Evaluating Cloud Services using Multicriteria Decision Analysis

Pedro Miguel Almeida Carvalho Costa

Thesis to obtain the Master of Science Degree in

Information Systems and Computer Engineering

Examination Committee

Chairperson:	Prof. Nuno João Neves Mamede
Supervisor:	Prof. Miguel Leitão Bignolas Mira da Silva
Co-Supervisor:	Prof. João Carlos da Cruz Lourenço
Member of the Committee:	Prof. André Ferreira Ferrão Couto e Vasconcelos

July 2013

Abstract

For many years, most organizations have made significant investments in software and hardware and have incurred in also substantial fixed costs to maintain proprietary datacenters. Nowadays, Cloud Computing is revolutionizing the Information Technology (IT) industry by enabling organizations to turn fixed costs into variable costs by buying a service instead of owning their own assets. The potential of Cloud services for cost reduction and other benefits has been capturing the attention of organizations. However, a difficult decision arises when an IT manager has to select a Cloud service because there are no established guidelines or any structured form to do it. When several, and often conflictive, criteria should be taken into account to compare different Cloud services, the decision is even more difficult. To address this problem, we propose a method that consists in a set of evaluation criteria to Cloud services, which includes 19 evaluation criteria divided in six groups, using MACBETH, which is an interactive Multiple Criteria Decision Analysis (MCDA) approach used to build a quantitative value model based on qualitative pairwise comparison judgments. To guide our research we used the Design Science Research Methodology (DSRM), which is often used to address IT and organizational problems. Our proposal demonstration was accomplished within a City Council in Portugal to evaluate and compare two Cloud services: Google Apps and Microsoft Office 356. In order to evaluate the results of these experiments we used interviews with practitioners, feedback from the scientific community, Moody and Shanks Quality Framework, and Österle principles.

Keywords: Cloud Computing, Cloud Services, Information Technology Services, MACBETH, Multiple Criteria Decision Analysis, and Service Quality.

Resumo

Durante anos muitas organizações têm investido na aquisição de software e hardware incorrendo em custos fixos para manter centros de dados próprios. Hoje em dia, a Cloud Computing está a revolucionar a indústria das Tecnologias de Informação (TI) fazendo com que as organizações possam transformar custos fixos em variáveis adquirindo serviços a terceiros ao invés de os manterem. O potencial dos serviços Cloud para reduzir custos e outros benefícios têm chamado à atenção às organizações. Porém, a decisão difícil surge quando um decisor tem que selecionar o serviço Cloud que pretende adquirir porque não existem guias nem forma estruturada para o fazer. Quando vários, não independentes, critérios são tomados em conta para comparar os diferentes serviços Cloud, esta decisão torna-se ainda mais difícil. Para reduzir este problema, propomos um método que consiste num conjunto de critérios para avaliar serviços Cloud, composto por 19 critérios divididos por seis grupos, usando o MACBETH, que é uma abordagem interactiva de análise de decisão usando múltiplos critérios (ADMC) usada para construir um modelo de valor quantitativo baseado em julgamentos qualitativos com comparações de elementos. Neste trabalho usamos a metodologia *Design Science Research* (DSR), que é usada para endereçar problemas de TI e organizacionais. A demonstração da proposta foi realizada numa Câmara Municipal em Portugal onde foram avaliados e comparados dois serviços de Cloud: Google Apps e Microsoft Office 365. Para fazer a avaliação dos resultados usámos entrevistas com especialistas, comentários da comunidade científica, a *framework* de Moody and Shanks e os princípios de Österle.

Palavras-chave: Análise de Decisão com Múltiplos Critérios, Computação na Nuvem, MACBETH, Qualidade de Serviço, Serviços Cloud e Serviços de Tecnologias de Informação.

Acknowledgements

I would like to express my gratitude to everybody who has part of my journey and contributed to complete my thesis work. First and foremost, my father, mother, and family for all their support, love, and sacrifice.

To my supervisor, Prof. Miguel Mira da Silva, who always accompanied and supported me with a great commitment and pleasure. I thank him for all the knowledge, motivation, inspiration, critical thinking, and all discussions that gave me considerably experience and improved my research results.

To Prof. João Lourenço for his commitment, effort, help, and knowledge in the final part of my thesis.

To Eng. João Paulo Santos, Eng. Carlos Mendes, and Eng. Rúben Pereira for their detailed and relevant suggestions, information provided, advises, support, and the needed formal ground for my thesis. Moreover, I would like to thanks to my colleagues for their support and help in different parts of my thesis.

I also would like to thank all people and organizations that helped me to collect all information used in evaluation and demonstration of my proposal and also for their wise opinions.

Finally, I would like to thanks to my friends for their presence in all the different moments of my life and for their comprehension when I abdicated for festive moments with them.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AHP	Analytical Hierarchy Process
AIIA	Australian Information Industry Association
APDEX	Application Performance Index
ASP	Application Service Provider
BPO	Business Process Outsourcing
CapEx	Capital Expenses
CLP	Cloud Legal Project
DM	Decision Maker
DSRM	Design Science Research Methodology
EC	European Commission
eSCM-CL	eSourcing Capability Model – Client Organizations
IaaS	Infrastructure as a Service
ICTSQ	Information and Communication Technology Service Quality
IS	Information Systems
IT	Information Technology
ITO	Information Technology Outsourcing
KPI	Key Performance Indicators
MACBETH	Measuring Attractiveness by a Categorical Based Evaluation Technique
MAVF	Multi-Attribute Value Function
MAVT	Multi-Attribute Value Theory
MAUF	Multi-Attribute Utility Function
MAUT	Multi-Attribute Utility Theory
MCDA	Multiple Criteria Decision Analysis
NIST	National Institute of Standards and Technology
OpEx	Operational Expenses
PaaS	Platform as a Service
QoS	Quality of Service
SaaS	Software as a Service

SLA	Service Level Agreement
SME	Small and Medium Enterprises
SMI	Service Measurement Index
SMICloud	Service Measurement Index Cloud Framework
TCO	Total Cost of Ownership

Chapter 1

Introduction

The Outsourcing is divided in two big areas: Information Technology Outsourcing (ITO) and Business Process Outsourcing (BPO). The global ITO market has increased each year since 1989, when a global ITO was only $9 to $12 billion market. In 2007, the global ITO market was estimated to be between a $200 to $250 billion market. The BPO market in 2008 was less than ITO market, but growing at faster rates. It is expected that Outsourcing revenues will increase year after year. It is very clear that Outsourcing of Information Technology (IT) and business services is moving into becoming an almost routine part of management and representing in many major organizations and government agencies the greater percentage of their IT expenditure (Willcocks and Lacity, 2012).

On the other hand, IT industry is evolving and today a new offer is in place, Cloud Computing, which is revolutionizing and changing the IT industry. For decades, most organizations have invested significant capital to own their software and hardware, and Cloud Computing offers a radically different paradigm where organizations contract their services from the Cloud rather than owning them. So the decision maker (DM) in organizations need to take responsibility for deciding in adopting Cloud Computing or choose another IT sourcing more appropriate for their business and strategy (McAfee, 2012).

But, despite the maturity growing of the Cloud Computing concept within organizations, and the growing adoption of such services, most DMs continued to express some concerns (IDC Portugal, 2012). This shows that they are with doubts about what, when, and how they should migrate to the Cloud, because there are no guidelines to Cloud Computing (Conway and Curry, 2012).

Concordantly, we started our problem as: **decision makers in organizations do not know how to evaluate Cloud services**, leading to a postponing about the decision to migrate their services to Cloud.

To address this problem, our research proposes a method to support the decision-making to evaluate the Cloud services, which consists in a set of evaluation criteria to evaluate Cloud services compound by 19 evaluation criteria divided in six groups that IT practitioners identified as best criteria, using MACBETH (Measuring Attractiveness by a Categorical Based Evolution Technique) method, which is an interactive Multiple Criteria Decision Analysis (MCDA) approach used to build a quantitative (numerical) value model (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986) based on quantitative (non numerical) pairwise comparison judgments (Bana e Costa and Vansnick 1994, 1996; Bana e Costa et al., 2005). In short, our objective is to propose a **method to evaluate Cloud services using MACBETH**.

In order to demonstrate the use of our proposal, we applied it within a public organization, a City Council in Portugal. We chose this organization as it had doubts about what was the best Cloud service that meet their organization strategy and business, which is precisely the problem we intend to help solve with our proposal. The Cloud services evaluated and compared in the demonstration were Google Apps and Microsoft Office 365. At the end of the process we obtained an overall value score for each option, which depicted their overall attractiveness for the City Council.

To evaluate the proposal and its results we used interviews with customers, providers, and experts of Cloud services, the appraisal from the scientific community through the submission of scientific publications, Moody and Shanks Framework (Moody and Shanks, 2003) to evaluate the produced models, and the four principles proposed by Österle et al. (2011) to validate if our proposal met the objectives. After the evaluation we conclude that the proposal method is suitable for evaluating Cloud services

Regarding the communication, we submitted two papers in international conferences, whereupon one was accepted and the other one is under revision. In addition, we made some presentations.

1.1 Research Methodology

The research methodology we used in this work is the **Design Science Research Methodology** (DSRM). Towards the end of the 1990s began growing in popularity for use in scholarly investigations in Information Systems (IS). DSRM is conducted in two steps: build and evaluate. In contrast with behaviour research, design-oriented research builds a conception and posteriorly seeks to build the system according to the defined model taking into account restrictions and limitations (Österle et al., 2011). DSRM addresses research through the building and evaluation of artifacts designed to meet identified business needs (Hevner et al., 2004).

This methodology is a system of principles, practices and procedures required to carry out a study. IS can draw advantage from DSRM by often using theories from diverse disciplines, such as social science, engineering, computer science, economics, and philosophy to address problems at the intersection of IT and organizations (Hevner et al., 2004). Several researchers have succeeded in integrating design as a major component of research in order to solve relevant organizational problems (Hevner et al., 2004; Peffers et al., 2008).

To overcome these organizational problems, DSRM proposes the creation and evaluation of IT artifacts that may include **constructs** (vocabulary and symbols), **models** (abstractions and representations), **methods** (algorithms and practices), and/or **instantiations** (implemented and prototype systems) (Hevner et al., 2004). These are concrete prescriptions that IT researchers enable to understand the problems inherent in developing and implementing successfully IS within organizations. In this research the proposed artifact is a **method** because the artifact provides guidance on how to solve the identified problem, i.e., how to search the solution space.

The construction and evaluation of the designed artifacts follows a sequential order, illustrated in Figure 1. This process is iterative and is divided into six different steps, which are explained below.

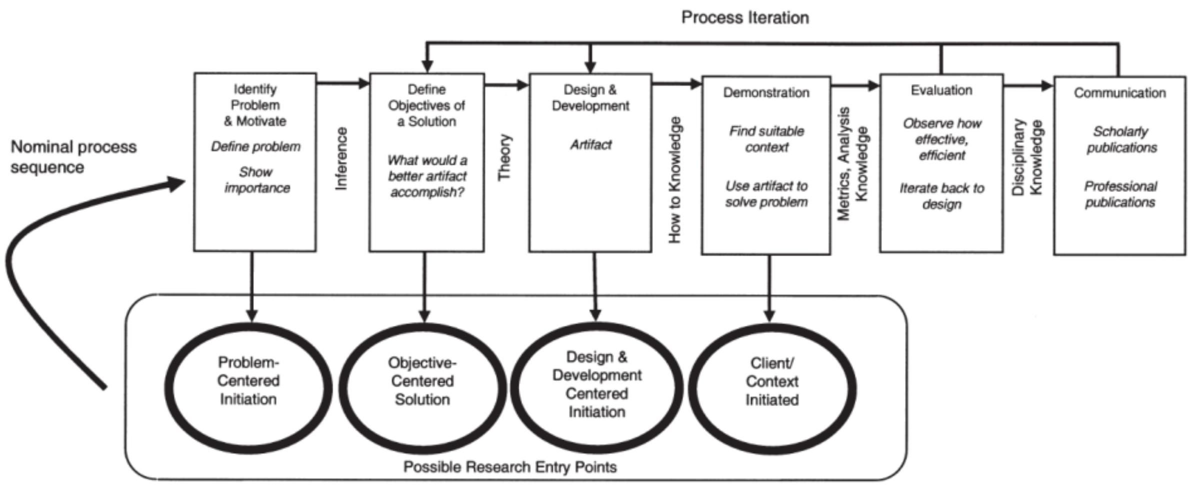

Figure 1. DSRM Process Model (Peffers et al., 2008)

Our research entry point was the problem, so we had a problem-centered initiation. The steps in the previous diagram are described as follows:

1. **Problem identification and motivation**: defines a specific research problem and justifies the value of a solution. The problem definition will be used to develop an artifact that can effectively provide a solution. In this research, we considered an unsolved problem, without an effective or efficient solution. Our research problem becomes that DMs in organizations do no know how to evaluate Cloud services (**Chapter 2**).
2. **Definition of the objectives of a solution**: to infer the objectives of a solution from the problem definition, related work, and knowledge of what is possible and feasible. The objectives should be inferred rationally from the problem specification and can be quantitative, such as the description of how a new artifact is expected to solve problems, or qualitative,

such as terms in which a desirable solution would be better than the current ones. The objective of the solution is the creation of a systematic approach to evaluate Cloud services and our related work tried to show what are the flaws of the actual methodologies and frameworks (**Chapter 3**).

3. **Design and development**: creates the artifact. A design research artifact can be any designed object in which a research contribution is embedded in the design. To create that artifact we need to determine that artifact's desired functionality and its architecture based on related work. It may extend the knowledge base or apply existing knowledge in new ways. Our artifact is an evaluation method for Cloud services (**Chapter 4**).
4. **Demonstration**: shows the use of the artifact to solve one or more instances of the problem. This could involve its use in experimentation, simulation, case study, proof, or other appropriate activities. Our demonstration was made in a Portuguese City Council, which had several doubts about what Cloud service should choose and with our proposal we tried to help them (**Chapter 5**).
5. **Evaluation**: consists to observe and measure how well the artifact supports a solution to the problem. The utility, quality, and efficacy of the design artifact are demonstrated via well-executed evaluation methods. These methods may include interviews with practitioners, surveys, simulations, and the appraisal of the scientific community, among other strategies. At the end of the activity, researchers can decide whether to iterate back to improve the artifact. In our evaluation we made interviews with practitioners, we got feedback from the scientific community, and we used Moody and Shanks Quality Framework and Österle principles (**Chapter 6**).
6. **Communication**: concerns with the presentation of the whole research to relevant audiences. This step was accomplished with the submission of papers for international conferences and with the presentation of the master thesis (**Chapter 7**).

1.2 Thesis Structure

This document is divided in eight different chapters, described as follows:

1. **Introduction** (Chapter 1) gives a general context about the thesis, shows the research methodology, and describes the structure of the document.
2. **Problem and Motivation** (Chapter 2) describes in detail the research problem and motivation.
3. **Related Work** (Chapter 3) presents a brief overview of the literature on the research area, and describes the needed concepts that underlie the proposal, which was crucial for our proposal's coherence. In addition, shows the objectives that our proposal wants to address.
4. **Proposal** (Chapter 4) identifies the objectives of the solution and explains the artifact produced.

5. **Demonstration** (Chapter 5) presents the artifact demonstration in an organization and the conclusions about the results.
6. **Evaluation** (Chapter 6) provides an explanation of the evaluation strategy used to access the artifact, the analysis of the artifact, the evaluation results, and a discussion about the results.
7. **Communication** (Chapter 7) describes where and how we obtain the appraisal from the scientific community.
8. **Conclusion** (Chapter 8) presents a summary of the main conclusions, limitations, and contributions of the thesis, and some proposals for the future work.
9. **Appendixes** present:
    A. Mapping SMI Attributes to Literature.
    B. Judgments Matrices

Chapter 2

Problem

This chapter corresponds to the **problem identification and motivation** step of DSRM, which defines the specific research problem and justify the value of a solution. Although, the problem becomes clear from the introduction of this document, it is important to structure it and explain its justification and motivation.

Currently, Cloud Computing is experiencing a strong adoption in the market and this trend is expected to continue, but the risks and costs in integration with the Cloud is one of the top concerns that IT managers in most organizations in world had about Cloud Computing (Weeden and Valiente, 2012).

As Paolo Malinverno, vice-president at Gartner, said:

> *"The opportunities for Cloud Computing value are valid all over the world, and the same is true for some of the risks and costs. However, some of Cloud Computing's potential risks and costs – namely security, transparency, and integration – which are generally applicable worldwide take on a different meaning in Europe."*[1]

Cloud Computing is still in its beginning and is quite far from mature. Therefore it remains unclear as to extent to which fears of Cloud are reasonable in long term and if the total cost of ownership (TCO) is favorable to costumer comparing with owning the systems (Willcocks and Lacity, 2012). The doubts about Cloud are instilled by those with the stake in an old-computing model against the new. Figure 2 shows that opinion of organizations about the stigma on Cloud Computing are related with key challenges like security and privacy, offshore data housing, lock-in, and compliance.

[1] Gartner Says Cloud Adoption in Europe Will Trail U.S. by at Least Two Years: http://www.gartner.com/it/page.jsp?id=2032215

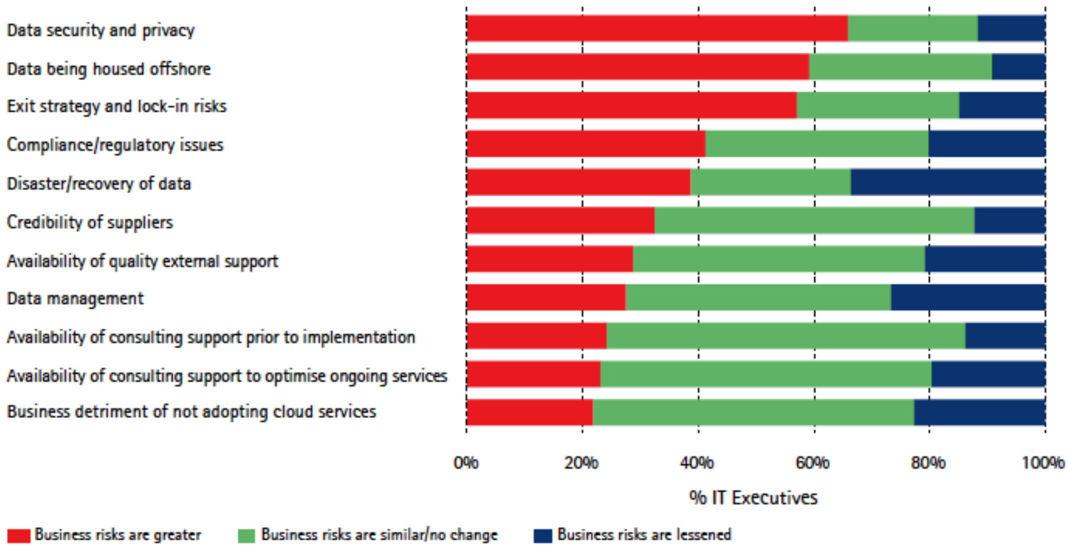

Figure 2. Comparative risks of Cloud business services (Sample: 214 Enterprises) (Willcocks and Lacity, 2012)

However, when assessing the challenges of Cloud Computing, there is a need to avoid focusing wholly on these concerns, because the technology is changing, legislation is uncertain, and decisions must be taking into account multiple criteria to address the business and strategic objectives of organizations. One of the key challenges that hold back business from adopting Cloud services, even if they found it cost effective is that, by migrating to the Cloud, they move their information and services out of their control (Dastjerdi, Tabatabei, and Buyya, 2012). To reinforce this idea Australian Information Industry Association (AIIA) advice for the risks:

> *"Over-empathizing data sovereignty and possible loss of control will play to agencies not wanting to change the status quo and this may compromise higher level aims to reduce cost and increase efficiencies."* (Twentyman, 2012, p. 3)

In addition, is common that different DMs weighed risk differently, and this difference is most noticeable when comparing business and IT executives. They agreed that security and offshoring of data were significant new risks introduced by moving to Cloud. In practice, they have greater concerns over all contractual issues. However, IT executives also weighted some issues as much greater risks than did the business executives and they also saw the business detriment of not adopting Cloud service as a much lower risk than did their senior business counterparts. We can see this in Figure 3, which points to a major challenge in organizations as they move to Cloud: How to bridge the Cloud 'risk perception gap' that exists in the organizations between business and IT executives (Willcocks and Lacity, 2012).

