perspectives that we hope to learn more about as a result of the needs assessment. (OMNI, 1996).

4.1.2.3 Focus Group Limitations

A focus group has several limitations. Firstly, focus groups are susceptible to facilitator bias, which can undermine the validity and reliability of findings. Secondly, the discussions can be sidetracked or dominated by a few vocal individuals. Finally, focus groups generate important information. However, such information often has limited generalization to a whole population. (KRUEGER, 2002).

4.1.3 Phase 3 - Experimentation

This research phase aims to reach two specific objectives: Firstly, OBJ3 - "Build the prototype for integrated data visualization". Secondly, OBJ4 - "Test and validate conceptual framework" by querying the built prototype. Results of experimentation phase performed in this research are presented both at "Chapter 6 - Implementation" and at "Chapter 7 - Validation".

In a general perspective, this phase includes the following stages: (1) Scoping, (2) Planning, (3) Operation, (4) Analysis and interpretation and (5) Results presentation. According to Wohlin et al. (2012), experiments are launched when we want control over the situation and want to manipulate behavior directly, precisely and systematically. Also,
experiments involve more than one treatment to compare the outcomes. This type of manipulation can be made in an off-line situation where the events are organized to simulate their appearance in the real world.

For Gil (2002), experiments consist of: 1) Define a study object; 2) Select variables capable of influencing it; 3) Define the means to control variables; 4) Observe the effects that variables can cause at the object studied.

Wohlin et al. (2012) also state that experiments may alternatively be made on-line, which means that the investigation is executed in the field in a real life context. It is possible to consider the current situation to be the baseline (control), which means that the baseline represents one level (or value) of the independent variable, and the new situation to be the one we want to evaluate.

4.2 Research Object

With headquarters in Brazil and operations in more than 30 countries, VALE is a global company, with a mission to transform natural resources into prosperity and sustainable development. VALE is leader in the production of iron ore and the second largest producer of nickel. Its primary business is mining. Ores are the ingredients for various items, such as: cell phones, airplanes, building structures, coins, among others.

To ensure the company’s organic growth, Exploration and Mineral Projects Department is in charge of a comprehensive geological research program all over the world focused on the discovery of large mineral deposits, according to VALE’s criterias. But, as large deposits are difficult to be discovered, many years of study are required until reserves areas can be discovered.

The work of Exploration and Mineral Projects Department comprises geological, technological and engineering studies to promote discoveries and evaluate the feasibility of the extraction, beneficiation and transportation of the ore, considering technical, economic, environmental and social features. All disciplines must work in an integrated way so as to create long-term value and provide shared benefits to all stakeholders.

4.2.1 Portfolio Management Enabling Strategy

According to PMI (2013), a portfolio is a collection of programs, projects, or operations managed as a group to achieve strategic objectives. For Jenner, Commerce e Kilford (2011), a portfolio is the totality of an organization’s investment in the changes required to achieve its strategic objectives. This way, the portfolio components may not necessarily be interdependent or have related objectives. Portfolio components are quantifiable, that is, they can be measured, ranked and prioritized. (PMI, 2013).
Chapter 4. Methodology

For Cooper, Edgett e Kleinschmidt (2001), **Portfolio Management** is a dynamic decision process, whereby a business’s list of projects is constantly up-dated and revised. According to Jenner, Commerce e Kilford (2011), it can be defined as a co-ordinated collection of strategic processes and decisions that together enable the most effective balance of organizational change and business as usual. In this process, new projects are evaluated, selected and prioritized; existing projects may be accelerated, killed or de-prioritized; and resources are allocated and re-allocated to the active projects. (COOPER; EDGETT; KLEINSCHMIDT, 2001).

Kolisch, Meyer e Mohr (2005) states that portfolio management embraces two aspects: candidate projects must be selected and the speed with which they are to be accomplished must be decided. High-intensity projects obtain priority access to resources, thereby reducing the project lead time as well as permitting more resources to be assigned as needed. For PMI (2008), the aim of portfolio management is to:

1. guide investment decisions and their appropriate mix;
2. provide decision-making transparency;
3. increase likelihood of realizing desired return on investment

For PMI (2013), Portfolio Management includes interrelated organizational process by which an organization evaluates, selects, prioritizes and allocates its limited internal resources to best accomplish organizational strategies consistent with its vision, mission and values. Thus, there is a direct correlation between the portfolio management and the business strategy. Business that had installed a systematic, explicit portfolio management system have: higher value projects; better balance; the right number of projects and a strategically aligned portfolio. (COOPER; EDGETT; KLEINSCHMIDT, 2001).

In order to ensure alignment between its strategy and its overall research program, VALE uses the Portfolio Management techniques. However, as stated by PMI (2013), the organizational strategy and objectives are translated into a set of initiatives that are influenced by many factors such as market dynamics, customer and partner requests, shareholders, government regulations, and competitor plans and actions. Furthermore, corroborating Ruegg (2007), VALE has numerous business lines that require sophisticated information systems and computational techniques to track all lines and model the risk and uncertainty associated portfolio strategies.

Thus, VALE’s competitive intelligence team, which operates in the portfolio management context, would benefit from a solution able to gather, integrate and share data for external scan analysis.
5 Competency Questions

In order to answer this research’s questions we decided to perform a focus group to identify competency questions. This chapter describes both the procedures and the results derived from the focus group performed in this research aiming to attend the specific objective OBJ2 - "Identify competency questions".

We applied the focus group methodology, which is understood as a way of collecting qualitative data, which involves engaging a small number of people in informal group discussions, 'focused' around a particular topic or set of issues. (HøYLAND; HAUGEN; THOMASSEN, 2014). This qualitative research approach is very effective in determining people’s views, feelings and opinions. (SIMIGIU, 2014). The main advantage of focus group interviews lies in the informal nature of the method, where instead of asking questions to each participant, the moderators actively encourages interaction between group members. (HøYLAND; HAUGEN; THOMASSEN, 2014).

5.1 Focus Group Design

The general purpose of this focus group was to identify the information used by competitive intelligence professionals in order to assess potential places for investing. Specifically, this focus group intended to understand both the kind and the sources of information used.

Focus group have been defined as a carefully planned discussion designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment. (HUTH; FüSSL; RISER, 2014). Both the group situation and the guidance by a moderator stimulate discussions on the issues of interest and help to get insights into opinions and the motives that underlie attitudes and behavior. As a result of the social interaction of the group and its synergy, the data obtained is richer and deeper than those stemming from one-to-one interviews.