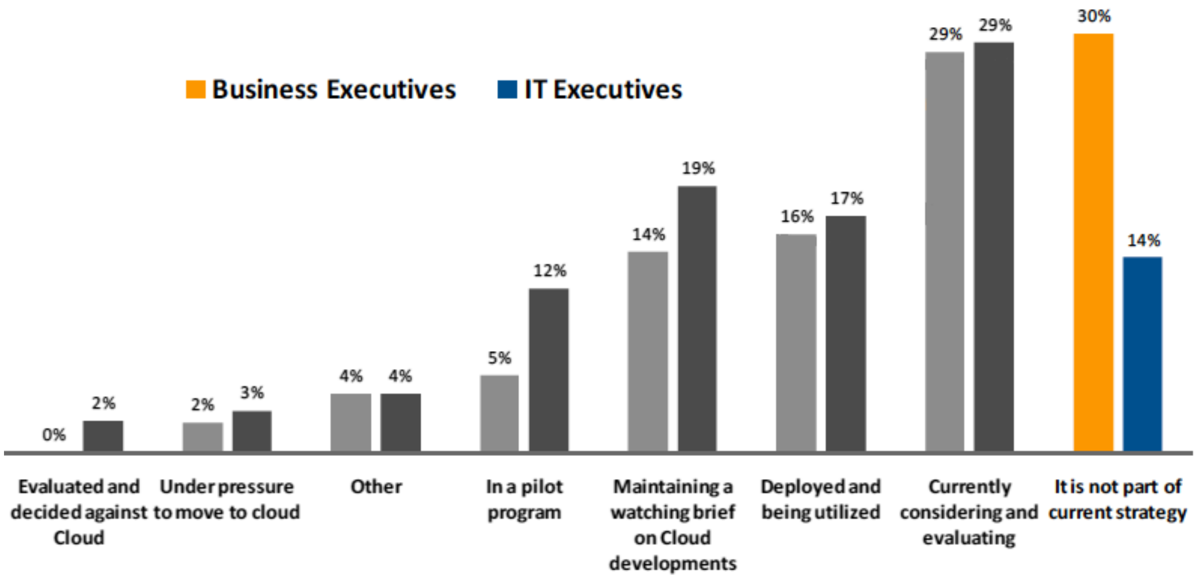

Figure 3. Cloud Evaluations (Source: HfS Research and the outsourcing unit at the London School of Economics, November 2010. Sample: 628 Enterprises)

Furthermore, most organizations do not talk on the record, and some information related with mistakes, made when organizations are implementing Cloud, was relayed in secondhand by the Cloud Computing vendors themselves (Staten, 2008).

In Portugal, the levels of Cloud adoption are also low (IDC Portugal, 2012). As illustrated by Figure 4 only few more than 30% of the companies had already adopted Cloud services, but, on the other hand, about 35% do not have plans to adopt Cloud. The other companies are planning the adoption but we do not know when they will truly do it.

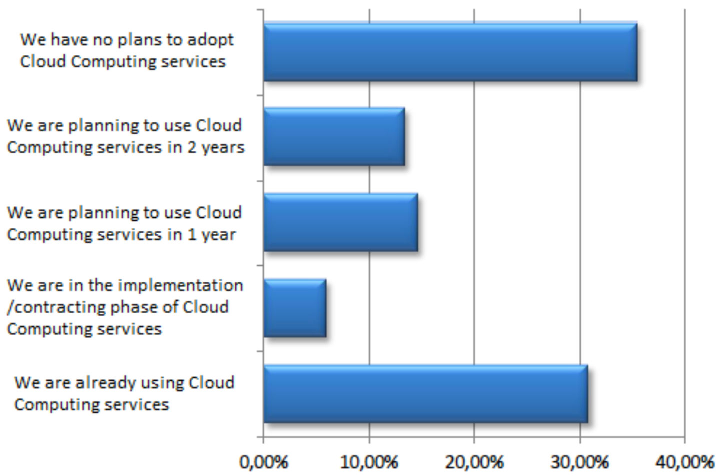

Figure 4. Cloud services adoption (Sample: 646 CIOs) (Adapted from: IDC Portugal, 2012)

Concerning about the critical function and services, 45% said that have no plans to Cloud Computing adoption and only 13% has already changed their critical functions and services to the Cloud (IDC Portugal, 2012). These is visible in Figure 5, which shows the incipient adoption of Cloud services, and that the most utilization of Cloud was for e-mail and videoconference, with the remaining applications having few utilization taxes. This also proves that DMs do not like the idea to place the critical functions and services in Cloud.

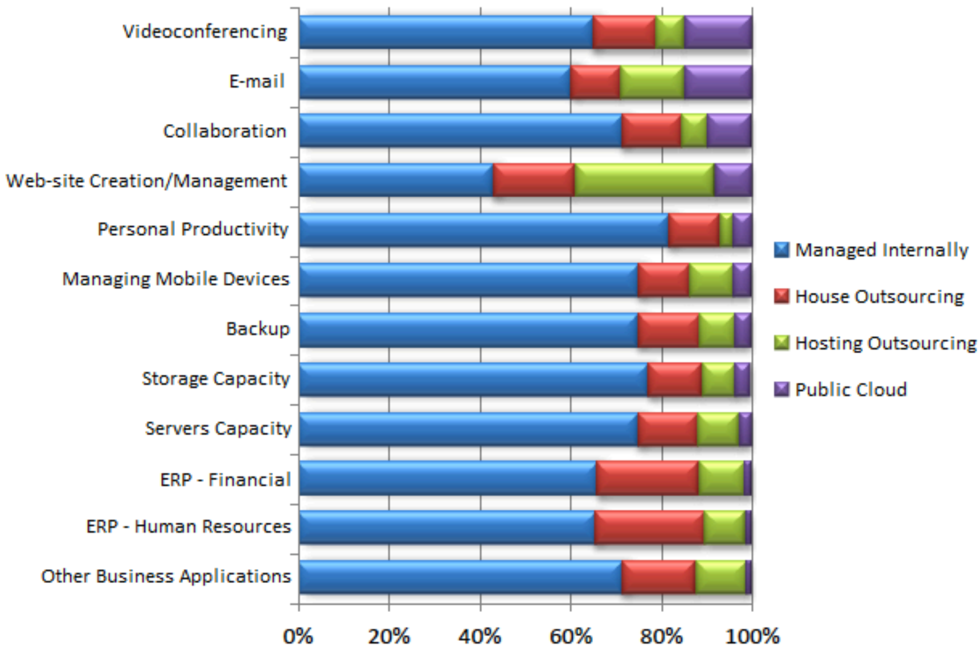

Figure 5. Applications already used in Cloud (Sample: 646 CIOs) (Adapted from: IDC Portugal 2012)

As we refer before, our research problem is about the DMs at most organizations in IT world have doubts about the Cloud adoption and what way they should follow now because they do not have the knowledge about the real benefits, risks, and cost effectiveness, and the evidences are clear. Associated with this, the resistance to change and the unknown of a Cloud solution are always connoted as more investment and risk, which causes the postponing of the decision to migrate their services or functions to Cloud. Therefore, organizations need a systematic solution to evaluate and review their business needs and weigh the potential gains and opportunities by Cloud against the challenges and risks, to make a well-planned and understood strategy (Conway and Curry, 2012).

Our main motivation behind this research problem is that a well-founded proposal may have big impact in the IT services market with respect to benefits, costs, and value that Cloud Computing added to IT industry. We can add important knowledge to DMs because despite being a highly discussed topic, there is not consensual opinion and criteria to address this problem.

In short, our research problem can be summarized, as **decision makers in organizations do not know how to evaluate Cloud services**.

Chapter 3

Related Work

This chapter covers half part of **definition of the objectives for a solution** of DSRM, in which we infer the goals of the solution from the related work. We begin with an overview about IT Outsourcing (Section 3.1). Then, we are going to give an overview of the existing evaluation models with we can compare the IT Outsourcing solutions, following by a discussion (Section 3.2). Afterwards, we are going to give a brief description about MCDA approaches, also followed by a discussion (Section 3.3). Finally, we are going to explain the objectives of a solution (Section 3.4).

3.1 IT Outsourcing

In this section we are going to explain sourcing types used in IT services in the business world. The main sourcing types are traditional Outsourcing and Cloud Computing.

3.1.1 Outsourcing

As we said before, the Outsourcing is divided in two big areas: ITO and BPO. ITO is a phrase used to describe the practice of seeking resources, or subcontracting, outside of an organization structure for IT functions. An organization would use ITO for functions ranging from infrastructure to software development, maintenance, and support. On the other hand, BPO can be defined as Outsourcing specific IT-intensive business process to an external provider who owns, administers, and manages it, according to a defined metrics. An organization could use BPO in terms of various back office functions, like human resources, finance and accounting, processing of insurance and credit transactions, and procurement (Penter, Pervan, and Wreford, 2008). The global ITO market has increased each year since 1989, when a global ITO was only $9 to $12 billion market. In 2007, the global ITO market was estimated to be between a $200 to $250 billion. The BPO market in 2008 was less than ITO market, but growing at faster rates (Willcocks and Lacity, 2012).

Outsourcing can be defined as developing a contract with an external organization that has the primary responsibility for the providing related management process. It is increasingly used to refer to subcontracting a set of processes or services, and it can be seen as a disaggregation of the organization itself and as a new way for create and distribute optimal value from and around the existent fundamentals of business. The main idea is that the organizations, in order to be successful, need to be focus in their core competences and outsource parts of their business, which they cannot do with highest level. When organizations choose the Outsourcing way, they can opt between six main types of Outsourcing (Willcocks and Lacity, 2012; Lacity and Willcocks, 2012):

- **Offshoring:** occurs when the Outsourcing is performed in a country that is not the one where the service will be developed and delivered. This can deliver lower contract prices when organization outsource for locations with weak local economy, but this also increase the communication costs between client and provider.
- **Onshoring:** implies the same of Offshoring but the Outsourcing supplier is in the same country as the client.
- **Nearshoring:** happens when the Outsourcing is made in a bordering or near country. It has the benefit of ensure lower communication cost between client and provider when comparing with Offshoring.
- **Rural Outsourcing:** is the practice of Outsourcing work to providers with delivery centers located in low-cost, non-urban areas. The model promises lower costs to clients than urban-based centers and higher quality of work, superior domain knowledge, higher cultural compatibility, closer proximity to the customer, and time zone advantages over offshore provider.
- **Impact Outsourcing:** it is like Rural Outsourcing but with a social responsibility strategy. The main concept is hire and train marginalized people to perform ITO or BPO work.
- **Insourcing:** also known as in-house. Is when an organizations uses especially internal labor and personnel, but other resources as well, to supply the operational needs of its organization.

There are basically four reasons whereby organizations decided to realize Outsourcing: economical, strategic, technological, and political (Mira da Silva and Martins, 2008). The **economical** motivations are the main one, because the main reason to realize Outsourcing is normally permits cost reduction with the decrease of IT service costs and increase of efficiency and effectiveness. The increase of competition in all industries also increases the necessity of cost reduction. The **strategic** reasons are related with the fact of organizations have the necessity to focus on the main business to take more adaptability on market changes. The **technological** reasons are related with the technological evolution that makes more difficult the actualization of organizations infrastructures. It is more easy and cheapest and permits be always updated with the latest innovations. Finally, the **political** reasons are used when we do not have certainties of our IT investments. When we outsource these services we put the responsibility on the provider and if something is wrong is provider fault and they have to answer for them (Mira da Silva and Martins, 2008; Hirschheim, George, and Wong, 2004; Lacity and Hirschheim, 1993).

Despite this, there are some risks and problems associated to ITO and BPO that organizations have to think about. The main risks and problems are: escalation, winner's course, communication, and rigid contracts. The **escalation** is defined by the continued commitment in the face of negative information about prior resource allocations coupled with uncertainty surrounding the likelihood of goal attainment. The **winner's course** happens when the supplier is unable to deliver what specified on the contract and makes no profit, this lead to a negative impact on the service provided to the client. This situation is usually created when the supplier might enter in a contract that favors the client in order to win the deal. The **communication** problems that can be the result of misunderstood contracts and can lead to an unexpected result. Finally, the **rigid contracts** leading to a lack of space to innovative from both parts and usually to legal issues.

Furthermore, ITO and BPO clients lose some flexibility to make decisions, they can lose some partners and have more difficult to innovate. On the other hand, is necessary having knowledge to manage correctly their providers (Mira da Silva and Martins, 2008). However, in order to manage their providers, clients must have certain knowledge and the source of successful knowledge sharing is an organizational ability to learn and acquire the needed knowledge from other organizations. Furthermore, organizations need to have in account that Outsourcing is neither good nor bad in it. The results from an outsource contract will depend of how the organizations minimize the risks and manage the contract (Cullen, Seddon, and Willcocks, 2005; Earl, 1996; Lee and Kim, 1999).

3.1.2 Cloud Computing

The emergence of the Cloud Computing phenomenon represents a fundamental change in the way IT services are invented, developed, deployed, scaled, updated, maintained, and paid for. The promise of Cloud Computing is to deliver all the functionality of existing IT services even as it dramatically reduces the upfront costs of computing that deter many organizations from deploying many cutting-edge IT services.

Cloud Computing is defined by the National Institute of Standards and Technology (NIST), and mostly used by researchers, as:

> *"a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"* (Hogan et al., 2011, p. 10).

Also to NIST, this Cloud model is composed of four deployment models, three services models, and five essential characteristics (Hogan et al., 2011).

The four **deployment models** present in NIST Cloud Computing are:

- **Private Cloud**: Cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers and may be owned, managed, and operated by the organization, a third party, or some combination of them.
- **Community Cloud**: Cloud infrastructure is shared with a specific community of consumers form organizations that have shared concerns and it may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them.
- **Public Cloud**: Cloud infrastructure is may be owned, managed, and operated by a business, academic, or government organization, or some combination of them for open use by the generic public.
- **Hybrid Cloud**: Cloud infrastructure is a composition of two or more distinct Cloud infrastructures that remain unique entities, but are bound together by standardized or proprietary technology that enable data and application portability.

The three **service models** descripted in NIST Cloud Computing are:

- **Software as a Service (SaaS)**: The provider's applications running on a Cloud infrastructure. The applications are accessible from various client devices through either a thin client interface, or a program interface. The consumer does not manage or control the underlying Cloud infrastructure.
- **Platform as a Service (PaaS)**: Provides an environment to deploy onto the Cloud infrastructure consumer-created or acquired applications. The consumer does not manage or control the underlying Cloud infrastructure, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.
- **Infrastructure as a Service (IaaS)**: Provides the capability to provision processing, storage, networks, and other resources where the consumer is able to deploy and run arbitrary software. The consumer does not manage or control the underlying Cloud infrastructure.

Finally, the five **essential characteristics** to NIST Cloud Computing are:

- **On-demand self-service**: A costumer can unilaterally provision computing capabilities as needed automatically without requiring human interaction with each service provider.
- **Broad network access**: Capabilities are available over the network and accessed through standards mechanisms that promote use by client platforms.
- **Resource pooling**: The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned according to consumer demand.
- **Rapid elasticity**: Capabilities can be elastically provisioned and released to scale rapidly outward and inward commensurate with demand.

- **Measured service**: Cloud systems automatically control and optimize resource usage by leveraging a metering capability at some level of abstraction appropriate to the type of service.

Traditionally, small and medium enterprises (SME) had to make high capital investment upfront for procuring IT infrastructure, skilled developers, and system administrators, which results in a high TCO (Garg, Versteeg, and Buyya, 2012). Based on this, and knowing that the economic appeal of Cloud Computing is often described as converting capital expenses to operating expenses (CapEx to OpEx), the phrase pay-as-you-go are more directly captures the economic benefit to the buyer (Armbrust et al., 2010). Along with that, organizations are also searching for functional and resources benefits (Conway and Curry, 2012). Accordingly, it is expected that Cloud Computing to be a $150 billion business by 2014, SMEs are expected to spend over $100 billion on Cloud by 2014 (Armbrust et al., 2012).

DMs have four desires when they are looking for Cloud adoptions. These desires are on a *'desires framework'*, which strips out the value-added benefits of Cloud allows organizations to focus on the specific differences, and thus make decisions on alongside the promised benefits (Willcocks and Lacity, 2012; Lacity and Willcocks, 2012). Their framework consists of four desired dimensions of the different offerings, showed in Table 1.

Desired Dimension	Definition
Equivalence	Provider services which are at least equivalent in quality to that experienced by a locally running service on a PC or server
Abstraction	Hide unnecessary complexity of the lower levels of the applications stack
Automation	Automatically manage the running of a service
Tailoring	Tailor the provided service for specific enterprise needs

Table 1. Cloud Desires Framework (Willcocks and Lacity, 2012)

Nevertheless, Cloud Computing has some issues that organizations have to examine deeply.

First of all, Cloud risks are related with data security and privacy, data being housed offshore, exit strategy and lock-in risks, compliance/regulatory issues, recovery data, credibility of suppliers, availability of quality external support, and business detriment of not adopting Cloud.

Then, DMs are facing with four main challenges when they are thinking about Cloud adoption. The first challenge is **weighing up the security and legal risks**, these encompass the most perceived risks for Cloud, and they are related with data security and privacy together with offshore data housing and security. The second challenge is **defining the relationship through contracting**, because these contracts are focused typically on the service level agreements (SLA) regarding security and service quality, but these SLAs are not robust enough for the Cloud business model. The third challenge is **the lock-in dilemma**, and this challenge has two forms: technology lock-in and institutional lock-in. Technology lock-in occurs when the cost of moving a service inhibits taking a business from one Cloud platform to another. Institutional lock-in occurs when users become attached to the technologies embedded in organization routines. Finally, the fourth challenge is **managing the Cloud** that has two key issues: maintaining strategic control and managing Cloud services. The strategic control is important because, once Cloud services have been introduced into the organizations, they can be updated and changed easily by technology providers without internal IT's control or direction. Managing Cloud service covers a broad of range topics like monitoring usage, SLAs, performance, robustness, and business dependency (Willcocks and Lacity, 2012; Lacity and Willcocks, 2012).

To reduce these challenges, the current efforts are focused on creating policies, regulations, and contract negotiation for Cloud services. Currently, institutes like NIST, European Commission (EC), and Cloud Legal Project (CLP) are creating policies and guidelines for security and privacy (Jansen and Grace, 2011; Robinson et al., 2011; Walden, 2011). Although in some states of the United States of America (USA) states and European Union (EU) countries the validity of such contracts and terms may be challenged under consumer protection laws, Cloud users may face obstacles to bring a claim for data loss and privacy breach against provider that seems local online but is based in another continent. Indeed, providers claim that their contracts are subject to the laws of the place where they have their main business (Bradshaw, Millard, and Walden, 2011). In addition, clients and providers have different perspectives about contracts so it is necessary to comply with laws and regulations (Hon, Millard, and Walden, 2012).

3.1.3 Discussion

As we said above, organizations have four main reasons to organizations opt for Outsourcing. These reasons are economical, strategic, technological, and political. All DMs have to reduce costs and increase efficiency, easily adapt to market changes, actualize infrastructures, innovate, and pay attention to laws and regulations, in these points Outsourcing can help them. Whereas, Cloud Computing is an evolution of traditional Outsourcing, there are some risks like flexibility to make decisions, lose some partners, innovation, performance, and others, which are more easy to manage with Cloud Computing that have premises based on on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. In the other hand, Cloud has some issues that organizations have to take into account, as security and privacy, compliance, SLAs, and other exposed above. Despite the differences of Cloud deployment upon traditional Outsourcing, which is similar to private Cloud, this choice will depend on each organization's key criteria and their needs to keep competitive.

3.2 Evaluation Models

In this section we will explain the actual methodologies or models that try to make the assessment or propose criteria to evaluate the IT Outsourcing and/or Cloud Computing services.

3.2.1 Information and Communication Technology Service Quality (ICTSQ)

The development of ICTSQ was based on many models that have been developed in the concern of quality of service (QoS). However, most of these models have not investigated the performance measurement for their QoS. To support their opinion they argue that the literature has only focused on analyzing and identifying various factors and attributes based on the context involved. Therefore, their studies tried to fill this gap by identifying the related key performance indicators (KPI), involved in measuring ICTSQ, and construct the measurement tool. Besides, their study lists the related factors and KPIs based on the component involved (Nor, Alias, and Rahman, 2008). The model is shown in Figure 6.

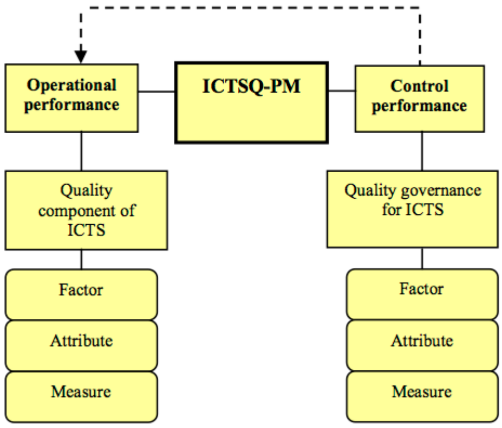

Figure 6. ICTSQ Model (Nor, Alias, and Rahman, 2008)

ICTSQ model only has components related with operational and control performance. The **operational performance** is defined as an accomplishment of routine functioning and activities of an organization, on the other hand, the **control performance** is an accomplishment of control process to ensure the functions such as plan, coordinate, and monitor are orderly implemented. These two components still contain three components each. The **variable**, which describes the component; the **factor**, which is defined as the measurement for variable to indicate the achievement fro the evaluation of performance; and the **attribute**, which determine metrics to this evaluation, and the measure that guides the performance score will be given (Nor, Alias, and Rahman, 2008).

Despite this, they proposed the examples of KPIs measures that they will pretend to use in their model. These KPIs are the mainly contribution to measure the performance of ICTSQ. Hence, for measures purposes their study will develop the KPI based on the context involved and the given example of KPIs are: reliability, staff training and skill, staff concern and attitude, responsive, proactive, technology usage, and service marketing (Nor, Alias, and Rahman, 2008).

The model has only concerns about performance and the KPIs measures are also related with performance. We think that only performance is insufficient to make the approach of our problem. Other issue is that anything says what measure is more important than other in order to help the DM.

3.2.2 ISO/IEC 9126: Information Technology – Software Product Evaluation – Quality Characteristics and Guidelines for their use

This International Standard defines six characteristics that describe software quality and provide a baseline for further refinement as well as the description of software quality and guidelines that describe the use of quality characteristics for the evaluation of software quality. These characteristics are (ISO/IEC, 1991):

- **Functionality**: which bears on the existence of a set of functions and their specified properties;
- **Reliability**: which bears on the capability of software to maintain its level of performance under stated conditions for a stated period of time;
- **Usability**: which bears on the effort needed to use;
- **Efficiency**: which bears on the relationship between the level of performance of the software and the amount of resources used;
- **Maintainability**: which bears on the effort needed to make specified modifications;
- **Portability**: which bears on the ability of software to be transformed from one environment to another.

3.2.3 Application Performance Index (APDEX)

APDEX is an open standard that defines a standardized method to report, benchmark, and track application performance. It is a critical metric of application performance that has direct implications for business revenue, customer retention, and user productivity. Aiming to help IT organizations to measure the service levels and customer satisfaction with insight into how well their applications perform from a business point of view based on response time, averaging samples, time values, and other measures, APDEX intends to be a better way to analyze and measure what matters. For this, APDEX proposes a numerical measure of user satisfaction with the performance of enterprise applications. It converts many measurements into one number on a uniform scale of 0 (no users satisfied) to 1 (all users satisfied). APDEX constitutes a comparable metric across applications, measurement approaches, and enterprises. This metric can also be applied to any source of end-user performance measurements and fills the gap between timing data and insight by specifying a uniform way to measure and report on the user experience (Sevcik, 2005; Apdex Alliance, 2007).

APDEX is a first user experience metric that makes a way to compare performance across application or other reporting groups he or she defines. This metric can easily understand and use to manage IT across many applications. But the greatest benefit is its ability to quickly show the alignment of application performance to the needs of the business. APDEX process (Sevcik, 2005) has three main phases through which the samples are treated. The first one is based on three zones of application responsiveness:

- **Satisfied**: the user is fully productive. This represents the time value (T seconds) below which users are not impeded by application responsive time. *Multiplier: T or less.*
- **Tolerating**: the user notices performance lagging within responses greater that T, but continues the process. *Multiplier: between T and 4T.*
- **Frustrated**: performance with a response time greater than T seconds is unacceptable, and users may abandon the process. *Multiplier: more than 4T.*

Then, the samples in these three zones are counting and APDEX score is calculated using the formula that consists in the number of satisfied samples plus half of the tolerating samples plus none of the frustrated samples, divided by all samples:

$$APDEX_T = \frac{Satisfied\ Count + \frac{Tolerating\ Count}{2}}{Total\ Samples} \qquad (1)$$

Finally, after calculating the score, APDEX defines a rating system when are displayed using quality colors, this rating are present in Figure 7.

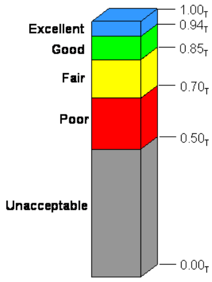

Figure 7. APDEX rating system (Apdex Alliance: http://www.apdex.org)

APDEX process needs a lagging process that includes the requirements elicitation and definition that can constitute a waste of time for enterprises. Nevertheless, many enterprises adapt APDEX to their business and use them in an unstructured form that does not bring significant advantages for the measurement of their provided services. Furthermore, we think that an evaluation based on single number is not useful because only it does not say what to do in case of an improvement needed, the used formula is not justified as appropriate, and in a typical IT enterprise we can have several service applications so we will have several APDEX indexes.

3.2.4 eSourcing Capability Model – Client Organizations (eSCM-CL)

eSCM-CL is a best practices capability model. This model has the best practices that allows client organizations to continuously evolve, improve, and innovate their capabilities to allow better relationships with their service providers, and to meet the dynamic demands of their business while managing the delivery of service providers. This capability has two purposes (Hefley and Loesche, 2006):

- To give client organizations guidance that will help them improve their sourcing capabilities throughout the sourcing life-cycle;
- To provide client organizations with an objective means of evaluating their capability.

This model is also composed by three dimensions (Hefley and Loesche, 2006):

- Life cycle: divided in several phases as analysis, initiation, delivery, and execution;
- Areas: areas grouped by governance, competencies, environment, and several life cycle phases;
- Level: five levels of maturity.

Additionally, the eSCM-CL gives to the client organizations a standard to use in the evaluations to achieve the certification, and is able to demonstrate their sourcing capabilities through certifications. This can be important to have differentiation among competitors (Hefley and Loesche, 2006).

eSourcing, referenced in this capability model, is related with two types of sourcing services. The first one is Task and Business Process Sourcing, which includes finance and accounting, engineering service, human resources, data management, customer care, medical/legal transcription and purchasing. The second one is IT Sourcing, related with application management, data center support, desktop maintenance, telecommunications network support, and application service provider (ASP) (Hefley and Loesche, 2006).

This model is one of the references in services and is very comprehensive and embracing but, furthermore, it does not make reference to what evaluation criteria we should adopt or what are the best criteria in our scenario, Cloud Computing.

3.2.5 Service Measurement Index (SMI)

SMI is a set of business-relevant KPIs that provide a standardized method for measuring and comparing a business service regardless of whether that service is internally provided or sourced from an outside organization. It is designed to become a standard method to help organizations measure Cloud-based business services based on their specific business and technology requirements. SMI is a hierarchical framework, the top level divides the measurement space into seven categories that answer to different questions, illustrated in Table 2, and each category is further refined by three or more attributes (Cloud Services Measurement Initiative Consortium, 2011).

Category	Questions
Accountability	Can we count on the provider organization?
Agility	Can it be changed and how quickly can it be changed?
Assurance	How likely is it that the service will work as expected?
Financial	How much is it?
Performance	Does it do what we need?
Security and Privacy	Is the service safe and privacy protected?
Usability	Is it easy to learn and to use?

Table 2. The seven top-level categories of SMI (Cloud Measurement Initiative Consortium, 2012)

The development of SMI is focused on two objectives: *(i)* construct a framework for organizing and classifying service measures; and *(ii)* create a standard way of describing and documenting service measures. To enable these calculations the seven categories are refined in attributes, presented in Table 3, and for each attribute they are creating a set of KPIs that describe the data to be collected for each metric (Cloud Measurement Initiative Consortium, 2012).

Category	Attributes
Accountability	Auditability; Compliance; Contracting Experience; Data Ownership; Ease of Doing Business; Governance; Ownership; Provider Business Stability; Provider Certifications; Provider Contract/SLA Verification; Provider Ethically; Provider Personnel Requirements; Provider Supply Chain; Security Capabilities; and Sustainability.
Agility	Adaptability; Capacity; Elasticity; Extensibility; Flexibility; Portability; and Scalability.
Assurance	Availability; Data Geographic/Political; Maintainability; Recoverability; Reliability; Resiliency/Fault Tolerance; Service Stability; and Serviceability.
Financial	Acquisition and Transition Cost; On-going Cost; and Profit or Cost Sharing.
Performance	Accuracy; Functionality; Interoperability; Service Response Time; and Suitability.
Security and Privacy	Access Control and Privilege Management; Data Geographic/Political; Data Integrity; Data Privacy and Data Loss; Physical and Environmental Security; Proactive Threat and Vulnerability Management; and Retention/Disposition.
Usability	Accessibility; Client Personnel Requirements; Installability; Learnability; Operability; Transparency; and Understandability.

Table 3. The SMI attributes

These attributes are after used in a method for weighing importance to DM using the rating scale of 0-10 at level of each category and each of the attributes within that category (Cloud Measurement Initiative Consortium, 2012).

3.2.6 Service Measurement Index Cloud Framework (SMICloud)

Based on SMI, SMICloud offer a comparative evaluation framework of Cloud services. SMICloud helps Cloud customers to find the most suitable Cloud provider and therefore can initiate SLAs. The SMICloud framework provides feature such as service selection based on QoS requirements and ranking of services based on previous user experiences and performance of services. It is a decision making tool, designed to provide assessment of Cloud services in terms of KPIs and user requirements. Customers provide two categories of application requirements: essential and non-essential (Garg, Versteeg, and Buyya, 2012).

To realize this model they tackled several challenges to evaluating QoS and ranking Cloud providers. These challenges are: how to measure various SMI attributes of a Cloud service; and how to rank Cloud services based on attributes. The first challenge is most important, and has some value adding to SMI. This value are the creation of the truly KPIs promised to this framework, but the main focus is on the IaaS services. In the second challenge, to rank Cloud services based on multiple KPIs, they proposed a ranking mechanism based on Analytic Hierarchy Process (AHP). This process has three phases: problem decomposition, judgment of priorities, and aggregation of these priorities (Garg, Versteeg, and Buyya, 2012).

In order to evaluate this proposal, we think that is an interesting addition to SMI but does not have all sufficient KPIs to cover all SMI framework and to compare all Cloud services. Additionally, this proposal is only focus on IaaS providers.

3.2.7 Discussion

Considering that the SMI is the best approach, we validated the value added by SMI to DMs and ensured that SMI categories and attributes were referenced in literature, we chose 10 referenced papers related with Cloud Computing and their measurement criteria to reduce the challenges and issues related with the theme.

Category	(Buyya, Yeo, and Venugopal, 2008)	(Nor, Alias, and Rahman, 2008)	(Armbrust et al., 2009)	(Alhamad, Dillon, and Chang, 2010)	(Kundra, 2011)	(NIST, 2011)	(Samimi and Patel, 2011)	(Conway and Curry, 2012)	(Garg, Versteeg, and Buyya, 2012)	(Low and Chen, 2012)
Accountability	✓	✓	✓		✓	✓	✓	✓	✓	✓
Assurance	✓	✓	✓	✓	✓	✓			✓	✓
Agility	✓		✓	✓	✓	✓	✓		✓	✓
Performance		✓	✓		✓	✓		✓	✓	✓
Security and Privacy			✓	✓	✓	✓		✓		
Financial				✓			✓		✓	✓
Usability	✓						✓		✓	

Table 4. Mapping SMI categories to literature

First, to have a bigger picture, we mapped the seven SMI categories with these 10 papers. This map is present in Table 4. Second, to access greater detail, we drill-down to attributes, which are the most important measure criteria in SMI. These maps showed which categories and attributes are more important in literature and which ones can be hidden or improved.

After analyzing the Table 4, we could conclude which were the most important SMI categories. We could see that **accountability**, **assurance**, and **agility** are very important categories to evaluate Cloud services and **performance** and **security and privacy** are also important. On the other hand, without many references, we had **financial** and **usability** categories that presented less worries and could be improved to be more enlightening.

As far as 52 attributes are concerned, only three attributes were referenced five or more times by the chosen literature. These attributes were **reliability,** referenced seven times; **scalability,** referenced six times; and **availability,** referenced five times. The other attributes were referenced less than five times and might belong to the less important attributes for Cloud evaluations (see **Appendix A – Mapping SMI attributes to literature**).

Hereupon, with this literature analysis we also discovered several attributes that could be inserted on SMI. These attributes are: confidentiality; customizability; data transfer; ease to use; efficiency; innovation; integration; laws and regulations; lock-in; quality of use; resources; risks; return on investment; supportability; standard architecture; testability; throughput; and value creation. In this list, we can see that attributes like **confidentiality**, **efficiency**, **laws and regulations**, and **supportability** are the most mentioned and could easily be included on SMI.

After this mapping, we consider that SMI framework needs some improvements to become a standard method to help organizations measure Cloud-based business services. One of these improvements is the redundant attributes that this framework has. We think that SMI has several attributes that we can eliminate and that make it more objective. This was made evident in SMI attributes spoken previously. The other improvement is the missing attributes previously listed, which some of them might have some preponderance in measure Cloud services. The final improvement is related to the fact of nothing mention what measure is more important than other.

Taking into account all methodologies and frameworks and the fact of all organizations have to measure and/or evaluate the services that want to acquire. According to several factors related to actuation scope, complexity, and criteria insufficiency or importance, the models and frameworks referenced above and other literature contents less relevant, do not resolve our research problem. The main flaw of all actual models is the missing or redundant criteria to evaluate Cloud services, which mean a poor assessment and future problems in organizations. Several models, as ICTSQ, ISO/IEC 9126, eSCM-CL, SMI, and SMICloud list the criteria but do not give them the comparative importance that will help DMs when Cloud solutions have different offers. Furthermore, some of these models propose KPIs for each criterion but some of those are calculated based on unclear formulas or are only focus on performance criteria, which is insufficient to evaluate Cloud services, as we saw in ICTSQ, APDEX, SMI, and SMICloud. The other flaws are related to complexity, present in several models used in some IT organizations, as APDEX and eSCM-CL, which makes the process harder to complete and increases process lagging, leading to the need for changes to these models made in-house by those organizations, a fact that can constitute a waste of time for organizations. Table 5 resumes all flaws of the actual solutions.

Flaws	Models					
	ICTSQ	ISO/IEC 9126	APDEX	eSCM-CL	SMI	SMICloud
No criteria			✗	✗		
Missing or Redundant Criteria	✗	✗			✗	✗
No Comparative Importance	✗	✗			✗	
KPIs with Unclear Formulas	✗		✗			✗
Complexity			✗	✗		

Table 5. Mapping the flaws with the actual models

3.3 Multiple Criteria Decision Analysis

Decision has inspired reflection of many thinkers since ancient times. MCDA considers multiple criteria in decision-making environments. In organizations, professional settings there are typically multiple conflicting criteria that need to be evaluated in making decisions. Cost is usually one of the main criteria in all organizations together with measure of service quality that conflicts with the cost (Bana e Costa, De Corte, and Vansnick, 2005).

A decision problem typically involves balancing multiple, and often conflicting, criteria. MCDA consists in "a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that mater" (Belton and Stewart, 2002).

In this section we explain the methodologies that are considering the most used as regards to MCDA.

3.3.1 Multi-Attribute Value Theory (MAVT)

The intention in this approach is to construct a means of associating a real number with each alternative in order to produce a preference order on the alternatives consistent with the DM's judgments. MAVT approach thus constructs preferences that are required to be consistent with a relatively strong set of axioms but in the end result will generally be much less rigidly precise than may be suggested by the axioms (Belton and Stewart, 2002).

The component of preference modeling is achieved by constructing a marginal value functions for each criteria. A fundamental property of the partial value function must be that the alternative *a* is preferred to *b* in terms of criterion *i* if and only if $v_i(a) > v_i(b)$. A central feature of MAVT is that the properties required of the partial value functions and the forms of aggregation used are critically interrelated. The additive aggregation is constructed in the form (Belton and Stewart, 2002):

$$V(a) = \sum_{i=1}^{m} w_i\, v_i(a), \text{with} \sum_{i=1}^{m} w_i = 1 \text{ and } w_i > \ 0 \text{ for } i = 1, \ldots, m \qquad (2)$$

where w_i represents the weight of criterion *i*.

The weights parameters w_i have a very specific algebraic meaning. People are often willing to specify the relative importance of criteria on numerical scales, but these may not necessarily coincide with the algebraic meaning in the additive value function. The weights here are simply scaling constants, which render the different value scales commensurate (Belton and Stewart, 2002).

The simplicity and transparency of additive aggregation is particularly appealing. However, additional assumptions are required in order to justify this additive aggregation. The most important requirement is that the additive aggregation model (2) can only be applied if the criteria are mutually preferential independent (Dyer and Sarin, 1979).

3.3.2 Multi-Attribute Utility Theory (MAUT)

MAUT can be viewed as an extension of MAVT, relating to the use of probabilities and expectations to deal with uncertainty. MAUT can be based on different sets of axioms that are appropriate for use in different contexts. The axioms that are appropriate for risky choice do not have to be satisfied in order to use multiple attribute models of preference for cases that do not explicitly involve risk (Dyer, 2005).

The attribute values, z_i, are not fully determined by choice of the alternative but may also be influenced by random factors. The consequences of each alternative are thus described in terms of a probability distribution on the attribute vector *z*. A utility function, $U(z)$, must ensure that alternative *a* is preferred to alternative *b* if and only if the expectation of $U(z^a)$ is greater than $U(z^b)$, where z^a and z^b are the random vectors of attribute values associated with alternatives *a* and *b* respectively. This condition is called expected utility hypothesis (Belton and Stewart, 2002).