What distinguishes group interviews from one-to-one in-depth interviews is their capacity to capture the dynamics of group interaction and to exploit this in attempting to understand a topic. (HYDE et al., 2005). The execution of focus group depends on two professionals: moderator and facilitator. The first one is responsible for encouraging the participants to answer questions and interact during the meeting. The second one is responsible for observe the meeting, manage the overall time and take notes. After the meeting these professionals should discuss the impressions and perceptions about the meeting.
5.1.1 Participants

Participants were selected based on their experience with competitive intelligence. Basically, three types of professionals were selected: strategy specialists, engineers and health, safety & sustainability professionals. Following guidelines proposed by Hoyland, Haugen e Thomassen (2014), we recruited a purposeful sample from personnel who had more than five years of experience with strategic analysis. The participants were selected through consultation with the managers of the section and the leaders of each profession.

As suggested by Huth, Füssel e Risser (2014), participants of focus group discussions were selected on the basis of purposiveness for a given topic rather than representativeness. They were able to provide relevant information on the topic of interest and provide insights that were personally important to them.

5.1.2 Written Invitation

A written invitation was sent to selected professionals, presenting the topic, local, date and time of focus group. An example of invitation is presented both in English and in Portuguese at Appendix B. Few days before the focus group, the invited professionals were contacted by phone, aiming to reinforce the meeting idea. Invitation letters were sent to 11 (eleven) professionals and 07 (seven) of them attended the meeting.

According to Hoyland, Haugen e Thomassen (2014), the size of our focus group adheres to the common norm of four (minimum) to twelve (maximum) participants per group. For these authors, a smaller number of participants has greater potential, since pauses are more likely to occur, which allows everyone to share their insights. Picture 10 shows the classification of participants regarding their function in the company. Picture 11 shows the classification of participants by age, professional experience and years working at Vale. Lastly, picture 12 shows the classification of participants regarding their specialties.

5.2 Focus Group Execution

In order to achieve the focus group purpose, we scheduled a meeting two hours long shortly after office hours. This way, the participants could expose their thoughts without haste. As suggested by (HYDE et al., 2005), in this interview, a moderator facilitated the interview, while an assistant moderator was primarily responsible for the audio-recording and noted the order in which participants spoke.

Few days before the meeting, we generated the focus group script and other forms to support the overall process. The focus group script was generated with the purpose to guide the facilitator and the moderator during the meeting. It was designed to provide triggers for discussion rather than be used as a prescriptive structure for the interview.
Figure 10 – Focus Group - Participants Roles  
Source: Author

(HYDE et al., 2005). The script focuses on the main theme we identified at the start of this section with a broad opening question under this theme facilitative of group interactions.

Moderator and assistant moderator were given additional information specific to focus group interviewing, including written guidelines. (HYDE et al., 2005). This included details about the role of the moderator in establishing ground rules, making the environment a safe place for group members, making the group feel at ease, the importance of a non-judgmental stance, and the need to moderate dominant group members. Following guidelines proposed by Huth, Füssl e Risser (2014), the moderator reminded participants that there were no right or wrong statements and that any information was welcome, including opposed opinions among the participants.

The questions used in the focus group were carefully chosen aiming at achieving the general purpose of focus group. Firstly, an opening question was selected intending to create an open and comfortable environment to the participants. Introductory questions were used to bring the participants closer to main topic of focus group. We selected 02 (two) questions with this purpose. A transition question was used to stimulate participants’ memories aiming to focus their attention to the key questions which were asked next. We defined 06 (six) key questions whose answers could answer directly the focus group main question. Finally, we used 03 (three) ending questions to capture any thoughts not mentioned before and to close the questions section.
As suggested by Høyland, Haugen e Thomassen (2014), the ending of focus group session occurred when all discussion triggers had been pursued and the participants had no further thoughts on the particular themes. Focus group script used as guideline for this stage is presented in Appendix A.

5.2.1 Data Collection

The group consisted of different profiles of professionals (age, expertise, etc.) and discussion was held until no new ideas were raised any more. As suggested by Huth, Füssl e Risser (2014), a list of relevant issues was gathered that can be considered as important to a wider group.

While the meeting was audio recorded, the facilitator had to have a space for note taking. So, a note taking form was generated based on script so that facilitator could take notes during the focus group meeting. Another form generated was the consent form. Through this form the participants declared their consent to record the audio of the
meeting. The consent form is presented in the Appendix C.

A topic guide was then followed, in order to evoke the research themes during the discussion. A topic guide was specifically created for this study, approaching the key areas of interest. Following the suggestions of Huth, Füssl e Risser (2014), the topic guide was designed to elicit conversation among the participants on the determinants strategic management issues. The focus group session was audiotaped and a member of the research team took detailed notes of the discussions.

5.2.2 Ethics

We informed all potential participants of the study prior to the focus group session. At the start of focus group session, we also informed the participants of the study and gave them a written information form, including information on the aim of the study, anonymity issues, and a field for signing informed consent. We only proceeded with the focus group session once all participants had agreed to the study.

Following Huth, Füssl e Risser (2014) suggestion, the moderator assured that the data would be treated confidentially and asked the participants to sign an informed consent form. The participants were informed that the data would be used by researchers. Appendix D presents the form given to participants detailing the commitments to them.

5.2.3 Interview

The initial interview guide was revised during the data gathering phase to reflect emerging theoretical insights. The notes taken during the focus group discussions were used for the analysis of the collected data. Whenever clarifications of these notes were
needed, the audiotapes were consulted. Patterns were identified through careful reading and listening of the data.

5.2.4 Limitations

According to Høyland, Haugen e Thomassen (2014), a limitation of this study concerns the nature of qualitative findings, which is often unique to the particular context in which the study is being conducted. However, the fact that we were able to compare our findings more or less directly to existing research, conducted in different settings, suggests that our findings should have a certain transferability potential to other strategic management settings.

5.3 Focus Group Results

Focus group studies provide data that is of crucial importance for the design of subsequent quantitative studies, which will be able to focus on truly relevant issues when analysing the topic with a larger population. (HUTH; FüSSL; RISSER, 2014).

5.3.1 Datasources

During the focus group, participants mentioned the existence of specific and generic datasources. "Specific Datasources" may vary depending on the country, business or competitor being analyzed. "Generic Datasources" can be used in multiple analysis, independently of country, business or competitor. Participants also separated datasources applicable to multiple segments from those applicable uniquely to their segment. Picture 13 presents a matrix organizing datasources by category and applicability. This picture indicates that results of this focus groups can be used to formulate and design valid large-scale studies that will close relevant knowledge gaps. Table 9 presents a list of Generic Datasources cited by participants as references for obtaining information regarding external environments, independently of search in course.