This is not identical to the additive forms of value function, $V(z)$, although the functions must be related to each other via some strictly increasing transformation, since $U(z^1) > U(z^2)$ if and only if $V(z^1) > V(z^2)$. The distinction between $U(z)$ and $V(z)$ arises because the requirements of the expected utility hypothesis and those of seeking simple forms of aggregation, as additive, may be incompatible (Belton and Stewart, 2002).

MAUT is the simplest method that combines various preferences in the form of multi-attribute utility functions (MAUF). In MAUT, utility functions for each criterion are combined with weighted functions. The advantage of using MAUT is that the problem is constructed as a single objective function after successful assessment of the utility function. This, it becomes easy to ensure the achievement of the best compromise solution based on the objective function. MAUT is an elegant and useful model of preference suitable for applications involving risky choices. Very often the utility is the sum of the marginal utilities that each criterion assigns to the considered action (Dyer, 2005).

3.3.3 Outranking Methods

Outranking methods are applied directly to partial preference functions, which are assumed to have been defined for each criterion. These preference functions may correspond to natural attributes on a cardinal scale, or may be constructed in some way, as ordinal scales and do not need to satisfy all of the properties of the value functions, only the ordinal preferential independence would still be necessary (Belton and Stewart, 2002).

In Outranking methods, for two alternatives *a* and *b*, where $z_i(a) \geq z_i(b)$ for all criteria *i*, we can say that *a* outranks alternative *b* if there is sufficient evidence to justify a conclusion that *a* is at least as good as *b*, taking all criteria into account. There are two aspects that distinguish Outranking methods from the others (Belton and Stewart, 2002):

- The emphasis is on strength of evidence for the assertion that *a* is at least as good as *b*, rather than on strength of preference by itself.
- When neither *a* nor *b* outranks the other, a state of indifference is not necessarily implied. In comparing two alternatives, three situations may arise: a definitive preference for one alternative over the other, indifference, or incomparability.

Once preferences for each criterion have been modeled this evidence is aggregated across criteria in order to summarize the current state of information. This process is explicitly dynamic and have to obey to such rules that there are not uniquely defined, leading to the existence of many varieties of Outranking methods. Nevertheless, two recurring themes running through most of these methods (Belton and Stewart, 2002):

- If *a* is demonstrably as good as or better than *b* according to a sufficiently large weight of criteria, then this is considered to be evidence in favor of *a* outranking *b* (concordance principle).
- If *b* is very strongly preferred to *a* on one or more criteria, then this is considered to be evidence against *a* outranking *b* (discordance principle).

The definitions of concordance and discordance do not yet define a full aggregate model. Two broad approaches can be identified. In the crisp approach, a final conclusion is based on thresholds for the concordance and discordance measures respectively. Alternative *a* is asserted to outrank *b* if the concordance measure exceeds a minimum threshold level and if the discordance does not exceed the threshold. In the fuzzy approach, no define conclusion is stated, but the strength of evidence is summarized by value between 0 and 1 obtained by multiplying the concordance measure by a factor that tends to zero as discordance tends to its maximum value (Belton and Stewart, 2002).

3.3.4 Analytical Hierarchy Process (AHP)

The AHP is a method for MCDA developed by (Saaty, 1980), has in its implementation many similarities with multi-attribute value function (MAVF) approach. Both approaches are based on evaluating alternatives in terms of an additive preference function.

The initial steps in using the AHP are to develop a hierarchy of criteria (value tree) and to identify alternatives. Then, it uses pairwise comparisons in comparing alternatives with respect to criteria (scoring) and criteria within families (weighting), and uses the ratio scales for all judgments. Alternatives are not differentiated from criteria, but are treated as the bottom level of the hierarchy and all comparisons follow the same procedure. Afterwards, the DM is required to respond to a series of pairwise comparison questions, which leads to an implied numerical evaluation of the alternatives according to each criterion (Belton and Stewart, 2002).

In this approach, a number of absolute levels of performance on each criterion are defined, and it is these levels rather than the alternatives, which are compared pairwise, to generate numerical scores for each level of performance. Values for each alternative are derived from those of the absolute performance levels for each criterion to which it most closely corresponds. The numerical values used by AHP are a scale of 1 to 9 (1: Equally preferred; 3: Weak preferred; 5: Strong preferred; 7: Demonstrated preferred; and 9: Absolute preferred). Intermediate values are used when DM hesitate between two of the descriptors (Belton and Stewart, 2002; Saaty, 2005).

Once all pairs of alternatives have been compared this way, the numeric values corresponding to the judgments made are entered into a pairwise comparison matrix. To determine a set of relative priorities amongst *n* alternatives, only *n – 1* judgments are in principle needed. Then, it is necessary to synthesizing these judgments in a comparison vector. This vector is the eigenvector corresponding to the maximum eigenvalue of the pairwise comparison matrix (Belton and Stewart, 2002; Saaty, 2005).

A vector of relative preferences is determined by comparing the alternatives with respect to each of the criteria at next level of the hierarchy. The next step is to compare all criteria, which share the same parent using the same pairwise comparison procedure, deserving a vector indicating the relative contribution of the criteria to the parent. The DM is asked to compare the criteria and working aggregates the judgments upwards from the bottom of the hierarchy (Belton and Stewart, 2002; Saaty, 2005).

3.3.5 Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH)

MACBETH is an approach for multicriteria value measurement (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986) that only requires qualitative judgments about differences of value to help an individual or a group quantifying the relative attractiveness of alternatives. Note that every MCDA methodology based on the additive value model to be quantitatively meaningful should be unaffected by transformations of quantities involved, as it the case with MACBETH. A necessary condition is that each value scale should be unique up to a positive affine transformation, as it is with a value difference scale (Goodwin and Wright, 2004; Bana e Costa, De Corte, and Vansnick, 2012).

Suppose that *J* is a DM, which make the judgments about attractiveness of the elements, in a set of stimulus *S*. *S* must be finite and has a structure *P*, which models the relative attractiveness of the elements in *S*, only for *J*.

So, $x, y \in S, x P y$ then *J* judge *x* more attractive than *y*. Based on this, MACBETH is an interactive approach to help *J* to determine, in each element *x* of *S*, a real number *v(x)* that quantifies the attractiveness of *x*, creating a scale *v* about *S*. MACBETH asks to *J*, for each pair of stimulus *x* and *y* of *S*, that judge the difference of attractiveness (Bana e Costa, De Corte, and Vansnick, 2012).

Unfortunately, the construction of an interval value scale is usually a difficult task. In MACBETH, when a DM compares two or more things, firstly by giving a judgment to their relative attractiveness and secondly, if the two elements are not deemed to be equally attractive, by expressing a qualitative judgment about the difference of attractiveness between the most attractive of the two elements and the others. Moreover, to ease the judgmental process, MACBETH have six semantic categories of difference of attractiveness: very weak, weak, moderate, strong, very strong, or extreme, or a succession of these (Goodwin and Wright, 2004; Bana e Costa, De Corte, and Vansnick, 2012).

Then, MACBETH verifies if exists a numeric scale about *S* that satisfies the following rules:

- Rule 1 : $" x , y \in S: \; v(x) > v(y) \Leftrightarrow x$ is more attractive than *y*;
- Rule 2 : $" k , k' \in \{1,2,3,4,5,6\}, "(x , y) \in Ck, " (w , z) \in Ck': k > k'+1 =>$

  $v(x) - v(y) > v(w) - v(z)$

Afterwards, when the judgment matrix is consistent, MACBETH determines a scale for these judgments by solving a linear programming problem (Bana e Costa, De Corte, and Vansnick, 2005a; Bana e Costa, De Corte, and Vansnick, 2012).

As the answers are entered into the MACBETH decision support system (Bana e Costa, De Corte, and Vansnick, 2005a) it automatically verifies their consistency and subsequently generates a numerical scale by solving a linear programming problem that is representative of the DM's judgments. Through a similar process it permits the generation of weighting scales for criteria (Bana e Costa and Vansnick, 1999; Bana e Costa, De Corte, and Vansnick, 2011, 2012).

Afterwards, obtaining overall value scores for all the alternatives it is then possible to make sensitivity and robustness analyses, which will allow elaborating an informed recommendation.

Finally, we want to highlight that the MACBETH approach preconizes Keeney's approach (Keeney, 1992) 'value-focused thinking' philosophy: first discover the DM's values, identify the criteria that are relevant for the problem, and only then create alternatives that might satisfy these criteria. 'Value-focused thinking' is worth considering for important decisions where there is a need to think deeply about what are fundamental values of an individual or organization (Goodwin and Wright, 2004), instead of the traditional, less creative, 'alternative-focused thinking' philosophy.

3.3.6 Discussion

The Outranking methods differ from the value functions approaches in that there is no underlying aggregative value function, so they do not produce an overall preference scale for each of the alternatives. The output of an analysis is not a value for each alternative, but an outranking relation on the set of alternatives. An alternative *a* outranks alternative *b* if taking into account of all available information regarding the problem and the DM's preferences, there is a strong enough argument to support a conclusion that *a* is at least as good as *b* and no strong argument to the contrary.

The MAVT and AHP apply the additive aggregation model (2) where the weights define tradeoffs between performances on the different criteria. In the case of MAVT, all partial value functions are scaled using two fixed reference levels on the performance (e.g. "best" and "worst" performances, or "good" and "neutral" performances) to which are assigned two arbitrary numbers (e.g. 100 and 0, respectively to the upper reference level and to the lower reference level). AHP generates global scores to represent the overall preference upon the alternatives, which is a wanted feature. However, there are known issues regarding this method concerning the appropriateness of the conversion from the semantic to the numeric scale used in AHP. The partial scores are sum to 1, and imply that the meaning of weight is the relative worth, not of the swings, but of the total score on different criteria, thus it can only be defined by all the alternatives under consideration. Furthermore, AHP assumes, which all comparisons can be made on a ratio scale that implies the existence of a natural zero. This means that if comparing alternatives we could state our preference for one over the other as a ratio, which makes good sense only to compare countable things (Bana e Costa and Vansnick, 2008; Dyer, 1990).

A MACBETH advantage over the other methods for multicriteria value measurement is that it only requires qualitative judgments to score alternatives and to weight criteria, instead of requiring quantitative judgments. Furthermore, its decision support system M-MACBETH (Bana e Costa, De Corte, and Vansnick, 2005a) that implements MACBETH, is able to compute the overall values of the alternatives by applying the additive model, and to make extensive sensitivity and robustness analysis. Finally, MACBETH is also very used when organizations have to think deeply about their decisions. These reasons justified the adoption of the MACBETH approach as our evaluation method.

3.4 Objectives

This section covers the other half part of **define the objectives of a solution** phase of DSRM, in which we infer the objectives of a solution from the problem definition and knowledge of what is possible and feasible.

Looking back to the problem we stated in **Chapter 2** that the main issue is that there are no consensual criteria to evaluate and compare Cloud services, and following the line of thought also concluded that there are no guidelines and structure form to make a complete assessment to Cloud services.

The main objective of this proposal is to present some mechanism that enables all organizations, from big to SMEs organizations, to address this problem. The solution can be divided into two parts:

- Creation of a set of evaluation criteria for Cloud services based on the literature and on the feedback from practitioners in Cloud Computing.
- Using this set of criteria together with the MACBETH multicriteria decision analysis approach to value measurement to create a method that can be applied to several organizations.

What is intended to achieve with the sum of both parts is a method that allows organizations to evaluate and compare Cloud services offers, which is more affordable, understandable, easier to use and less dependent on external consultancy allowing for all organizations, mainly SMEs, to conduct a most of this process.

There are some other more specific objectives that the proposal must be met. So, the proposal should:

- Clarify DM doubts and fears about what Cloud solution is better to their business and strategy.
- Be general for all Cloud services and for all kind of organizations.
- Be easier to apply upon the moment of the application on an organization.
- Provide results that are easily understandable by the DM.
- Be from, the moment an organization decides to purchase a Cloud service, to the moment where the results of the services comparison are obtained, more affordable than actual methods.
- Be less expensive than the currently frameworks and methodologies, allowing cost reduction and improving efficacy and effectiveness of Cloud contracts.

These objectives allowed us to set a roadmap for the design and development of the proposed method, which was decided based on the problem, the necessities we identified and some principles we believe are essential to any research like abstraction, benefit, justification, and originality.

Chapter 4

Proposal

This chapter corresponds to the **design and development** phase of DSRM, in which we present a DSRM artifact. This artifact aims at solving the identified problem (Chapter 2) and achieves the objectives defined in previous chapters.

This chapter is divided in three sections, which correspond to the sequential process that composes our proposal. In the first step, we need to structure the model, which implies selecting evaluation criteria and their descriptors of performance (Section 4.1). The second step is the evaluation, in which a value function is built for each criterion and criteria weights are assessed (Section 4.2). Finally, in the third step the results of the model are analyzed and sensitivity and robustness analyses are performed before giving a selection recommendation (Section 4.3).

To address the problem stated (Chapter 2), we need several different criteria in order to evaluate the alternatives. Thereunto, in our proposal, we use the M-MACBETH software (Bana e Costa, De Corte, and Vansnick, 2005), which permits the evaluation and comparison of the alternatives using the criteria previously approved by the DM.

M-MACBETH can generate scores for each alternative in each criterion, weights for the criteria, and quantify the relative attractiveness between the alternatives using the DM’s judgments. In addition, also allows performing extensive sensitivity and robustness analyses.

For our proposal, we define Cloud service as in (Hogan et al., 2011). In short, is any resource that is provided at distance and is hosted by a provider that made available to the clients over a network. These resources are related with all service models (SaaS, PaaS, and IaaS).

Hereupon, our method consists in three main steps and is summarized in Figure 8.

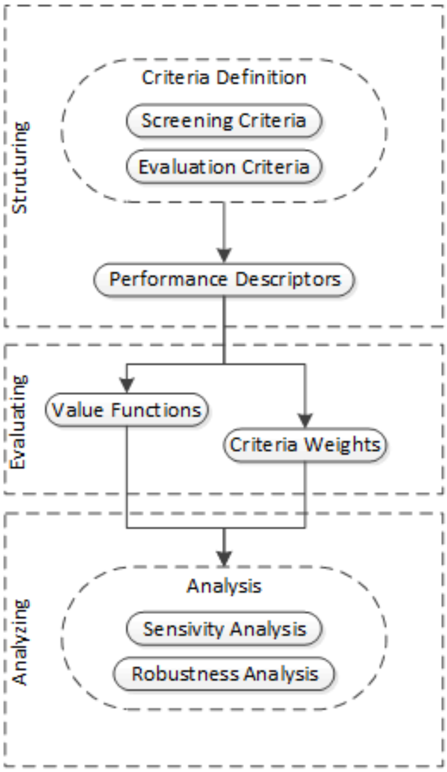

Figure 8. Proposed steps (Adapted from: (Bana e Costa, De Corte, and Vansnick, 2008))

Despite the fact that a DM can propose other criteria that are essential for his organization, the first part of the Figure 8, which mainly consists in evaluation criteria, provided as a criteria catalog. Note that “screening criteria” are non-compensatory criteria that establish thresholds for the acceptance of alternatives (e.g. to exclude from the evaluation alternative that have an unacceptable cost). However, the second and third parts have to be done for each problem because different DMs has different opinions and the results are mainly based in their opinions.

4.1 Structuring the Model

The decision-making process begins by structuring the problem, which consists in identifying the issues of concern for the DM. The DM fundamental points of view should be taken as evaluation criteria. Each criterion should be associated with a qualitative or quantitative descriptor of performance, to measure the extent to which the criterion can be satisfied.

As we saw (Chapter 3), actual approaches have some flaws liable of improvements mainly related with redundant, less objective, and missing criteria. Our objective in the first step was create a set of criteria through mapping SMI framework, because the intend of SMI is definition of a set of KPI to provide a method for measuring and comparing any business service, with other approaches that also proposed some criteria to measure Cloud services (Section 3.2.7). Through this mapping we could conclude what are the mainly criteria to evaluate Cloud services. Using these results we construct the first draft of our evaluation criteria.

Despite this, this draft based in literature needed the organization feedback. To complete our critical and improvement to SMI, we made 12 interviews (Section 6.5.2) with Cloud clients, providers, and experts in well-chosen organizations to obtain feedback from the industry so that we could identify the criteria that really matter to evaluate the pros and cons of Cloud services in real organizations.

Based on the previous work, our purpose was creating a set of evaluation criteria more qualitative than quantitative and for this we proposed criteria with different ranges but which we consider critical. Some of the initial proposed criteria were redundant or preferential dependent on other criteria. To eliminate these issues some initial criteria were merged or excluded from the criteria catalog. Figure 9 shows the 19 evaluation criteria obtained in this process, which were organized in six groups.

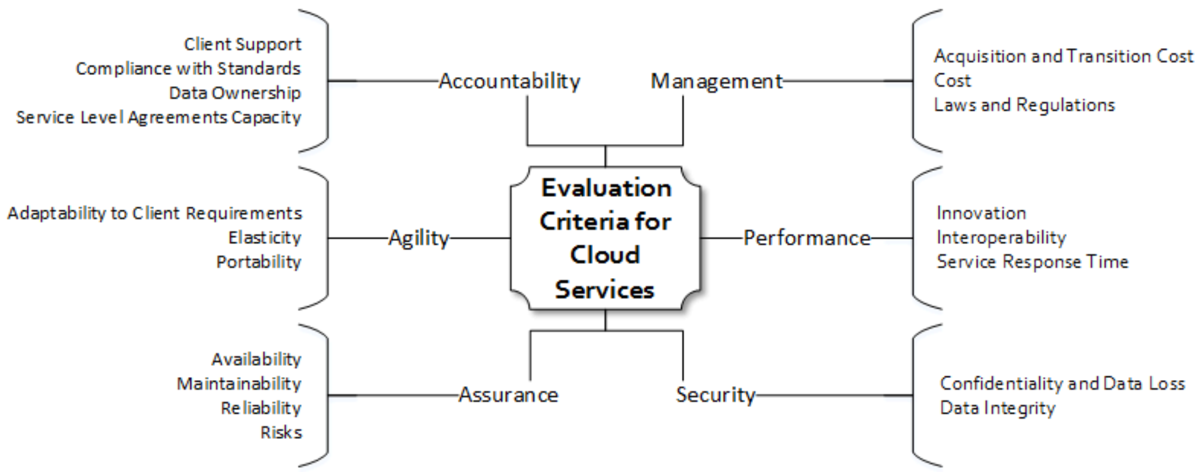

Figure 9. Evaluation Criteria Hierarchy

The groups of criteria are defined as:

- **Accountability**: measures characteristics related to service provider organization, independently of the service provided.
- **Agility**: measures the impact of service in terms of direction, change, strategy, or tactics with minimal disruption in organization strategy.
- **Assurance**: measures how likely the service will be available as specified.
- **Management**: measures the business strategic decisions of service option in market related with costs and regulations.
- **Performance**: measures the service expectations and how they really work.
- **Security**: measures the protection of service, data privacy, and confidentiality.

The meaning of each criterion is explained in Table 6.