5.3.2 Information

Picture 14 presents the cloud of all information cited by participants during focus group as typically used for the analysis of external environments, aiming to support the decision-making process. Participants also listed information classified as crucial to perform an external environment analysis, as can be seen next:

- General Risk
- Geological Fertility
Chapter 5. Competency Questions

- Licensing Requirements
- General Infrastructure
- Human Development Index
- Operational Costs

5.3.3 Difficulties

Participants reported difficulties to obtain information from countries which adopt non-latin alphabets, such as: Cyrillic and Arabic Alphabets, Ideograms, among others. Participants also mentioned difficulties to obtain data from competitors, such as: world presence, products specification, marketing analysis, etc. Another trouble reported by participants is when they try to obtain historical demand for products. Finally, trying to obtain a geo-referenced list of sacred places inside the country was mentioned as another barrier by participants.
Chapter 5. Competency Questions

Figure 13 – Focus Group - Matrix of Datasources
Source: Author
<table>
<thead>
<tr>
<th>DATASOURCE</th>
<th>WEBSITE ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOOGLE</td>
<td><a href="https://www.google.com.br/">https://www.google.com.br/</a></td>
</tr>
<tr>
<td>DOING BUSINESS PROJECT</td>
<td><a href="http://portugues.doingbusiness.org/">http://portugues.doingbusiness.org/</a></td>
</tr>
<tr>
<td>UNITED NATIONS</td>
<td><a href="http://www.un.org/">http://www.un.org/</a></td>
</tr>
<tr>
<td>WORLD HEALTH ORGANIZATION</td>
<td><a href="http://www.who.int/en/">http://www.who.int/en/</a></td>
</tr>
<tr>
<td>WORLD TRADE ORGANISATION</td>
<td><a href="http://www.wto.org/">http://www.wto.org/</a></td>
</tr>
<tr>
<td>CENTRAL INTELLIGENCE AGENCY</td>
<td><a href="https://www.cia.gov/index.html">https://www.cia.gov/index.html</a></td>
</tr>
<tr>
<td>DELLOITE</td>
<td><a href="http://www.deloitte.com/">http://www.deloitte.com/</a></td>
</tr>
<tr>
<td>LONDON METAL EXCHANGE</td>
<td><a href="http://www.lme.com/">http://www.lme.com/</a></td>
</tr>
<tr>
<td>EDGARD - ELECTRONIC DATA GATHERING, ANALYSIS, AND RETRIEVAL SYSTEM</td>
<td><a href="https://www.sec.gov/edgar/aboutedgar.htm">https://www.sec.gov/edgar/aboutedgar.htm</a></td>
</tr>
<tr>
<td>SEDAR - SYSTEM FOR ELECTRONIC DOCUMENT ANALYSIS AND RETRIEVAL</td>
<td><a href="http://www.sedar.com/homepage_en.htm">http://www.sedar.com/homepage_en.htm</a></td>
</tr>
<tr>
<td>REUTERS</td>
<td><a href="http://br.reuters.com/">http://br.reuters.com/</a></td>
</tr>
<tr>
<td>ABNT</td>
<td><a href="http://www.abnt.org.br/">http://www.abnt.org.br/</a></td>
</tr>
<tr>
<td>CRU</td>
<td><a href="http://www.crugroup.com/">http://www.crugroup.com/</a></td>
</tr>
<tr>
<td>SNL</td>
<td><a href="http://www.snl.com/">http://www.snl.com/</a></td>
</tr>
<tr>
<td>BROOKHUNT</td>
<td><a href="http://www.brookhunt.com">http://www.brookhunt.com</a></td>
</tr>
<tr>
<td>SCIENCE DIRECT</td>
<td><a href="http://www.sciencedirect.com/">http://www.sciencedirect.com/</a></td>
</tr>
<tr>
<td>SCOPUS</td>
<td><a href="http://www.scopus.com/home.url">http://www.scopus.com/home.url</a></td>
</tr>
<tr>
<td>U.S. CENSUS BUREAU</td>
<td><a href="http://www.census.gov/">http://www.census.gov/</a></td>
</tr>
<tr>
<td>U.S. GEOLOGICAL SURVEY</td>
<td><a href="http://www.usgs.gov/">http://www.usgs.gov/</a></td>
</tr>
<tr>
<td>STEEL BUSINESS BRIEFING</td>
<td><a href="https://www.steelbb.com/pt/?PageID=1">https://www.steelbb.com/pt/?PageID=1</a></td>
</tr>
<tr>
<td>NASA</td>
<td><a href="http://www.nasa.gov/">http://www.nasa.gov/</a></td>
</tr>
<tr>
<td>ESRI</td>
<td><a href="http://www.esri.com/">http://www.esri.com/</a></td>
</tr>
<tr>
<td>GREENPEACE</td>
<td><a href="http://www.greenpeace.org/brasil/pt/">http://www.greenpeace.org/brasil/pt/</a></td>
</tr>
<tr>
<td>FRASER INSTITUTE</td>
<td><a href="https://www.fraserinstitute.org/">https://www.fraserinstitute.org/</a></td>
</tr>
<tr>
<td>WIKIPEDIA</td>
<td><a href="http://www.wikipedia.org/">http://www.wikipedia.org/</a></td>
</tr>
</tbody>
</table>

Table 9 – Focus Group - Generic Datasources
<table>
<thead>
<tr>
<th>Data Category</th>
<th>Key Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic Data</td>
<td>- AIDS INDEX&lt;br&gt;- CORRUPTION INDEX&lt;br&gt;- COUNTRY GOVERNANCE REGIME&lt;br&gt;- COUNTRY RELIGION&lt;br&gt;- COUNTRY ECONOMIC SEGMENTS&lt;br&gt;- COUNTRY SIZE&lt;br&gt;- ETHNICITY&lt;br&gt;- FORECASTED POPULATION&lt;br&gt;- HDI</td>
</tr>
<tr>
<td>Macro Economic Data</td>
<td>- COUNTRY RISK LEVEL&lt;br&gt;- COUNTRY'S MINERAL PRODUCTION&lt;br&gt;- EDUCATION EXPENDITURE&lt;br&gt;- EXCHANGE&lt;br&gt;- FORECASTED DEMANDA&lt;br&gt;- FORECASTED GDP&lt;br&gt;- FORECASTED PRICES&lt;br&gt;- FORECASTED SUPPLY&lt;br&gt;- GDP&lt;br&gt;- HEALTH EXPENDITURE&lt;br&gt;- INDUSTRIALIZATION LEVEL&lt;br&gt;- INITIAL COSTS&lt;br&gt;- PER CAPTA GDP&lt;br&gt;- RATINGS</td>
</tr>
<tr>
<td>Data About Competitors</td>
<td>- COMPETITORS FINANCIAL PERFORMANCE&lt;br&gt;- COMPETITORS GEO-REFERENCED LOCATION&lt;br&gt;- COMPETITORS TOTAL INVESTMENT&lt;br&gt;- COMPETITORS INVESTMENT PLACES&lt;br&gt;- COMPETITORS PRODUCTS SPECIFICATIONS&lt;br&gt;- TECHNOLOGY USED BY COMPETITORS</td>
</tr>
<tr>
<td>Stocks and Shares</td>
<td>- COUNTRY'S MAIN ECONOMIC GROUPS&lt;br&gt;- IMPORT AND EXPORT TRADES</td>
</tr>
<tr>
<td>Internal Data</td>
<td>- CAPEX AND OPEX&lt;br&gt;- CUSTOMERS DISTANCE&lt;br&gt;- SUPPLIERS DISTANCE</td>
</tr>
<tr>
<td>Geo-Referenced Data</td>
<td>- AIRPORTS&lt;br&gt;- HOTELS&lt;br&gt;- MINERAL RESOURCES&lt;br&gt;- NEIGHBOURS COUNTRIES&lt;br&gt;- PORTS&lt;br&gt;- RAILWAYS&lt;br&gt;- SACRED PLACES GEO-REFERENCED LOCATION&lt;br&gt;- TOPOGRAFY</td>
</tr>
<tr>
<td>Legal Issues</td>
<td>- ENVIRONMENTAL LAW&lt;br&gt;- HEALTH LAW&lt;br&gt;- LEGAL OBBLIGATIONS&lt;br&gt;- MINERAL LEGISLATIONS&lt;br&gt;- TAXES&lt;br&gt;- TECHNOLOGY OR PROCESS PATENTS</td>
</tr>
<tr>
<td>Infrastructure Data</td>
<td>- ARABLE LAND&lt;br&gt;- EARTHQUAKE RISK&lt;br&gt;- ENERGY DISTRIBUTION GRID&lt;br&gt;- ENERGY MATRIX&lt;br&gt;- FLOODS RISK&lt;br&gt;- HURRICANES RISK&lt;br&gt;- POPULATION WITH ACCESS TO WATER&lt;br&gt;- TSUNAMIS RISK&lt;br&gt;- VOLCANES RISK&lt;br&gt;- WAR RISK</td>
</tr>
</tbody>
</table>

Figure 14 – Focus Group - Relevant Information
Source: Author
6 A Framework for Linked Enterprise Data

This chapter describes the steps involved in the construction of the prototype used to attend the specific objective OBJ3 - 'Build the prototype for integrated data visualization'. The chapter is organized as following: firstly, the conceptual framework of the prototype is presented; secondly, the dataset built to attend the requirements of the project is detailed; finally, the visualization application is itemized.

6.1 Conceptual Framework

The conceptual framework is presented in figure 15 which intends to highlight the main elements without further specification details. Table 10 describes components and functions according to objectives, theoretical background and methodologies of this Master’s Thesis.
<table>
<thead>
<tr>
<th>Layer</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>This layer refers to data sources available for integration. Only structured data will be considered</td>
</tr>
<tr>
<td>Wrapper</td>
<td>This layer provides access to specific types of data, distributes queries over data sources and aggregates results. It encompasses adaptors layer</td>
</tr>
<tr>
<td>Integration</td>
<td>This layer provides all the mapping functionality (semantics) and an API that is independent of a specific data source and query language. It encompasses the query layer</td>
</tr>
<tr>
<td>Presentation</td>
<td>This layer enables an interface to access data. The competency questions will be validated through this interface</td>
</tr>
</tbody>
</table>

Table 10 – Functions of Conceptual Framework

6.2 Implementation

6.2.1 Data Layer

The data layer was implemented in MS-SQL Server 2012. This implementation enabled creation of the dataset which was used later to validate competency questions.

Categories were created based on Nettleton (2014). In order to satisfy all kinds of variables identified during the focus group, we added four categories, including one specifically used to identify entities. Table 11 presents categories used in data layer.

The dataset was populated with information gathered from various sources, as identified on focus group, described on chapter 5. Table 12 presents all sources used in dataset with their web address.

Table 12 – Implementation - Data Layer Sources

<table>
<thead>
<tr>
<th>Source Name</th>
<th>Source Reference</th>
</tr>
</thead>
</table>

Continued on next page
Table 12 – Continued from previous page

<table>
<thead>
<tr>
<th>Source Name</th>
<th>Source Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>OurAirports</td>
<td></td>
</tr>
<tr>
<td>KPMG</td>
<td></td>
</tr>
<tr>
<td>Transparency International</td>
<td></td>
</tr>
<tr>
<td>Federal Reserve</td>
<td></td>
</tr>
</tbody>
</table>

From all information identified during focus group, only those available on web
<table>
<thead>
<tr>
<th>Category Name</th>
<th>Category Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A - Demographic data</td>
<td>Demographic data includes general statistics about the population and characterize different groups and subgroups.</td>
</tr>
<tr>
<td>B - Macro-Economic data</td>
<td>The economic data about a given business sector includes the increase or decrease in total sales, sales for given products and services, and sales in a given sector with respect to previous periods (quarters, years).</td>
</tr>
<tr>
<td>C - Data about competitors</td>
<td>In order to compete in the marketplace, a good-quality database must be maintained, with information about the characteristics of competitors, their products and services, and their summarized business data</td>
</tr>
<tr>
<td>D - Stocks and shares</td>
<td>If a company is listed in a stock exchange, if it is interested in following the share prices of companies in its sector, or if it is interested in investing, then the stock exchange is a key source of data for that company</td>
</tr>
<tr>
<td>E - Internal Data</td>
<td>Data about a business’s products, services, and customers, together with feedback on business activities from surveys, questionnaires, and loyalty and customer cards</td>
</tr>
<tr>
<td>F - Spatial Data</td>
<td>Data or information that identifies the geographic location of features and boundaries on Earth, such as natural or constructed features, oceans, and more</td>
</tr>
<tr>
<td>G - Legal Issues</td>
<td>Issues about lawsuits, response to new legislations or laws, new regulations, and fines</td>
</tr>
<tr>
<td>H - Infrastructure Data</td>
<td>Infrastructure helps determine the success of manufacturing and agricultural activities. Investments in water, sanitation, energy, housing, and transport also improve lives and help reduce poverty. And new information and communication technologies promote growth, improve delivery of health and other services, expand the reach of education, and support social and cultural advances</td>
</tr>
<tr>
<td>System Data</td>
<td>Category used to identify entities in the system</td>
</tr>
</tbody>
</table>

Table 11 – Implementation - Data Layer Categories

in open data format were selected to compound this dataset. This categorization was performed following guidelines proposed by Berners-Lee (2009). Table 13 presents the final list of variables created to support this study along with categories and description.