Accountability	
Client Support	The extent to which the service provider is capable to clearly define the methods to support the client and communicate and report failures
Compliance with Standards	The extent to which the service provider is certificated to follow the standards, processes, and polices committed to
Data Ownership	The extent to which the client has ownership over data
Service Level Agreements Capacity	The extent to which the service provider is able to negotiate the SLAs established
Agility	
Adaptability to Client Requirements	The extent to which the service provider is able to adjust the service to changes in client requirements
Elasticity	The extent to which the service is capable of adjust resources to meet client's demand
Portability	The extent to which the service can be ported to another service provider
Assurance	
Availability	The extent to which the client can use the service without interruptions
Maintainability	The extent to which the service provider is able to maintain the service without affecting the service up time
Reliability	The extent to which the service is capable to operate without failure under unfavorable condition
Risks	The extent to which the service provider can identify and treat risks related with IT to improve client confidence in risk mitigation and contingency plan execution
Management	
Acquisition and Transition Cost	Cost to acquire the rights and ability to use the service and to move the service to a new service provider
Cost	Cost to operate the service
Laws and Regulations	The extent to which the service is subject to a specific legislation which restricts or not the service adoption and gives some benefits or disadvantages
Performance	
Innovation	The extent to which the service is able to include innovative features without compromising the service and make updates to reach new technologies and IT evolution at any moment
Interoperability	The extent to which the service is able to interact with other services
Service Response Time	The extent to which the service is able to provide a fast response to a client request
Security	
Confidentiality and Data Loss	The extent to which the service provider restricts the information sharing to authorized people and promptly detects and reports any failure of these protections
Data Integrity	The extent to which the service provider maintains data stored and backups in its correct form so that clients may be confident that it is accurate and valid

Table 6. Evaluation Criteria

According to the analysis done these criteria are enough and necessary to evaluate most Cloud services, but it is always possible to add or delete criteria to meet some specific needs. This set of criteria meets the needed characteristics, which are (Keeney, 1992): **essential**, **controllable**, **complete**, **measurable**, **operational**, **decomposable**, **no redundant**, **concise**, and **intelligible**.

We associated a descriptor of performance to each criterion. A descriptor of performance is a ranked set of performance levels (quantitative and qualitative) that aims to operationalize the criterion, i.e., to measure the extent to which the alternatives satisfy the criteria. The descriptors of performance also contribute to the intelligibility of criteria, clarifying their meaning, and ensure the respect of the fundamental requirement of isolability.

A descriptor of performance can be classified on three axes: **natural**, if directly measures the performances of the alternatives; **proxy**, if measures indirectly the performances of the alternatives; or **constructed**, if the performances of the alternatives are measured by combination of characteristics: **quantitative**, **qualitative**, and **discrete**. A needed requirement for a multicriteria value measurement (Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986) is the definition of two reference levels in the performance scale of each criterion. We opted to define a “good” level, which is a very attractive level, and a “neutral” level, which is a performance level neither positive nor negative, in each criterion.

The descriptors of performance associated with the evaluation criteria for Cloud services, as well as two reference levels on each criterion, are presented in Tables 7 to 25. There we can find **qualitative** descriptors of performance, like the descriptor for criteria “Client Support**”** present in Table 7, which is a discrete descriptor, and **quantitative** descriptors of performance, like the descriptor for criterion “Data Ownership”, which is a natural and continuous descriptor, present in Table 9.

Accountability

There are four criteria in the accountability group, with the descriptors of performance presented in Tables 7, 8, 9 and 10.

Client Support	
Performance Levels	
The service provider has defined methods to support the client and is able to communicate and report service failures	L1
The service provider has defined methods to support the client but is not able to communicate and report service failures	L2 = Good
The service provider has no defined methods to support the client and is able to communicate and report service failures	L3 = Neutral
The service provider has no defined methods to support the client and is not able to communicate and report service failures	L4

Table 7. Descriptor of performance for the “Client Support” criterion

Compliance with Standards	
Performance Levels	
The service provider follows all the standards, processes, and policies	L1 = Good
The service provider follows the most important standards, processes, and policies	L2
The service provider follows some of the standards, processes, and policies	L3 = Neutral
The service provider does not follow any standards, processes, and policies	L4

Table 8. Descriptor of performance for the “Compliance with Standards” criterion

Data Ownership	
Descriptor: Level of rights (in %)	
Reference Levels	
90%	Good
50%	Neutral

Table 9. Descriptor of performance for the "Data Ownership" criterion

Service Level Agreements Capacity	
Performance Levels	
The service provider is able to negotiate all terms of the SLAs	L1 = Good
The service provider is able to negotiate some terms of the SLAs	L2 = Neutral
The service provider is not able to negotiate SLAs	L3

Table 10. Descriptor of performance for the "Service Level Agreements Capacity" criterion

Agility

In agility group there are three criteria with the following descriptors of performance presented in Tables 11, 12, and 13.

Adaptability to Client Requirements	
Performance Levels	
The service provider is able to include all client requirements in the service	L1
The service provider is able to include core or important client requirements in the service	L2 = Good
The service provider is able to include client requirements if they not require any modification in service (e.g. interface modifications	L3 = Neutral
The service provider is not able to include client requirements in the service	L4

Table 11. Descriptor of performance for the "Adaptability to Client Requirements" criterion

Elasticity	
Descriptor: Level of added resources (in %)	
Reference Levels	
100%	Good
50%	Neutral

Table 12. Descriptor of performance for the "Elasticity" criterion

Portability	
Performance Levels	
The service can be ported to other service provider without disruption	L1 = Good
The service can be ported to other service provider but cannot move all the data	L2 = Neutral
The service cannot be ported to other service provider	L3

Table 13. Descriptor of performance for the "Portability" criterion

Assurance

In assurance group there are four criteria with the following descriptors of performance presented in Tables 14, 15, 16, and 17.

Availability	
Descriptor: Amount of time without interruptions per day (in %)	
Reference Levels	
99%	Good
97%	Neutral

Table 14. Descriptor of performance for the "Availability" criterion

Maintainability	
Performance Levels	
The service maintenance does not affect the service up time	L1 = Good
The service maintenance stops the service	L2 = Neutral

Table 15. Descriptor of performance for the "Maintainability" criterion

Reliability	
Performance Levels	
The service can operate without failures under worst unfavorable conditions (e.g. natural disasters)	L1
The service can operate without failures under common unfavorable conditions (e.g. power failure)	L2 = Good
The service can operate under unfavorable conditions but some components may not work	L3 = Neutral
The service cannot operate under unfavorable conditions	L4

Table 16. Descriptor of performance for the "Reliability" criterion

Risks	
Performance Levels	
The service provider has an effective risk identification and treatment, and contingency plan	L1 = Good
The service provider has an effective risk identification and treatment but no contingency plan	L2 = Neutral
The service provider has no risk identification, no risk treatment, and no contingency plan	L3

Table 17. Descriptor of performance for the "Risks" criterion

Management

In management group there are three criteria with the following descriptors of performance presented in Tables 18, 19, and 20.

Acquisition and Transition Cost	
Descriptor: Amount (in €)	
Reference Levels	
€0	Good
€1000	Neutral

Table 18. Descriptor of performance for the “Acquisition and Transition Cost” criterion

Cost	
Descriptor: Amount (in € per user/month)	
Reference Levels	
€10	Good
€20	Neutral

Table 19. Descriptor of performance for the “Cost” criterion

Laws and Regulations	
Performance Levels	
The service is subject to laws and regulations to protect clients against all kind of irregularities in the client’s country	L1
The service is subject to laws and regulations to protect clients against all kind of irregularities in the provider’s country	L2 = Good
The service is subject to laws and regulations only to protect clients against data losses in the client’s country	L3
The service is subject to laws and regulations only to protect clients against data losses in the provider’s country	L4 = Neutral
The service is subject to no laws and regulations	L5

Table 20. Descriptor of performance for the “Laws and Regulations” criterion

Performance

In performance group there are three criteria with the following descriptors of performance presented in Tables 21, 22, and 23.

Innovation	
Performance Levels	
The service is able to make all updates to new technologies and to include innovative features automatically	L1 = Good
The service is able to make updates to new technologies but not automatically	L2 = Neutral
The service is not able to make innovation and updates	L3

Table 21. Descriptor of performance for the “Innovation” criterion

Interoperability	
Performance Levels	
The service is able to interact with other services	L1 = Good
The service is able to interact with other services from the same service provider	L2 = Neutral
The service is not able to interact with other services	L3

Table 22. Descriptor of performance for the “Interoperability” criterion

Service Response Time	
Descriptor: Time between service request and response (in seconds)	
Reference Levels	
0,5 s	Good
2 s	Neutral

Table 23. Descriptor of performance for the "Service Response Time" criterion

Security

In security group there are two criteria with the following descriptors of performance presented in Tables 24 and 25.

Confidentiality and Data Loss	
Performance Levels	
The information is restricted to authorized people and a failure is promptly detected and reported	L1
The information is restricted to authorized people and a failure is promptly detected but no reported	L2 = Good
The information is restricted to authorized people but there is no detection and reported failures	L3 = Neutral
The information can be accessed by unauthorized people and there are no failure detection and report	L4

Table 24. Descriptor of performance for the "Confidentiality and Data Loss" criterion

Data Integrity	
Performance Levels	
The data stored is accurate and valid and backups are updated to the second	L1 = Good
The data stored is accurate and valid and backups are updated weekly	L2
The data stored is accurate and valid and backups are updated monthly	L3 = Neutral
The data stored is accurate but no backups	L4
The data stored is not accurate and valid	L5

Table 25. Descriptor of performance for the "Data Integrity" criterion

We consider that these criteria with their descriptors of performance can be used as a catalog for the all Cloud service evaluations and comparisons. Although some changes to that template must meet all requirements set forth above.

4.2 Evaluating the Alternatives

In this step, the preferential judgments of a DM will be used to build a value function for each criterion and to weight the criteria. A value function serve to transform performances (factual data) in value scores on a criterion, which requires (subjective) judgments of value from the DM upon differences of performances on that criterion. The weights allow transforming partial value scores of the several criteria in commensurate value scores. Assessing weights requires the DM to judge trade-offs between criteria, using the “neutral-good” swings defined upon each criterion performance scale.

To build a value function for a criterion the two reference performance levels, **neutral** and **good**, are assigned the scores 0 and 100, respectively, establishing two anchors on the value scale. The DM are then asked to judge the differences in attractiveness between each two levels of performances on the criterion by choosing one, or more, of the MACBETH semantic categories: **very weak**, **weak**, **moderate**, **strong**, **very strong**, and **extreme**. For each criterion, the process are initiated by asking the difference of attractiveness of changing from the **neutral** performance level to the **good** performance level and followed by asking the difference between each two of the other performance levels.

The M-MACBETH decision support system generates a numerical value scale by solving a linear programming problem based on the set of qualitative judgments inputted in the matrix of judgments. The generated scale is then subjected to an analysis in terms of the proportions of the resulting scale intervals.

The next step is weighting the criteria. The DM is first asked to rank the “neutral-good” swings by their overall attractiveness. The decision analyst started by asking: “if you could choose just one criterion to change from a neutral performance to a good performance which criterion would you choose?”. The question procedure continues until the final ranking of “neutral-good” swings is achieved.

The process continues with a new question. The DM is now asked to judge the overall attractiveness of each “neutral-good” swing. Subsequently, the DM is asked to pairwise compare the most attractive swing to the second most attractive swing. The pairwise comparison continues between the most attractive swing and each one of other swings until filling the MACBETH weighting matrix. This is enough for MACBETH to create the weighting scale by linear programming. Finally, the DM has to check the resulting weights in order to validate them.

4.3 Analyzing the Results

In the final step of our proposed method, the performance of the alternatives on the criteria are inputted in M-MACBETH. The software then transforms these performances into value scores, using the value functions previously built, and calculates the overall value score for each alternative by weighted summation of its partial value scores (using the additive aggregation model presented in equation 2), allowing ranking the alternatives by overall attractiveness.

Sensitivity and robustness analysis are then made in order to understand how the results can be affected by “small” changes in the model parameters.

In the sensitivity analysis we verify what changes may be produced in the results obtained by the evaluation model when a criterion weights varies, within the limits allowed by the DM's weighting judgments. This analysis is made by the M-MACBETH and is very useful to show the DM the consequences of changing weights.

The robustness analysis consists in comparing the pairs of alternatives considering their global attractiveness in uncertain conditions. It can be used to examine the implication into the global results when we simultaneous vary all or some of the parameters of the model. This analysis allows detecting the existence of dominance, in classical sense, between alternatives or additive dominance.

An alternative *a* **dominates**, in the classical sense, an alternative *b* if is not worse than *b* in any criterion and *a* is better than *b* in at least on criterion. This kind of dominance is not influenced by the weights of the criteria and it is represented in M-MACBETH by red triangle, meaning that the alternative in row dominates the alternative in column.

We say that alternative *a* **additively dominates** alternative *b* if *a* is always globally more attractive than *b* through the use of an additive aggregation model under a set of information constraints (Bana e Costa, De Corte, and Vansnick, 2005a). For example, when the criteria weights $w = [w_1, \ldots, w_n]$ are subject to variations within the limits $\left(\overline{w}_1 \leq w_1 \leq \underline{w}_1, \ldots, \overline{w}_n \leq w_n \leq \underline{w}_n\right)$, M-MACBETH solves two linear programing problems (max *P* and min *P*) with the following structure:

$$P = \sum_{i=1}^{n} w_i \left[v_i(a) - v_i(b) \right] \qquad (3)$$

subject to:

$$\begin{aligned} &\overline{w}_1 \leq w_1 \leq \underline{w}_1 \\ &\qquad \vdots \\ &\overline{w}_n \leq w_n \leq \underline{w}_n \\ &\sum_{i=1}^{n} w_i = 1 \\ &w_i \geq 0 \quad (i = 1, \ldots, n) \end{aligned}$$

We can conclude that:

- if min P > 0: the alternative *a* **additively dominates** alternative *b*;
- or if max P < 0: the alternative *b* **additively dominates** alternative *a*;
- otherwise, there is no dominance relationship between *a* and *b*.

The symbol that represents additive dominance in M-MACBETH is the green cross, meaning that the alternative in row additively dominates the alternative in column.

In addition we have three "conditions" for the robustness analysis, either in the **local information**, regarding the criteria values scales, or the **global information**, regarding the criteria weights. **Ordinal**, considers the ranking of the options within each criteria, or the ranking of the criteria "neutral-good" swings. **MACBETH**, considers the judgments inputted in the M-MACBETH matrices of judgments, either for producing value scales or criteria weights. Finally, **cardinal**, uses the specific criteria value scales, or the criteria weights scale, validated by the DM

Having these analysis completed a recommendation of the alternative that should be selected can then be made.

A step-by-step application of the proposal is shown in **Chapter 5**.

Chapter 5

Demonstration

This chapter corresponds to the **demonstration** phase of DSRM, which consists in demonstrate the artifact to solve one or more instances of the problem.

The main objective of our proposal is to construct a method to evaluate and compare Cloud services. We choose a City Council in Portugal because its CIO had doubts concerning the selection of a Cloud service to buy. The situation of this Portuguese City Council meets our research problem for which there is not a valid solution.

The DM in the field study here reported was the CIO of the City Council. The author of his dissertation acted as decision analyst in the decision-aid process.

Due to the advantages of Cloud Computing, the DM decided to put some services, like e-mail and productivity, in Cloud. Thereunto, we identified the two best services that covers the DM needs. The services alternatives were **Google Apps** and **Microsoft Office 365**.

Google Apps characterized as a Cloud-based productivity suit that helps you connect and get work done from anywhere on any device. In addition, characterized as simple to setup, use, and manage, allowing you to work smarter and focus on what really matters. On the other and, **Microsoft Office 365** has the slogan that gives you virtually anywhere access to the Microsoft Office tools you know and rely on, plus business-class IT services that are easy to administer. It is your complete Office in the Cloud.

5.1 Structuring the Model

This step begins with meetings with the DM in order to define the screening criteria and the evaluation criteria.

We made an exhaustive research related with the evaluation criteria for Cloud services and we developed a criteria catalog that can be used by all organizations. Changes can be made to the set of criteria to include in the evaluation model in order to meet specific organization needs (Section 4.1).

In our demonstration, the DM validated all criteria, accepting them as the essential criteria to his problem. However, the inclusion of criterion “Acquisition and Transaction Cost” raised some doubts because in this kind of Cloud services there is a month for service experimentation without any acquisition cost. The value true in Figure 10 shows the selected evaluation criteria grouped in six families.

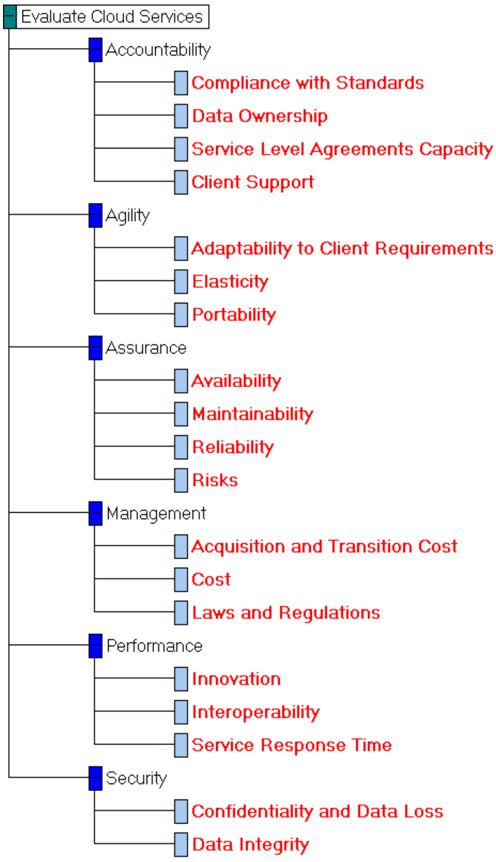

Figure 10. Value Tree

5.2 Evaluating the Alternatives

To evaluate the alternatives in consideration, **Google Apps** and **Microsoft Office 365**, we asked the DM to validate the descriptors of performance and reference levels (Section 4.1), which after a rigorous checking were accepted by the DM.

Afterwards, we asked to DM to judge the differences in attractiveness between the levels of performance within each criterion using the MACBETH approach. In general, the DM did not have any difficulty to make these judgments. Then, M-MACBETH software generated a value scale for each criterion based on the judgments inputted in the matrices of judgments. Figures 11 and 12 show the value scales and the MACBETH judgments matrices for the criteria “Availability” and “Data Integrity”, which respectively have a **quantitative** descriptor of performance and a **qualitative** descriptor.

Figure 11, which is related to a **quantitative** criterion, presents the DM’s judgments matrix for the criterion “Availability”, where we can see that, for example, the difference in attractiveness between 100% amount of time without interruptions per day and 99% was judged “weak”, whereas the differences between 99% and 98%, 98% and 97%, 97% and 96% were deemed “moderate”, which means that the DM values has less difference between 100% and 99% than the other mentioned differences. The resulting value scale depicts such reasoning, because the difference of value between 100% and 99% is 33 value scale, whereas the other mentioned differences resulted in 50 value united. We can also observe that this criterion has an increasing continuous value function: when the percentage of available time per day increases the value for the “Availability” criterion also increases.

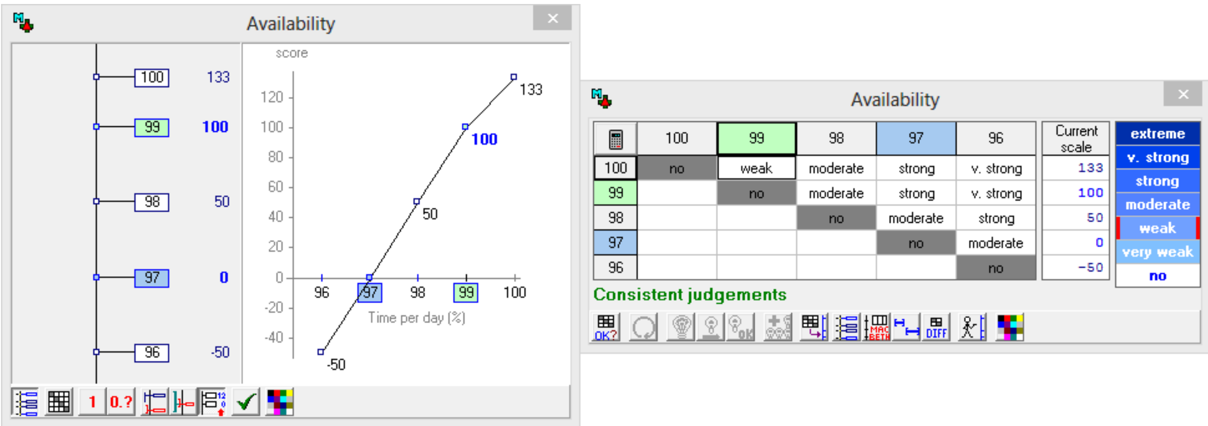

Figure 11. MACBETH judgments matrix and value function for the “Availability” criterion

Figure 12 presents the DM’s judgments matrix for the criterion “Data Integrity” which is a qualitative criterion with a discrete value function. The performance levels for this criterion are present above in Table 25. Analyzing the value scale of this criteria, we can conclude, for example, that the difference of value between “L1” (the data stored is accurate and valid and backups are updated to the second) and “L2” (the data stored is accurate and valid and backups are updated weekly) is worth the same for the DM as the difference of value between “L2” and the neutral level “L3” (the data stored is accurate and valid and backups are updated monthly), i.e., 50 value unites, but the difference in value between “L3” and “L4” (the data stored is accurate but no backups) is worth for the DM 1.5 times the value of the difference between “L1” and “L2” or “L2” and “L3”.