Table 13 – Implementation - Data Layer Variables

<table>
<thead>
<tr>
<th>Variable Friendly Name</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airports</td>
<td>Recordset containing airports from country</td>
</tr>
<tr>
<td>Annual GDP Growth</td>
<td>Annual GDP Growth</td>
</tr>
<tr>
<td>Annual Population Growth</td>
<td>Annual population growth rate (%)</td>
</tr>
<tr>
<td>Arable Land</td>
<td>Arable land</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Variable Friendly Name</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corruption Index</td>
<td>Corruption Index - Relates to the degree to which corruption is perceived to exist among public officials and politicians by business people and country analysts.</td>
</tr>
<tr>
<td>Country Capital Name</td>
<td>Name of Capital of Country</td>
</tr>
<tr>
<td>Country Code - ISO3</td>
<td>International Country code - 3 digits</td>
</tr>
<tr>
<td>Country Common Name</td>
<td>Common name of Country</td>
</tr>
<tr>
<td>Country Currency Code</td>
<td>International Currency code</td>
</tr>
<tr>
<td>Country Currency Name</td>
<td>Name of currency</td>
</tr>
<tr>
<td>Country Data Completeness</td>
<td>Indicates the completeness of data</td>
</tr>
<tr>
<td>Country Formal Name</td>
<td>Formal name of Country</td>
</tr>
<tr>
<td>Education Expenditure</td>
<td>Public expenditure on education as % of GDP</td>
</tr>
<tr>
<td>Employment Index</td>
<td>Employment-to-population ratio</td>
</tr>
<tr>
<td>Exchange Rate</td>
<td>Exchange rate to US$</td>
</tr>
<tr>
<td>Export Trade</td>
<td>In the case of the general trade system, export flows come from: (1) The free circulation area, premises for inward processing or industrial free zones; (2) Premises for customs warehousing or commercial free zones. There are three types of exports: (a) Domestic goods originating in the free circulation area or in industrial free zones; (b) Domestic goods comprised of compensating products after inward processing; (c) Foreign goods in the same state as previously imported. In the case of the special trade system under the relaxed definition, the export flows come only from the free circulation area, premises for inward processing or industrial free zones. There are three types of exports: (a) Domestic goods originating in the free circulation area or industrial free zones; (b) Domestic goods comprised of compensating products after inward processing; (c) Foreign goods in the same state as previously imported. There are two possible destinations: (1) The rest of the world; (2) Premises for customs warehousing or commercial free zones.</td>
</tr>
<tr>
<td>Female Literacy Rate</td>
<td>Adult literacy rate - Female 15+ yr</td>
</tr>
<tr>
<td>Forest Proportion</td>
<td>Proportion of land area covered by forest</td>
</tr>
</tbody>
</table>
Table 13 – Continued from previous page

<table>
<thead>
<tr>
<th>Variable Friendly Name</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>Gross domestic product is an aggregate measure of production equal to the sum of the gross values added of all resident institutional units engaged in production (plus any taxes, and minus any subsidies, on products not included in the value of their outputs). The sum of the final uses of goods and services (all uses except intermediate consumption) measured in purchasers’ prices, less the value of imports of goods and services, or the sum of primary incomes distributed by resident producer units.</td>
</tr>
<tr>
<td>Geothermal Electricity</td>
<td>Net installed capacity of electric power plants, geothermal (public &amp; self-producer)</td>
</tr>
<tr>
<td>HDI</td>
<td>Human Development Index (HDI)</td>
</tr>
<tr>
<td>Health Expenditure</td>
<td>Central government expenditure allocated to health</td>
</tr>
<tr>
<td>HIV Rate</td>
<td>Adult HIV/AIDS prevalence rate - 15 to 49 Yr</td>
</tr>
<tr>
<td>Hospital Beds Rate</td>
<td>Hospital beds (per 10 000 population)</td>
</tr>
<tr>
<td>Hydro Electricity</td>
<td>Net installed capacity of electric power plants, hydro (public &amp; self-producer)</td>
</tr>
</tbody>
</table>

*Continued on next page*
Table 13 – Continued from previous page

<table>
<thead>
<tr>
<th>Variable Friendly Name</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Import Trade</td>
<td>In the case of the general trade system, import flows come from the rest of the world or from customs transit, i.e., goods redirected from customs transit to remain in the economic territory. There are three types of imports: (a) Foreign goods (other than compensating products after outward processing); (b) Foreign goods comprised of compensating products after outward processing; (c) Domestic goods in the same state as previously exported (so-called re-imports). These imports are brought into: (1) The free circulation area, premises for inward processing or industrial free zones; (2) Premises for customs warehousing or commercial free zones. In the case of the special trade system, under the relaxed definition, the import flows come from: (1) The rest of the world or from customs transit; (2) Premises for customs warehousing or commercial free zones. There are three types of imports: (a) Foreign goods (other than compensating products after outward processing); (b) Foreign goods comprised of compensating products after outward processing; (c) Domestic goods in the same state as previously exported. The goods are imported into free circulation.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Internal Score</th>
<th>Country’s score considering internal parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet Users</td>
<td>Percentage of individuals using the Internet</td>
</tr>
<tr>
<td>Life Expectancy</td>
<td>Life expectancy at birth</td>
</tr>
<tr>
<td>Male Literacy Rate</td>
<td>Adult literacy rate - Male 15+ yr</td>
</tr>
<tr>
<td>Marginal Tax Rates</td>
<td>Marginal tax rates</td>
</tr>
<tr>
<td>Minerals</td>
<td>Recordset containing minerals from country</td>
</tr>
<tr>
<td>Nuclear Electricity</td>
<td>Net installed capacity of electric power plants, nuclear (public &amp; self-producer)</td>
</tr>
<tr>
<td>Per Capta GDP</td>
<td>GDP per head calculated as the aggregate of production (GDP) divided by the population size</td>
</tr>
<tr>
<td>Population Projection</td>
<td>Population projections by 2030</td>
</tr>
<tr>
<td>Population Supplied by Water</td>
<td>Total population supplied by water supply industry</td>
</tr>
</tbody>
</table>