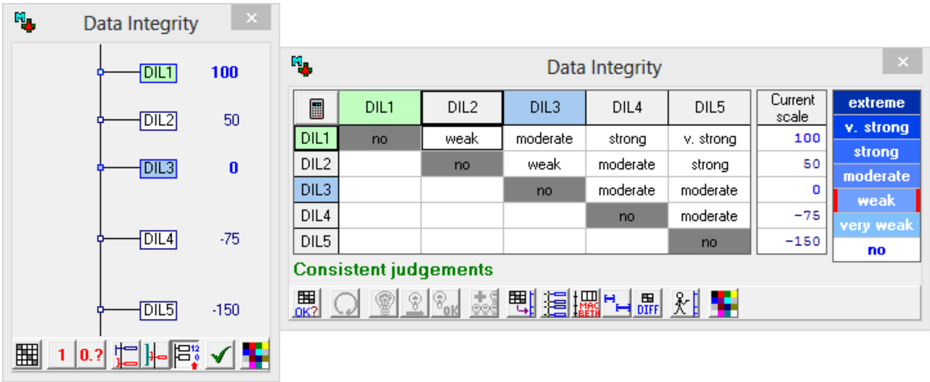

Figure 12. MACBETH judgment matrix and scale for the “Data Integrity” criterion

The MACBETH matrices of judgments and value functions for the other criteria can be seen in **Appendix B – Judgments Matrices**.

To finalize the evaluation model we also had to assess the criteria weights. For that purpose the DM was asked to: *(i)* sort the criteria by decreasing order of importance of their “neutral-good” swings; and *(ii)* fill in the weighting judgment matrix. Figure 13 shows the weighting references, and also shows the window that supports our question procedure.

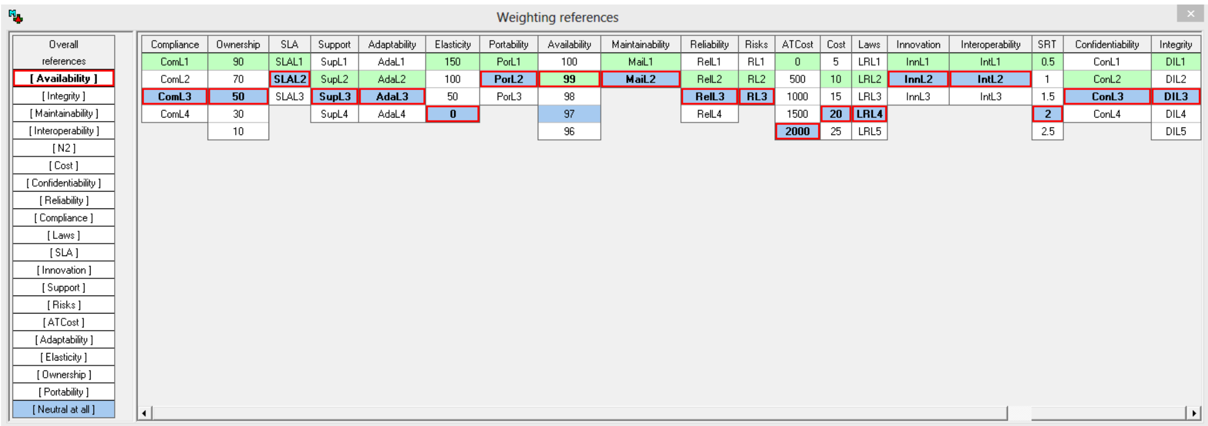

Figure 13. Weighting references

In Figure 16, each criterion name within brackets (in the “Overall references” column) represents a fictitious alternative that has a good performance in the named criterion and neutral performances in the remaining criteria. We have 19 criteria so we must have 19 fictitious alternatives built in the same manner plus the fictitious alternative [Neutral at al.], which has neutral performances in all the criteria. For example, we can see in Figure 13 that the fictitious alternative [Availability] has neutral performances in all the criteria with the exception of the criterion “Availability”, where it has a good performance (99%). The fictitious alternatives are already present sorted by decreasing overall attractiveness in the “Overall references” column.

Making pairwise judgments of difference of attractiveness with 20 fictitious alternatives would be a very demanding task for any DM. So we asked the DM to make comparisons using only the criteria within each group of criteria, on group at a time, and in the end we asked him to make pairwise comparisons with a subset of criteria that included one criterion for each group (we have chosen for this comparison the criteria with the highest weight in each group). Having this done we made some extra calculations to globally scale the weights to obtain the final weights that are presented in the histogram of Figure 14.

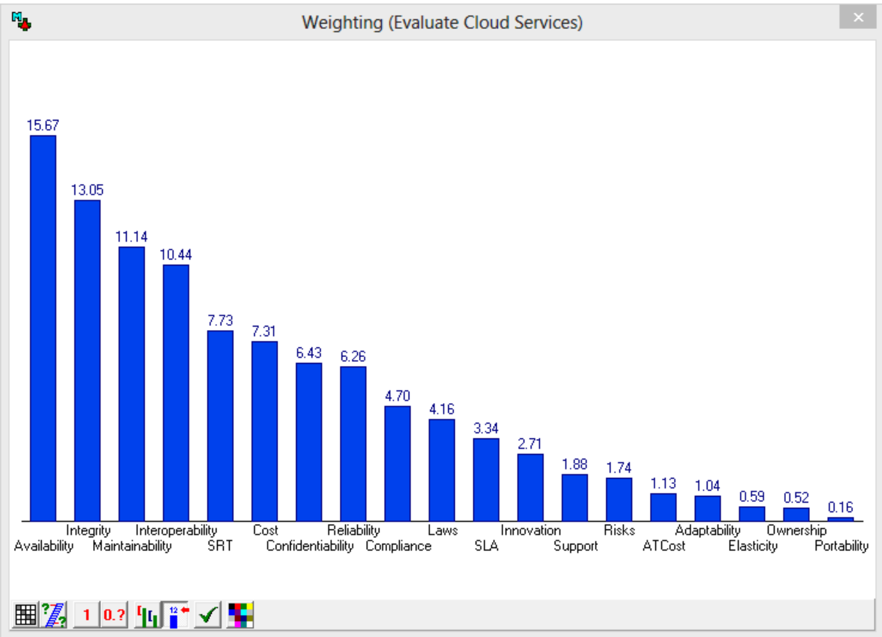

Figure 14. Histogram for weighting scale

Next, the DM validated the proposed weighting scale by answering some questions. We asked, for example, if the “neutral-good” swing on the criterion “Data Integrity” is worth two times the “neutral-good” swing on criterion “Confidentiality and Data Loss” which the DM agreed. Note that the weights of these criteria are approximately 13% and 6.5%, respectively.

5.3 Analyzing the Results

Figure 15 shows the performances of the alternatives, **Google Apps** (Google) and **Microsoft Office 365** (Microsoft) on all the evaluation criteria.

Table of performances

Options	Compliance	Ownership	SLA	Support	Adaptability	Elasticity	Portability	Availability	Maintainability	Reliability	Risks	ATCost	Cost	Laws	Innovation	Interoperability	SRT	Confidentiability	Integrity
Google	ComL1	100	SLAL3	SupL1	AdaL4	50	PorL2	99.9	MaiL1	RelL1	RL2	0	4	LRL1	InnL1	IntL1	1	ConL3	DIL1
Microsoft	ComL1	100	SLAL3	SupL1	AdaL4	50	PorL2	99.9	MaiL2	RelL2	RL1	0	12.3	LRL1	InnL2	IntL2	1	ConL1	DIL1

Figure 15. Performances of the alternatives for each criterion

Based on the performance of the alternatives and the judgments previously built with the DM, the M-MACBETH software creates a value profile for each alternative.

Figures 16 shows the performance profile for **Google Apps**.

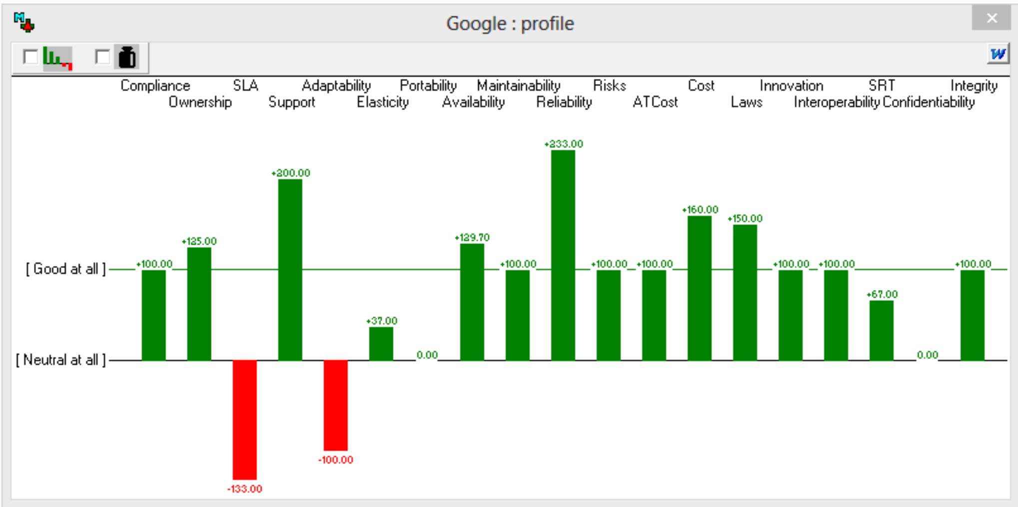

Figure 16. Performance profile of Google Apps

Figures 17 show the performance profile for **Microsoft Office 365**.

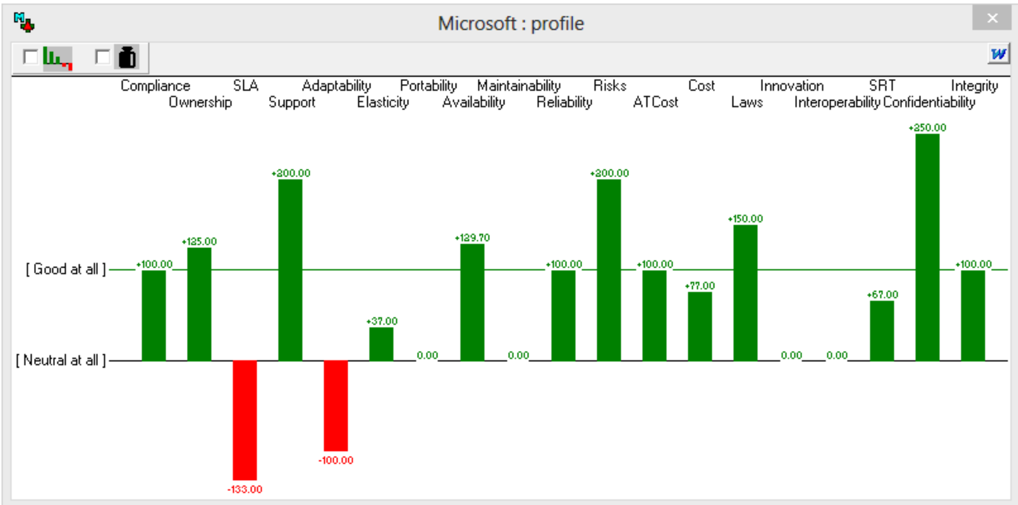

Figure 17. Performance profile of Microsoft Office 365

We can visualize in Figures 16 and 17 that both alternatives have negative scores on criteria "Service Level Agreements Capability" and "Adaptability to Client Requirements". We can also conclude that **Google Apps** has neutral score (zero) on criteria "Portability" and "Confidentiality and Data Loss", whereas **Microsoft Office 365** has neutral score on criteria "Portability", "Maintainability", "Innovation", and "Interoperability". Both alternatives have positive scores on the other remaining criteria. **Google Apps** has scores equal to or higher than the "good" reference level on 13 criteria, whereas that only happens on 10 criteria for **Microsoft Office 365**.

Using the performance of the alternatives – **Google Apps** and **Microsoft Office 365** – the value functions built for the criteria and the criteria weights assessed with the DM the M-MACBETH decision support system calculated the overall value scores for all the alternatives using the additive aggregation model (see column "Overall" in Figure 18).

Table of scores

Options	Overall	Compliance	Ownership	SLA	Support	Adaptability	Elasticity	Portability	Availability	Maintainability
Google	102.08	100.00	125.00	-133.00	200.00	-100.00	37.00	0.00	129.70	100.00
[Good at all]	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Microsoft	81.21	100.00	125.00	-133.00	200.00	-100.00	37.00	0.00	129.70	0.00
[Neutral at all]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Weights :		0.0470	0.0052	0.0334	0.0188	0.0104	0.0059	0.0016	0.1567	0.1114

Reliability	Risks	ATCost	Cost	Laws	Innovation	Interoperability	SRT	Confidentiality	Integrity
233.00	100.00	100.00	160.00	150.00	100.00	100.00	67.00	0.00	100.00
100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
100.00	200.00	100.00	77.00	150.00	0.00	0.00	67.00	250.00	100.00
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.0626	0.0174	0.0113	0.0731	0.0416	0.0271	0.1044	0.0773	0.0643	0.1305

Figure 18. Overall scores

Analyzing the results presented in Figure 18, we see that **Google Apps** ranked first with 102.08 overall value units and that **Microsoft Office 365** has 81.21 overall value units. Only **Google Apps** obtained an overall score higher than the score of the hypothetical alternative "[Good at all]" (i.e., a fictitious alternative that has a good performance in all the criteria), which shows that **Google Apps** is a very attractive alternative for the DM. **Microsoft Office 365** also is an attractive alternative, because its overall score is closer to the score of the hypothetical alternative "[Good at all]" than to the score of the hypothetical alternative "[Neutral at all]".

Figure 19 shows the differences in value between the two alternatives for each criterion.

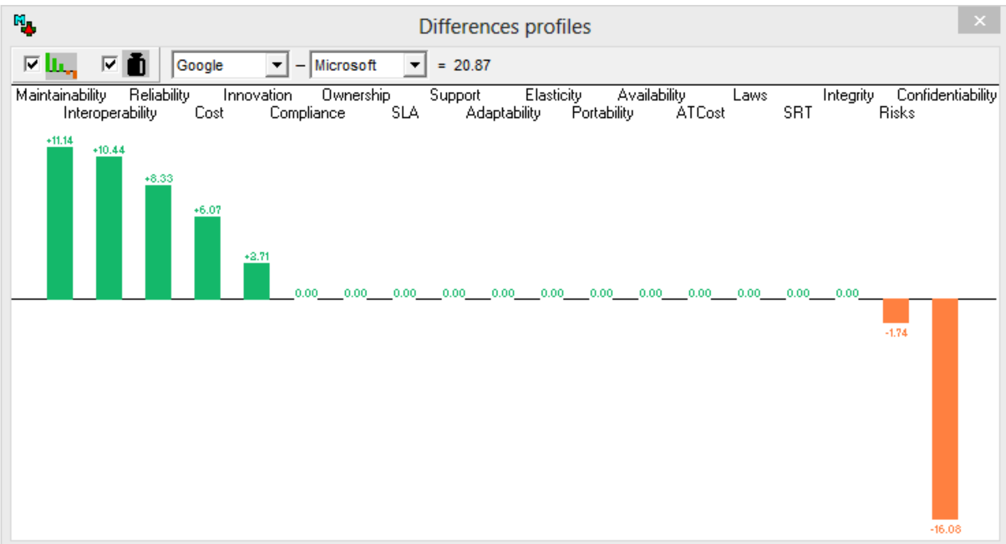

Figure 19. Difference profiles between Google Apps and Microsoft Office 365

In Figure 19, each bar shows the weighted value score difference between **Google Apps** and **Microsoft Office 365** on one criterion. According to that, the positive differences, illustrated by green bars, represent the score differences favorable to **Google Apps**, and, on the other hand, the negative differences, illustrated by red bars, represent the score differences favorable to **Microsoft Office 365**. From the graph in Figure 19 we can observe that **Google Apps** is superior to **Microsoft Office 365** in five criteria and **Google Apps** is inferior to **Microsoft Office 365** in two criteria: “Risks” and “Confidentiality and Data Loss”.

A sensitivity analysis on the weight of criterion “Risks” showed that the weight of this criterion would need to be increased from 1.74% to 18.7% to see **Microsoft Office 365** be ranked first. A similar analysis showed that the weight of criterion “Confidentiality and Data Loss” would need to be increased from 6.43% to 13.6% to see **Microsoft Office 365** as the winner alternative (see Figure 20). However, the DM did not consider plausible these changes on the weights.

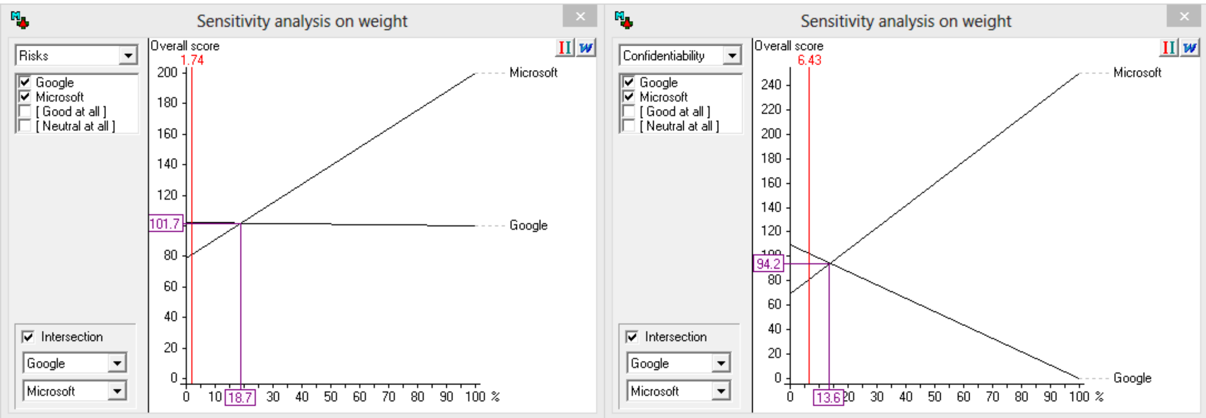

Figure 20. Sensitivity analysis for the “Risks” and “Confidentiality and Data Loss” criteria

Regarding the criterion “Compliance with Standards”, the sensitivity analysis showed that only if the weight of the criteria were 100% the **Microsoft Office 365** could catch **Google Apps** (see Figure 21).