Continued on next page
Table 13 – Continued from previous page

<table>
<thead>
<tr>
<th>Variable Friendly Name</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reexport Trade</td>
<td>Re-exports are exports of foreign goods in the same state as previously imported; they are to be included in the country exports. They are also recommended to be recorded separately for analytical purposes, which may require the use of supplementary sources of information in order to determine the origin of re-exports, i.e., to determine that the goods in question are indeed re-exports rather than the export of goods that have acquired domestic origin through processing.</td>
</tr>
<tr>
<td>Reimport Trade</td>
<td>Re-imports are goods imported in the same state as previously exported; they are to be included in the country imports. They are also recommended to be recorded separately for analytical purposes, which may require the use of supplementary sources of information in order to determine the origin of re-imports, i.e., to determine that the goods in question are indeed re-imports rather than the import of goods that have acquired foreign origin through processing.</td>
</tr>
<tr>
<td>Solar Electricity</td>
<td>Net installed capacity of electric power plants, solar (public &amp; self-producer)</td>
</tr>
<tr>
<td>Thermal Electricity</td>
<td>Net installed capacity of electric power plants, thermal (public &amp; self-producer)</td>
</tr>
<tr>
<td>Time to Start a Business</td>
<td>Time required to start a business (days)</td>
</tr>
<tr>
<td>Total Electricity</td>
<td>Total net installed capacity of electric power plants, public &amp; self-producer</td>
</tr>
<tr>
<td>Total Population</td>
<td>Total Population (in thousands)</td>
</tr>
<tr>
<td>Wave Electricity</td>
<td>Net installed capacity of electric power plants, wave (public &amp; self-producer)</td>
</tr>
<tr>
<td>Wind Electricity</td>
<td>Net installed capacity of electric power plants, wind (public &amp; self-producer)</td>
</tr>
<tr>
<td>Years of Schooling</td>
<td>Mean years of schooling</td>
</tr>
</tbody>
</table>

Next section describes wrapper layer.
6.2.2 Wrapper Layer

Wrapper layer was implemented through D2RQ Platform, which is a system for accessing relational databases as virtual, read-only RDF graphs. Its architecture is represented in Figure 16. According to Cyganiak (2012), D2RQ Platform consists of:

- the **D2RQ Mapping Language**, a declarative mapping language for describing the relation between an ontology and an relational data model.
- the **D2RQ Engine**, a plug-in for the Jena Semantic Web toolkit, which uses the mappings to rewrite Jena API calls to SQL queries against the database and passes query results up to the higher layers of the frameworks.
- **D2R Server**, an HTTP server that provides a Linked Data view, a HTML view for debugging and a SPARQL Protocol endpoint over the database.

D2RQ Platform has D2R Server, which is a tool for publishing the content of relational databases on the Semantic Web. (CYGANIAK, 2012). After installing and configuring D2RQ Platform, the dataset presented in Data Layer section was made available through SPARQL endpoint. Figure 16 illustrates this feature.

Next section describes integration layer.

6.2.3 Integration Layer

In order to meet this study objectives, a Semantic Repository was built and a portion of RDF file is illustrated in figure 18. The main purpose of the Semantic Repository was to create an API independent of data source.

Integration layer allows variables to be categorized and related to a datasource, a measuring unit and to one or more attributes. The RDF model also allows the creation of relationship between entities and the association with variables. There are three steps to implement the data model: 1) create entities; 2) create variables and attributes; 3) create relationships between entities and variables.

Initially, a class of entities should be created in the Entity table. Classes should be identified through the value of ClassOfEntities field. In order to achieve this study’s purposes, a single class was created representing countries. After that, instances should be created, representing each country individually. All instances should be related to the country class through RootEntityDBCode field.

Secondly, variables should be created and related to categories, sources and units. An special category of variables, used uniquely to identify entities, should be created. Variables such as name, international code, among others, should use this special category.
Next, further categories, used for describing entities, should be created. Variables such as GDP, HDI, among others, should use these normal categories. Sources of variables should represent the external references of those variables. Units of variables should represent variables engineering units.

Thirdly, variables should be related to attributes. Each variable could have one or more attributes. Considering, as an example, a variable used to identify the name of an entity, the database schema should be populated with a variable representing the name; this variable should be associated to an attribute called value, which would receive the entity name itself.

Finally, each instance of country should be related to one or many variables and attributes, in a process called Assessment. At the end of this process, there should be relationships between instances of entities and instances of variables. Although this study is limited to the examination of country data, the database schema was designed to allow the creation of classes to represent regions, cities, neighborhood, etc. Database schema
also allows creation of filiation of entities.

Next section describes presentation layer.
6.2.4 Presentation Layer

In order to meet this study’s objectives, a web application was built so that users could validate data from Semantic Repository using their own criteria. The web application was named **DB4Trading** and was made available on **<www.DB4Trading.com>**. The main idea behind the application name was that the tool should be used as support for trading between companies.

**DB4Trading** web site was designed to be dynamically generated, based on parameters defined by administrators and exchanged through special files. Picture 19 presents the main page of website, highlighting the main parts of site. Table 14 describes each part of the site, considering the identifiers.

<table>
<thead>
<tr>
<th>ID</th>
<th>Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Title</td>
<td>Reserved for exhibition of website title</td>
</tr>
<tr>
<td>2</td>
<td>Menus</td>
<td>Reserved for navigation menus</td>
</tr>
<tr>
<td>3</td>
<td>Categories</td>
<td>Reserved for categories. From this area, users should define the relevant criteria to compare multiple countries</td>
</tr>
<tr>
<td>4</td>
<td>Map</td>
<td>Reserved for map exhibition. Countries would change their colors based on criteria selected by users.</td>
</tr>
<tr>
<td>5</td>
<td>Footnote</td>
<td>Reserved for footnote</td>
</tr>
</tbody>
</table>

Table 14 – Implementation - Presentation Layer Website Parts

![Figure 19 – Implementation - Presentation Layer Main Page](source: Author)
6.2.4.1 Site Structure

This section details the structure of DB4Trading web site. The site was developed using HTML, ASP and JavaScript. As the web site was designed to be dynamically generated, we used four files to exchange information about menus, categories, variables and data: MENUS.ASP, CATEGORIES.ASP, VARIABLES.ASP and DATASET.JS. All files were exported from Semantic Repository. Picture 20 illustrates interactions between the site and the files used to generate it, while Table 15 describes the purpose of each file.

<table>
<thead>
<tr>
<th>File</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>MENUS.ASP</td>
<td>Used to create and relate menus to HTML pages.</td>
</tr>
<tr>
<td>CATEGORIES.ASP</td>
<td>Used to create group of variables in categories area.</td>
</tr>
<tr>
<td>VARIABLES.ASP</td>
<td>Used to create variables and relate them to categories.</td>
</tr>
<tr>
<td>DATASET.JS</td>
<td>Used to exchange country data with application.</td>
</tr>
</tbody>
</table>

Table 15 – Implementation - Presentation Layer Files and Purpose

6.2.4.2 Heat Map

From the weight defined for each one of the variables in the categories area, the application was designed to identify countries whose information were more adherent. In order to identify countries, we used a specific Google API capable to highlight the country polygon. DB4Trading was designed to calculate the color of countries, marking
in red those classified as more relevant and leaving in blank those classified as less relevant. The Heat Map formula used to calculate the color of the country is presented in Picture 21. Table 16 explains each formula parameter.

\[
Score = \left[ \text{Index} \cdot \left( \frac{\sum_{i=1}^{k} X_i \cdot n_i}{\sum_{i=1}^{k} n_i} \right) \right]
\]

**Figure 21 – Implementation - Presentation Layer Heat Map Formula**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index</td>
<td>Index of completeness of data gathered of the country. It is used to balance the final score. The more information is obtained from a country, the more higher is this index.</td>
</tr>
<tr>
<td>i</td>
<td>Index of the analyzed variable. Used to implement the weighted average formula.</td>
</tr>
<tr>
<td>k</td>
<td>Total of variables considered in the calculation.</td>
</tr>
<tr>
<td>X</td>
<td>Normalized value of variables, varying from 0 to 1. The closer the value is of 1, the more adherent is the country.</td>
</tr>
<tr>
<td>n</td>
<td>Weights of variable, varying from 0 to 1. This value is set by user through the use of sliders in the categories area.</td>
</tr>
</tbody>
</table>

**Table 16 – Implementation - Presentation Layer Heat Map Formula Parameters**
7 Framework Validation

To validate the conceptual framework we decided to run specific queries against dataset and also create specific scenarios in visualization application. This was necessary because web application shows data from multiple countries based on heat map calculation while dataset is able to present data for each country individualized.

7.1 Semantic Repository

In this section we describe the queries runned against dataset and the results obtained. To validate the framework, the following competency questions were considered:

1. What are the countries with the highest human development index (HDI)?
2. What are the countries with the highest numbers of hospital beds?
3. What are the countries with the highest export trade volume?
4. What are the countries with the lowest tax burdens?
5. What are the countries with the lowest time for starting a business?

Table 17 presents SPARQL sentences created to answer the validation competency questions. Picture 22 shows the results obtained.

Table 17 – Framework Validation - Semantic Repository Queries

<table>
<thead>
<tr>
<th>Question #1</th>
<th>SPARQL Sentence</th>
</tr>
</thead>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Question</th>
<th>SPARQL Sentence</th>
</tr>
</thead>
</table>

Continued on next page
### 7.2 Visualization Application

In this section we describe the scenarios used to validate web application and the results obtained. We created the following scenarios:

1. What are the countries with the highest human development index (HDI) AND with the highest population AND with the highest gross domestic product (GDP)?

2. What are the countries with the highest export trade volume AND with the highest volume of Nuclear Power?

![Figure 22 – Framework Validation - Semantic Repository Queries Results](source: Author)
Firstly, figures 23 and 24 present the DB4Trading configuration for answer the first and second scenarios respectively. Lastly, figures 25 and 26 present results obtained for the first and second scenarios respectively.

Figure 23 – Framework Validation - DB4Trading Scenario 1 Parameters
Source: Author
Figure 24 – Framework Validation - DB4Trading Scenario 2 Parameters
Source: Author
Figure 25 – Framework Validation - DB4Trading Scenario 1 Results
Source: Author

Figure 26 – Framework Validation - DB4Trading Scenario 2 Results
Source: Author
8 Conclusion

Based on the results of this research, we concluded that it is possible to use Semantic Web technologies to gather and distribute information for external environment analysis. Assuming that information internally generated into an organization can be fully integrated to external information through open data patterns - even if internal information is not made available to external publics - this study shows that 96% of required information for external environment analysis can be gathered and distributed through Semantic Web technologies, corroborating Queyras e Quoniam (2006).

It is worth pointing out that about 38% of information obtained from external datasources in this research have no global reach. This way, depending on the analyzed country the percentage of available information can be lower. Governments and institutions should be encouraged to release their information using open data patterns in order to increase the percentual of information shared globally.

From the perspective of competitive intelligence professionals working on external environment analysis, we concluded that the results of this research can be used to formulate and design valid large-scale studies that will close relevant knowledge gaps. This research showed that most part of datasources and information used by competitive intelligence professionals are applicable to any segment.

This research indicated that competitive intelligence professionals have difficulties to obtain: a) data from countries which adopt non-latin alphabets; b) data from competitors, such as world presence, products specification, marketing analysis, etc; c) historical data about demand for products; d) geo-referenced list of sacred places inside countries.

Based on systematic literature review, we concluded that enterprises are already using Semantic Web technologies for support their operations, specially for internal datasources integration. However, enterprises still believe that Semantic Web technologies are complex and require high specialized teams.
References


References


MITJA, T. The knowledge management wheel. Management, Knowledge and Learning, 2011.