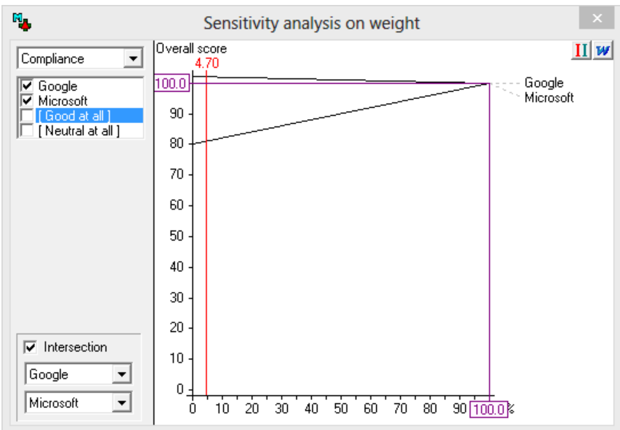

Figure 21. Sensitivity analysis for the “Compliance with Standards” criterion

Regarding to the “Maintainability” criterion it was evident (see Figure 22) that in any case **Microsoft Office 365** could be ranked first because **Google Apps** was first in all percentage of weight.

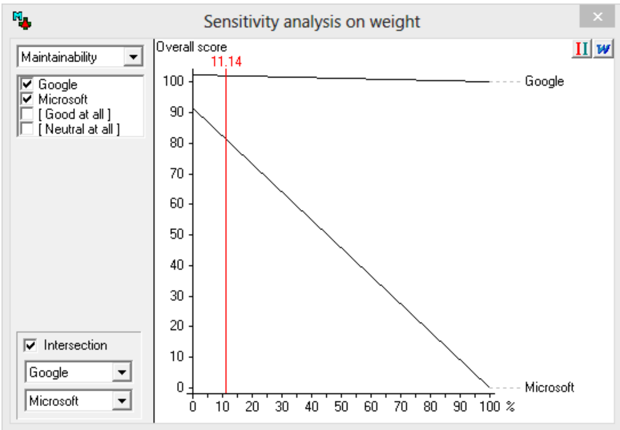

Figure 22. Sensitivity analysis for the “Maintainability” criterion

For the other criteria the previous two analyses repeated, i.e., **Google Apps** was ranked first and any variation of the weights of all criteria could not change this classification.

Finally, we tested the robustness of the results obtained with the evaluation model using the M-MACBETH robustness analysis feature described in Section 4.3.

In the first robustness analysis (shown in Figure 23) we only took into account **ordinal data** in **local** and **global information**, which allowed us to conclude that this information is insufficient to determine the best alternative.

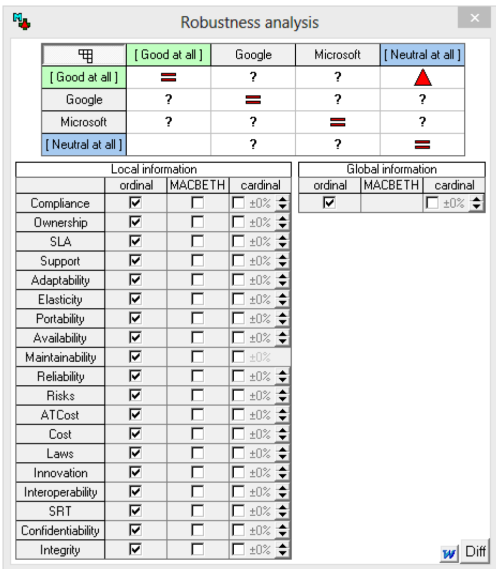

Figure 23. Robustness analysis with local and global information using ordinal scale

Then wed use the **MACBETH** judgments in **local information**, which did not change the result obtained in the previous analysis. Note that we could not use **MACBETH global information** because the criteria weights were not assessed using a single MACBETH judgments matrix, with all the criteria included, to avoid too many pairwise comparisons, as we explained in Section 5.1.3.

Figure 24 shows the robustness analysis made with M-MACBETH considering simultaneous variations ±1% on the weights of all criteria, though not allowing negative weights. This analysis revealed that **Google Apps** continues to be the most attractive alternative. A green cross in a cell in Figure 24 means that the alternative in row, **Google Apps**, additively dominates the alternative in column, **Microsoft Office 365**, i.e., within a variation of ±1% on the criteria weights **Google Apps** is always more attractive than **Microsoft Office 365**.

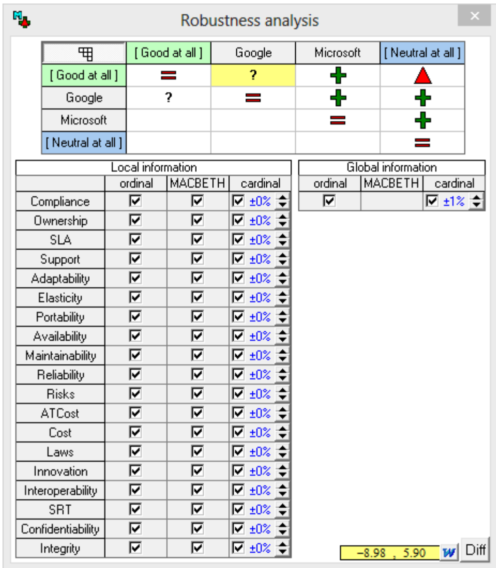

Figure 24. Robustness analysis with global (ordinal and cardinal) information with ±1% of global uncertainty

Since variations with ±1% of uncertainty on the weights may be not significant we tested the model with variations of ±3% (see Figure 25) and higher percentages. We can see in Figure 25 that with ±3% the "[Good at all]" fictitious alternative has lost its additive dominance upon **Microsoft Office 365**. But, which is more important, independently of the magnitude of the variations on the criteria weights, **Google Apps** is always more attractive than **Microsoft Office 365**, because "Google" always additively dominates "Microsoft".

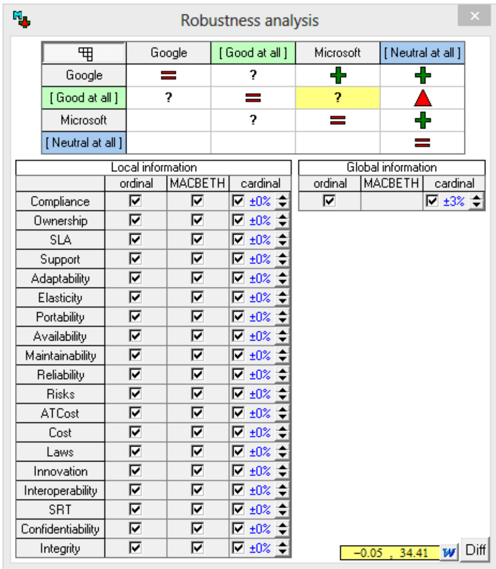

Figure 25. Robustness analysis with local (ordinal, MACBETH, and cardinal) and global (ordinal and cardinal) information with ±3% of global uncertainty

After concluding this multicriteria decision analysis process, we recommended the selection of **Google Apps,** because it was proved through our analysis that is the best alternative taking into account all the used criteria and the judgments of value expressed by the DM. In addition the sensitivity and robustness analysis showed that this is a robust selection.

5.4 Summary

This field study allowed us to test our proposal by applied them in order to resolve the research problem stated. The Portuguese City Council suffered from the same problem, as we found in literature, and our proposal helped them to overcome it.

Considering the results obtained with the evaluation model built with the DM, we recommend the selection of **Google Apps**. The DM recognized that the proposal was very useful and that the model developed and the decision analysis process were an important help to make his decision.

To summarize, we have proved that our proposed method can help a DM to make his decision, which answers our research problem.

Chapter 6

Evaluation

This chapter corresponds to the **evaluation** step of DSRM, which aims to observe and measure how well the artifact supports a solution of the problem. We are going to begin by providing a brief description of the Design Science Research Evaluation Framework that aims to offering a strategic view of the evaluation (Section 6.1). Afterwards, we need to define the criteria for evaluation. For this, we use the feedback from the interviews (Section 6.2), Moody and Shakes quality framework (Section 6.3), and the four principles proposed by (Österle et al., 2011) (Section 6.4). Finally, we are going to present the result of the evaluations based in the above sections (Section 6.5).

6.1 Design Science Research Evaluation Framework

Evaluation is one of the most crucial steps in DSRM because it is what verifies the contribution of the solution for the identified problem and its utility, quality, and efficacy. The five design evaluation methods in context of DSRM (Hevner et al., 2004) are: **observational**, **analytical**, **experimental**, **testing**, and **descriptive**. Although these divisions of evaluation paths are defined, not much more guidance was provided on how to accomplish them.

In order to fill this gap and taking into account prior research done in the area of DSRM evaluation, in (Pries-Heje, Baskerville, and Venable, 2004), proposed a framework to help design science researchers build strategies for evaluation and to achieve improved rigor in DSRM.

This framework consists on distinguishing evaluation in three dimensions. However, each dimension has two aspects. The first dimension is concerned with the moment in time of the evaluation. This evaluation can be done **ex ante**, meaning that the evaluation takes place before the artifact is developed, or **ex post**, meaning that the evaluation is conducted with the artifact already developed. The second dimension is related to the form of the evaluation, that can be **artificial**, meaning that the evaluation considers a solution in a contrived and non-realistic way, and **naturalistic**, meaning that the evaluation consists on exploring the performance of a solution within its real environments. Finally, the third dimension distinguishes the artifact between **design process**, which is the result of a particular process that can be considered tangible, and **design product**, which can be defined as the set of activities, tools, methods, and practices that can be used to guide the flow of production. Figure 27 resumes the three dimensions presenting the combinations possible.

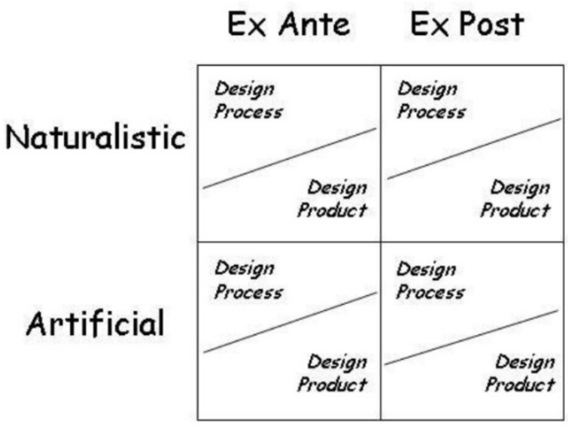

Figure 26. Framework for DSRM Evaluation (Pries-Heje, Baskerville, and Vanable, 2004)

The framework proposes a strategy to evaluate the artifact that is based on three questions:

- When does evaluation take place?
- How is it evaluated?
- What is actually evaluated?

6.2 Interviews with Practitioners

The objective of the interviews with practitioners is mainly to validate the research, its proposal, and results from demonstrations. This technique is used to collect qualitative information by setting up a situation that allows an interviewee the time and scope to talk about his opinions on a particular subject. The objective is to understand the interviewee's point of view rather than to make generalizations about behavior. There are three kinds of interviews (Bryman, 2012):

- **Unstructured**: when the researcher has a clear plan, but minimum control over how the respondent answers.
- **Semi-structured**: when the researcher uses a guide with questions and topics that must be covered.
- **Structured**: when the researcher has fixed questions and they are asked in a specific order.

6.3 Moody and Shanks Quality Framework

The Moody and Shanks Quality Framework is the result of research on how to evaluate and improve the quality of data models from the perspective of the multiple stakeholders and proposes the following quality factors (Moody and Shanks, 2003):

- **Completeness**: refers to whether the model contains all user and information requirements.
- **Integrity**: refers to whether the data model defines all business rules that apply.
- **Flexibility**: measures the ease with which the model can reflect changes in requirements without changing the model itself.
- **Understandability**: measure the ease with which the concepts and structures in the model can be understood.
- **Correctness**: refers to whether the model conforms to rules and conventions.
- **Simplicity**: refers to whether the model contains the minimum number of entities needed for the model.
- **Integration**: refers to whether the consistency of the model with the rest of the organization.
- **Implementability**: measures the ease with each the model can be implemented according to defined constraints.

6.4 Four principles of Österle et al. (2011)

These principles result from a memorandum written by 10 authors and supported by 111 full professors, with the objective of providing: rules for scientific rigor and improved guidance for researchers; criteria for journal and conference reviewers work; criteria for selection of young researchers and tenure procedures; criteria for evaluation of researchers and research organizations; and design-oriented information systems research in the international research community. In summary, it tries to provide a contribution to the rigor of research. These principles are (Österle et al., 2011):

- **Abstraction**: the artifact must be applicable to a class of problems.
- **Originality**: the artifact must substantially contribute to the advancement of the body of knowledge.
- **Justification**: the artifact must be justified in a comprehensible manner and must allow for its validation.
- **Benefit**: the artifact must yield benefit, either immediately or in the future, for the respective stakeholder groups.

6.5 Results of the Evaluation

In this section we are going to explain how we made the evaluation and what the conclusions were. We followed the previously order and the subsections are composed by Design Science Research Evaluation Framework (Subsection 6.5.1), interviews with practitioners (Subsection 6.5.2), Moody and Shanks Quality Framework (Subsection 6.5.3), and Four Principles of Österle et al. (Subsection 6.5.4).

6.5.1 Design Science Research Evaluation Framework

To evaluate our research we start by using the framework proposed in (Pries-Heje, Baskerville, and Venable, 2004), which is formulated as follows:

- **When did evaluation take place?** The evaluation strategy was **ex post** as the artifact was evaluated after their construction and demonstration.
- **How was it evaluated?** The evaluation strategy was **naturalistic** since it was conducted using a real artifact in a real organization facing real problems.
- **What was actually evaluated?** The developed artifact was considered a design science research artifact method and their evaluation represents an artifact **design process**, since it is defined as a set of activities that can be used to guide a procedure to help organizations.

6.5.2 Interviews with Practitioners

We had already designed our solution based on the main literature of that area, which gave us a strong theoretical viewpoint. In order to provide some practitioner viewpoint, we evaluated our proposal by performing 12 interviews with clients and providers of Cloud services at Portuguese organizations. With this evaluation we aim to know what criteria really matters to DM in real organizations when they are evaluating Cloud services.

We conducted these interviews with seven Cloud clients, which include public and private organizations from small to large organizations (European Commission, 2005), and five Cloud providers that include private organizations, from large and well-established organizations to small start-ups. We found the opinion of the providers useful since they interact with a high range of clients so they have knowledge about what criteria have more importance to clients. Detailed information about interviewees is provided in Table 26.

The interviews were semi-structured and consisted in a face-to-face or online meeting with the duration of 60 minutes each approximately. To support them we designed a questionnaire in order to lead the discussion in which the interviewees were asked to comparatively evaluate the importance of the initial proposal that includes six groups: **accountability**, **agility**, **management**, **performance**, **security**, and **usability**, which were divided in 30 criteria. This evaluation was made in a scale from one, less important, to six, extremely important. The interviewees were also asked to explain their classifications and to propose changes or addition to our criteria to evaluate Cloud services. In order to understand their Cloud experience, we collected information about what services they had in Cloud and their opinions.

Type	Sector	Employees	Role	Years of Experience
Client	Public Administration	130	CIO	20
Client	Public Administration	450	CIO	8
Client	Distribution and Retail	68	IT Manager	5
Client	Distribution and Retail	68	Financial Manager	5
Client	Banking	170	IT Manager	14
Client	Education	403	IT Manager	2
Client	Services	6400	IT Manager	8
Provider	Communications	6	CEO	3
Provider	IT Services	90000	ISV Partner Manager	18
Provider	IT Services	56	Customer Services Director	16
Provider	Consultant	15	Managing Partner	5
Provider	IT Services	90000	Solution Specialist	24

Table 26. Interviewees Details

After analyzing of gathered data we obtained the graphic present in Figure 28.

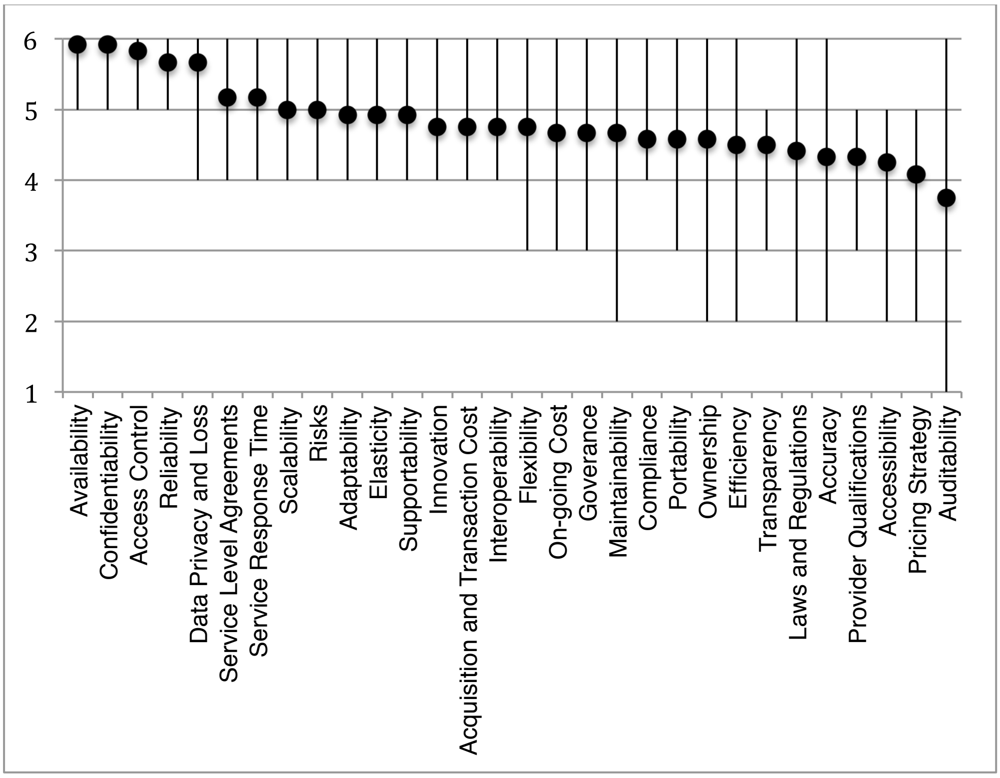

Figure 27. Results from the interviews

The lowest point of the lines represents the lowest value assigned, the highest point of the lines represents the highest value assigned, and the circle represents the results average of each criterion.

For example, considering the criteria **Service Level Agreements**, the lowest value assigned was 4, the higher value assigned was 6, and their average was 5.

By analyzing the graphic we could, on the one hand, conclude that criteria as **availability**, **confidentiality**, **access control**, **reliability**, and **data privacy and loss** were the most important in Cloud services and this conclusion is in consonance with literature. We could also see that our improvements are valuable for organizations whereas criteria as **confidentiality**, **risks**, **flexibility**, **supportability**, **innovation**, and **efficiency** were important for organizations too. On the other hand, the lowest importance was attributed to **auditability**, **pricing strategy**, **accessibility**, **provider qualifications**, and **accuracy**. Additionally, we can also conclude that for organizations the most important criteria were related with **security** and **performance**, which is not concordant with literature.

Through the analysis of the results we can observe that 30% of the criteria were considered extremely important, 67% very important, and only 3% were considered moderately important. Finally, we noted that 20% of the criteria had high variations in their evaluation and failed to generate a consensual opinion.

The interviews main goal was to validate our first model and assess if our proposal is adapted to organizations use when they are thinking about Cloud solutions. As a result, we observed that all organizations had an interest in our proposal and revealed the intention to use it.

Afterwards, taking into account the results of the interviews, we validated these results with a MACBETH expert, who had a lot of knowledge about decision analysis and we constructed the final proposal (Chapter 4).