References


Appendix
<table>
<thead>
<tr>
<th>Question #1</th>
<th>Have you ever been involved in environment scan analysis?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question #2</td>
<td>When you think about environment scan analysis, what are the facts that come into your mind?</td>
</tr>
<tr>
<td>Question #3</td>
<td>Which sources have you ever used to obtain information for environment scan analysis?</td>
</tr>
<tr>
<td>Question #4</td>
<td>Imagine that you were asked to do a survey using data from all analyzes of the external environment in which you participated. Which kind of information do you believe would find in most of these analyzes?</td>
</tr>
<tr>
<td>Question #5</td>
<td>In your opinion, what are the most relevant information if you needed to analyze the external environment, considering the politics and government of a region?</td>
</tr>
<tr>
<td>Question #6</td>
<td>In your opinion, what are the most relevant information if you needed to analyze the external environment, considering the social and economic indicators of a region?</td>
</tr>
<tr>
<td>Question #7</td>
<td>In your opinion, what are the most relevant information if you needed to analyze the external environment, considering the infrastructure and energy of a region?</td>
</tr>
<tr>
<td>Question #8</td>
<td>In your opinion, what are the most relevant information if you needed to analyze the external environment, considering competitors and suppliers of a region?</td>
</tr>
<tr>
<td>Question #9</td>
<td>In your opinion, what are the most relevant information if you needed to analyze the external environment, considering the community and the local risks of a region?</td>
</tr>
<tr>
<td>Question #10</td>
<td>Imagine that you were asked to do a survey using data from all analyzes of environment in which you participated. Please, reflect on the information that could have affected decisions if they were present in analyzes that you performed.</td>
</tr>
<tr>
<td>Question #11</td>
<td>If your company CEO asked your opinion about crucial information for understanding the external environment, what would you answer?</td>
</tr>
<tr>
<td>Question #12</td>
<td>Anyone would like to contribute with anything else?</td>
</tr>
</tbody>
</table>

Table 18 – Appendix - Focus Group Questions
APPENDIX B – Focus Group Invitation

FUMEC University invites you to Participate in a Focus Group:

Tools for Supporting
Competitive Intelligence

Date & Time: 2014/05/29 from 05pm to 07pm
Place: Ed Paraúnas – Meeting Room 02 – 3o Floor

Figure 27 – Appendix - Focus Group Invitation (English Version)
Source: Author

A Universidade FUMEC convida para Participação em Grupo Focal:

Ferramentas de Suporte à
Inteligência Competitiva

Data e Hora: 29/05/2014 de 17h às 19h
Local: Ed Paraúnas – Sala de Reuniões 02 – 3o Andar

Figure 28 – Appendix - Focus Group Invitation (Portuguese Version)
Source: Author
CONSENT TO PARTICIPATE IN FOCUS GROUP

I declare that I understood the purpose of the discussion group and the nature of the questions asked. I further declare that I agree with the audio recording of this meeting.

I declare that I agree to report experiences and suggestions that can provide service improvements and/or existing resources to obtain data used for competitive intelligence, recognizing that this step is crucial for strategic planning.

Declare to be aware that my participation is voluntary. I understand that I am free to leave the group at any time. I understand that I have the right not to participate or provide feedback at certain times of the discussion.

Declare that I know that the experiences and thoughts reported by me at the meeting may be shared with others within the Master Research in Information Systems and Knowledge Management of FUMEC University.

__________________________________________
Participant Name

__________________________________________
Signature

__________________________________________
Date

Figure 29 – Appendix - Focus Group Consent Form (English Version)
Source: Author
CONSENTIMENTO DE PARTICIPAÇÃO EM GRUPO FOCAL

Declaro que compreendi o propósito do grupo de discussão e a natureza das questões formuladas. Declaro ainda que concordo com a gravação do áudio desta reunião.

Declaro que concordo em relatar experiências e sugestões que possam proporcionar melhorias nos serviços e/ou recursos existentes para obtenção de dados utilizados para inteligência competitiva, reconhecendo que esta etapa é crucial para o planejamento estratégico.

Declaro estar ciente de que minha participação é voluntária. Compreendo que sou livre para deixar o grupo a qualquer momento. Compreendo que tenho o direito de não participar ou opinar em determinados momentos da discussão.

Declaro estar ciente de que as experiências e pensamentos relatados por mim durante a reunião poderão ser compartilhados com outras pessoas no âmbito da Pesquisa de Mestrado em Sistemas de Informação e Gestão do Conhecimento da Universidade FUMEC.

____________________________
Nome do Participante

____________________________
Assinatura

____________________________
Data

Figure 30 – Appendix - Focus Group Consent Form (Portuguese Version)
Source: Author
APPENDIX D – Focus Group Commitments
PARTICIPATION IN FOCUS GROUP

Thanks for your participation in the focus group to collect information used in environmental analysis (scan analysis). This procedure integrates the Master Research in Information and Management Systems Knowledge FUMEC University.

The main objective of the Master Research is help to improve the decision-making process in the context of portfolio management, by proposing and validating a framework capable of combining information from internal and external environments of an organization. We hope that the knowledge generated in the focus group orient the process of obtaining external information considered relevant.

As a participant in the focus group, you will receive a printed copy of the Master thesis in which will be published the results and findings of this research comprehensively.

The experiences and thoughts reported during the meeting will be shared with people directly involved in the research. However, for purposes of publication, the information provided will always be presented in an aggregated manner so as to prevent identification of participants.

If you have questions, comments or suggestions about the research use the data below to contact the researchers.

Thank You!

Vitor Afonso Pinto
Email: vitor.afonso.pinto@gmail.com

Dr. Fernando Silva Parreiras
Email: fernando.parreiras@fumec.br
Agradecemos sua participação no Grupo Focal para levantamento de informações utilizadas para análise de ambiente (*scan analysis*). Este procedimento integra a Pesquisa de Mestrado em Sistemas de Informação e Gestão do Conhecimento da Universidade FUMEC.

O principal objetivo da Pesquisa de Mestrado é contribuir para a melhoria do processo de tomada de decisão, no contexto de gestão de portfólio, através da proposição e validação de um *framework* capaz de combinar informações dos ambientes internos e externos de uma organização. Esperamos que o conhecimento gerado no grupo focal oriente o processo de obtenção de informações externas consideradas relevantes.

Como participante do grupo focal, você receberá uma cópia impressa da dissertação de mestrado na qual estarão publicados os resultados e as descobertas desta pesquisa de forma abrangente.

As experiências e pensamentos relatados durante a reunião poderão ser compartilhados com pessoas envolvidas diretamente na pesquisa. Entretanto, para fins de publicações, as informações prestadas sempre serão apresentadas de forma agrupada, de modo a impedir a identificação dos participantes.

Caso você tenha dúvidas, críticas ou sugestões sobre a pesquisa utilize os dados abaixo para entrar em contato com os pesquisadores.

Muito obrigado!

---

**Vitor Afonso Pinto**
Email: vitor.afonso.pinto@gmail.com

**Dr. Fernando Silva Parreiras**
Email: fernando.parreiras@fumec.br

---

Figure 32 – Appendix - Focus Group Commitments (Portuguese Version)
Source: Author