6.5.3 Moody and Shakes Quality Framework

Then, we made the demonstration phase in a Portuguese City Council and the following results were obtained from applying the Moody and Shakes Quality Framework to this demonstration. We also consider the feedback from the interviews in these results. The results of Moody and Shakes Quality Framework were:

- **Completeness**: the proposal was evaluated in terms of the missing information, i.e. criteria. In general, the proposal is complete since the mainly criteria to evaluate Cloud services are present and some those that are not. Each DM can include or remove criteria and their descriptors of performance to better assess their evaluation, respecting the decision analysis rules.
- **Integrity**: is highly dependent on the DM. There is no business rule or other constraint that prevents errors defining the criteria and their descriptors of performance of the proposal since it relies on interviews and observations.

- **Flexibility**: the proposal includes the possibility to change the criteria. Each organization has different business and strategy and their DM can add or remove criteria. The difficult to make changes is directly proportional to the number of the criteria that each DM want to take into account.
- **Understandability**: practitioners in general found the proposal obtained easy to understand since their language is close to traditional in Cloud services but they do not know the decision analysis process and this phase can be more difficult to execute without a guide.
- **Correctness**: practitioners in general found the proposal correct and valid for their intentions. However, it is also dependent on each organization.
- **Simplicity**: according to the practitioners, the proposal is simple to follow, and we verified that is simple to apply. The criteria presented in the proposal was strategically reduced taking into account all suggestions made in interviews phase and when mapped with the MCDA assumptions.
- **Integration**: the proposal is consistent with the organization's needs. Considering the City Council's CIO, which has doubts about what Cloud service should buy, the proposal could help him to make the a more informed decision.
- **Implementability**: practitioners showed interest in use the proposal but the implementability is dependent on law and policies in each organizations. They admit to use it as an auxiliary tool.

In summary, almost all the quality factors were accomplished. In general, factors as **completeness**, **flexibility**, **correctness**, **simplicity**, and **integration** were totally accomplished. However, on the one hand, **integrity** and **understandability** were half accomplish. The first factor since there is no rule and/or constraints to support all integrity. The second one, since the practitioners found the proposal not easy to understand at the beginning. However, after a needed adaption period they feel comfortable. On the other hand, the factor **implementability** was not verified. There are several bureaucracies to implement this kind of solutions.

6.5.4 Four principles of Österle et al. (2011)

Finally, we applied the four principles proposed by (Österle et al., 2011) to evaluate the artifact. These are:

- **Abstraction**: the artifact can be applied to any organization, which is in doubt between any Cloud services and give the option to add criteria to meet all requirements of each organization. All partitions agreed that it was the case.
- **Originality**: none of the practitioners had knowledge of any research or product similar to the proposed artifact, and none similar was found in the related work.
- **Justification**: the artifact is supported by the motivation of the problem and by the related work. Then, the artifact itself was described and justified in a comprehensible manner using

textual and graphical representations with clear steps and instructions. In addition, the demonstration also helped to further explain and justify it.

- **Benefit**: according to the practitioners, there would be valuable benefit, since it would provide an easier and complete evaluation of the Cloud services, which they are taking into account. This was also confirmed by the results of the interviews and demonstration, where in some cases the DM showed the intention to use it.

To sum up, **the four principles to evaluate a DSRM artifact were achieved**, thus showing the validity of the artifact.

Chapter 7

Communication

This chapter corresponds to the **communication** phase of DSRM, which consists in communicate the problem an its importance, the artifact, its utility and novelty, the rigor of its design, and its effectiveness to researchers and other relevant audiences such as practicing professionals.

In order to do so we decided to do this through two different ways: demonstrations to practitioners and experts and by submission of scientific publications.

Regarding the demonstrations, the audience was composed by the elements already identified in the evaluation. This communication results in further evaluation since by communicating our method we were approached with suggestions.

Regarding the scientific publications, we submitted two research papers to two international conferences:

- Costa, P., Santos, J., and Mira da Silva, M. (2013). Evaluation Criteria for Cloud Services, *IEEE 6th International Conference on Cloud Computing (IEEE CLOUD 2013)*, Santa Clara Marriott, California, USA. Ranking: B (ERA). **(Accepted)**
- Costa, P., Lourenço, J. C., and Mira da Silva, M. (2013). Evaluating Cloud Services using a Multiple Criteria Decision Analysis Approach, *11th International Conference on Service Oriented Computing (ICSOC 2013)*. Ranking: A (ERA). **(Under Revision)**

In the first paper was proposed a model based on a set of criteria to evaluate Cloud services that consists in six groups of 30 measurement criteria. Our study was supported in 12 interviews with experts of Cloud services and the results showed that the model was globally accepted and can help the decision-making about Cloud services. Note also, which this conference is reference in the Cloud Computing area.

Regarding the second paper, we used the criteria proposed in the first paper, which are scientifically validated, in order to demonstrate the practical value of our method. We applied our proposal to help to address the problem identified by the Portuguese City Council's CIO that has doubts about what productivity Cloud service he should purchase. We made our recommendations and our proposal was considered very useful to address the problem. The gathered feedback by the DM was very positive. Note also, which the ICSOC conference is ranked as A in ERA.

Chapter 8

Conclusion

In a world of growing Outsourcing business and technological evolutions, as Cloud Computing, organizations are forced to make improvements and the best choices in order to compete with any organization. Notwithstanding, organizations suffer from the lack of knowledge when choosing that type of source they will hire, which reveals a problem: the **decision makers in organizations do not know how to evaluate Cloud services**. Moreover, the literature and some studies from consulting enterprises consider that Outsourcing has changed and several options have benefits, risks, challenges, and issues. Considering this, DMs have an arduous task to make the best choice, but they do not have any structured form to make them.

Through the interviews and literature review, we conclude that DM are having many doubts about Cloud and do not know what to do because there are many different opinions about Cloud Computing. Their main concerns are related to the security of their services, the local of housing and accessibility to their data, change acceptance for human resources, and innovation. These concerns are in consonance with literature (Willcocks and Lacity, 2012).

We also found that the increased use of Cloud by each organization is in non-core services as e-mail, office, and communications. In addition, we found that the costs and price strategy are very important mostly for provider organizations. However, client organizations give much more importance to simple price strategy and better QoS, although it can be more expensive. So, the postponing of Cloud adoption is visible because organizations do not know the real Cloud concept and they have some fears and doubts about it.

Following this, our proposal is to help them with a **method to evaluate Cloud services using MACBETH** that simplifies the decision-making in all organizations undecided between some Cloud services. Our proposal can use the defined criteria to compare all options of Cloud services and give the best option based on the evaluation criteria selected by the DM.

To access the usefulness of our artifact we demonstrated it in the City Council. We choose this organization because it CIO had doubts between two Cloud services and our method can help them to reduce their doubts. All criteria and their descriptors of performance were validated by the DM in the City Council with the intention to adapt them to his needs.

To access the utility of the proposed artifact and its results we have used:

- The framework proposed in (Pries-Heje, Baskerville, and Venable, 2004);
- Interviews with practitioners;
- Demonstration of the utility of the method in a Portuguese City Council;
- The Moody and Shanks Quality Framework to access the quality of the produced models (Moody and Shanks, 2003);
- The four principles of Österle et al. (2011) to evaluate the artifact;
- The appraisal from the scientific community through the submission and presentation of papers.

These evaluations showed that the proposal is generic enough to be applied in different organizations. Furthermore, the other evaluations methodologies were positive, as well as practitioners showed a good acceptance and enthusiasm for this approach.

Finally, we communicated the method and our results to the proper audiences through demonstrations and by scientific publication to recognized conferences.

In the next sections we are going to detail our conclusions by presenting were the lessons learned during this research (Section 8.1), the limitations we identified (Section 8.2), the future work related to the context of this research that we believe would bring extra value (Section 8.3), and finally, which are the main contributions of this proposal (Section 8.4).

8.1 Lessons Learned

There were several aspects during this research that are interesting to mention. Some resulted from the related work, the proposal construction, and others resulted from the experience obtained during the demonstration and evaluation phase.

Regarding the lessons learned during the related work research, we observed that there is already a clear effort from several frameworks and methodologies to help organization in the process of choosing Cloud services but the problem stays without solution. Although all known benefits that Cloud services bring to each organization, DMs in these organizations continue to express their doubts and maintain their antiquated thoughts due to the fear of change. This feeling was created because does not exist any structured form or guidelines to do it. Also with related work, we learned that decision analysis methodology is a very useful strategy in IT and when a fundamental decision has to be taken.

Regarding the lessons learned during the proposal construction, these are also connected with related work. First of all, several organizations are facing the same problem when are thinking about Cloud adoption and this unanimity was confirmed with the interviews. The mainly criteria and concerns are shared by DM in all interviews. The other lesson learned was about the decision analysis, which is not well known by organizations and is an interesting way to help make decisions in organizations. These decisions can be very important, as choosing a Cloud service, or less important, as choosing a new printer. Using decision analysis in organizations can accelerate many processes and reduce costs.

Finally, regarding the lessons learned during the demonstration and evaluation phase, these were rather practical and quite easy to summarize. In the first place the accuracy given by the method is highly dependent on the opinion of DM. To obtain better results, DMs have to know precisely what they need before the beginning of the process. Finally, we observed that organizations, even stating that the proposal has value, could not overcome the resistance to change to start using our method.

8.2 Limitations

The limitations associated with our proposal are directly related to the demonstration and interviews done and the MACBETH software used.

Regarding to the first point, we cannot state definitively that our method is applicable to every industry, in organizations with every size, since we only did one demonstration in a public organization of medium size, but the feedback collected, with this demonstration and with the 12 interviews, gave us indications that it might be possible to apply to all organizations that intend to purchase any Cloud service. In addition, we also took into account the SaaS model in our demonstration, comparing Google Apps against Microsoft Office 365, and we did not compare any service from PaaS and IaaS model. Since each DM can add or remove criteria and change the descriptors of performance for each criterion, we believe that our method can be used in any Cloud model and by all industry.

Regarding the technical details related with the MACBETH software, it has two mainly limitations that can be improved. The first one is that it is an old software with old appearance that can be difficult to learn and use. The second one, and the most important, is that the software cannot support a hierarchical analysis. If we have a tree with three levels of criteria and we need to give weights to all criteria along the three, we have to calculate the missing weights and insert them. Furthermore, in order to make an assertive evaluation, it is necessary that a decision analyst help in the process because the DM does not have the necessary knowledge.

Finally, the last limitation is about the result of Moody and Shanks Quality Framework. The factors integrity, understandability, and implementability were not all accomplish.

8.3 Main Contributions

We believe that our proposal brings a valuable contribution in the context of Cloud services. We also believe that our method presents an answer to the research problem we raised since on the one hand, by using the decision analysis methodology together with the mainly criteria to compare and evaluate Cloud services, organizations are able to choose the best Cloud solution that meets their needs. On the other hand, our method enables a quickly identification of what Cloud service is appropriate and, consequently, dramatically reduce the total effort in this choice.

The proposed method provides a process of how to achieve the best Cloud service for each organization. We believe that these well-defined steps will also act like a guideline for all DMs in their organizations. Regarding the latter, the proposed method reduces the budget spent in personnel and facilities, making the decision process easier for SMEs and well-established organizations.

The method and the results obtained from the demonstration and from the interviews were also evaluated by both practitioners and experts. Their feedback was highly appreciated and resulted in several changes in the method and generally it went accordingly to our expectations. The method would be more affordable than those available today, it would be easier and it is innovative, making it a valuable contribution.

Finally, we believe that with our proposal we can reduce the postponing of Cloud adoption with the reduction of the DM doubts and fears.

8.4 Future Work

Regarding future steps we believe that further research can be performed in order to make our proposal more general and easy to apply. These future steps are highly related with the limitations previously mentioned.

First of all, we believe that have to be made more demonstrations of our method and applying it to more organizations to verify if it really can be used by all industries and by all kind of organizations, from large to public organizations and public or private.

The second aspect is related with the three Cloud models. We believe that a further research related the different Cloud models could be used to create the best variations to our method with criteria focus on each Cloud model, i.e. SaaS, PaaS, and IaaS. With these variations it would be possible to create a template for each Cloud model that would require fewer changes by each DM in different organizations.

Finally, the third point, in order to make the proposal easy to apply and to overcome the MACBETH software limitations, we believe that our proposal can be further developed, starting by the implementation of a software tool, and then evaluate its applicability by promoting a pilot where the analysis would be done by the DM.

Bibliography

Alhamad, M., Dillon, T., and Chang, E. (2010). Conceptual SLA Framework for Cloud Computing, *4th IEEE International Conference on Digital Ecosystems and Technology*.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. (2010). Above the Clouds: A Berkeley View of Cloud Computing, *EECS Department, University of California, Berkeley*.

Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. (2010). A View of Cloud Computing, *Communications of the ACM*, 53(4), 50-58.

Bana e Costa, C.A. and Vansnick, J.C. (1994). MACBETH – An interactive path towards the construction of cardinal value functions, *International Transactions in Operational Research*, 1(4), 489-500.

Bana e Costa, C.A. and Vansnick, J.C. (1997). Applications of the MACBETH approach in the framework of an additive aggregation model, *Journal of Multi-Criteria Decision Analysis*, 6(2), 107-114.

Bana e Costa, C. A. and Vansnick, J. C. (1999). The MACBETH Approach: Basic Ideas, Software, and an Application. In: Meskens, N. and Roubens, M. R. (Eds.), *Advances in Decision Analysis*, Kluwer Academic Publishers, pp. 131-157.

Bana e Costa, C. A. and Vansnick, J. C. (2008). A Critical Analysis of the Eigenvalue Mehtod used to Derive Priorities in AHP, *European Journal of Operational Research*, 187(3), 1422-1428.

Bana e Costa, C. A., De Corte, J. M., and Vansnick, J. C. (2005a). M-MACBETH Version 1.1 User's Guide. Retrieved from http://www.m-macbeth.com/.

Bana e Costa, C. A., De Corte, J. M., and Vansnick, J. C. (2005b). On the Mathematical Foundations of MACBETH. In: Figueira J., Greco, S., and Ehrgott, M. (Eds.), *Multiple Criteria Decision Analysis: State of Art Surveys*, Springer, pp. 409-442.

Bana e Costa, C. A., De Corte, J. M., and Vansnick, J. C. (2011). MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique), *Wiley Encyclopedia of Operations Research and Management Science*, John Wiley & Sons, 4, 2945-2950.

Bana e Costa, C. A., De Corte, J. M., and Vansnick, J. C. (2012). MACBETH. *International Journal of Information Technology & Decision Making*, 11(2), 359-387.

Bana e Costa, C. A., Lourenço, J. C., Chagas, M. P., and Bana e Costa J. C. (2008). Development of Reusable Bid Evaluation Models for the Portuguese Electric Transmission Company, *Decision Analysis*, 5(1), 22-42.

Belton, V. and Stewart, T. J. (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*, Kluwer Academic Publishers.

Bradshaw, S., Millard, C. and Walden, I. (2011). The Terms They Are A-Chargin'… Watching Cloud Contracts Take Shape, *Issues in Technology Innovation*, (7), 1-12.

Bryman, A. (2012). *Social Research Methods*, Oxford University Press.

Buyya, R., Yeo, C., and Venugopal, S. (2008). Market-oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Service as Computing Utilitites. In: *10th IEEE International Conference on High Performance Computing and Communications (ICHPCC 2008)*, pp. 5-13. Dalian, China.

Cloud Services Measurement Initiative Consortium (2011). Service Measurement Index, *CSMIC, Carnegie Mellon University Silicon Valley*.

Cloud Services Measurement Initiative Consortium (2012). Cloud Services Measurement Initiative Consortium (CSMIC) Overview, *CSMIC, Carnegie Mellon University Silicon Valley*.

Conway, G. and Curry E. (2012). Managing Cloud Computing: A Life Cycle Approach. In: *2nd International Conference on Cloud Computing and Services Science (CLOSER 2012)*, pp. 198-207. Porto, Portugal.

Cullen, S., Seddon, P., and Willcocks, L. (2005). *Managing Outsourcing: The lifecycle Imperative*, MIS Executive.

Dastjerdi, A., Tabatabei, S., and Buyya, R. (2012). A Dependency-aware Ontology-based Approach for Deploying Service Level Agreement Monitoring Services in Cloud, *Journal Software – Practice Experience*, 41, 501-518.

Dyer, J. S. (1990). Remarks on the Analytic Hierarchy Process, *Management Science*, 36(3), 249-258.

Dyer, J. S. (2005). MAUT – Multiattribute Utility Theory. In: Figueira J., Greco, S., and Ehrgoot, M. (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*. New York, Springer, pp. 265-295.

Dyer, J. S. and Sarin, R. K. (1979). Measurable Multiattribute Value Functions. *Operation Research*, 27(4), 810-822.

Earl, M. (1996). *The Risks of Outsourcing IT*, Sloan Management Review/Spring.

European Commission (2005). The New SME Definition: User Guide and Model Declaration, *European Commission*.

Garg, S., Versteeg, S., and Buyya, R. (2012). A Framework for Ranking of Cloud Computing Services, *Journal Future Generation Computer Systems*, 29(4), 1012-1023.

Goodwin, P. and Wright, G. (2004). *Decision Analysis for Management Judgement*. John Wiley & Sons.

Hevner, A., March, S., Park, J., and Ram, S. (2004). Design Science in Information Systems Research, *MIS Quarterly*, 28(1), 75-105.

Hirschheim, R., George, B., and Wong, S. F. (2004). Information Technology Outsourcing: The Move Towards Offshoring, *Indian Journal of Economics and Business*.

Hogan, M., Liu, F., Sokol, A., and Tong, J. (2011). NIST Cloud Computing Standards Roadmap, *National Institute of Standards and Technology*, Special Publication 500-291.

Hon, W. K., Millard, C., and Walden, I. (2012). Negotiating Cloud Contracts: Looking at Clouds from Both Sides Now, *Stanford Technology Law Review*, 16(1), 81-130.

IDC Portugal (2012). Situação Actual e Tendências de Adopção de Serviços Cloud Computing em Portugal, *IDC Survey* (in Portuguese).

ISO/IEC (1991). ISO/IEC 9126 – Information Technology – Software Product Evaluation – Quality Characteristics and Guidelines for their use.

Jansen, W. and Grace T. (2011). Guidelines on Security and Privacy in Public Cloud Computing, *National Institute of Standards and Technology*, Special Publication 800-144.

Keeney, R. L. (1992). *Value-Focused Thinking: A Path to Creative Decision-Making*, Harvard University Press.

Keeney, R. L. and Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, John Wiley & Sons.

Kundra, V. (2011). Federal Cloud Computing Strategy, *The White House*.

Lacity, M. and Hirschheim, R. (1993). *Information Systems Outsourcing: Myths, Metaphores, and Realities*, England: Wiley.

Lacity, M. and Willcocks, L. (2012). *Advanced Outsourcing Practice – Rethinking ITO, BPO, and Cloud Services*, Palgrave Macmillan.

Lee, J. N. and Kim, Y. G. (1999). Effect of Partnership Quality on IS Outsourcing Success: Conceptual Framework and Empirical Validation, *Journal of Management Information Systems*, 15(4), 29-61.

Low, C. and Chen, Y. (2012). Criteria for the Evaluation of a Cloud-based Hospital Information System Outsourcing Provider, *Springer Science.*

McAfee, A. (2012). What Every CEO Needs to Know About the Cloud, *Harvard Business Review*, 89(11), 124-132.

Mira da Silva, M. and Martins, J. (2008). *IT Governance – A Gestão da Informática*, FCA (in Portuguese).

Moody, D.L. and Shanks, G.G. (2003). Improving the Quality of Data Models: Empirical Validation of a Quality Management Framework, *Information Systems*, 28, 619-650.

Nor, R., Alias, R., and Rahman, A. (2008). The Performance Measurement Model for ICT Service Quality. In: *4th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2008)*, pp. 1-5. Dalian, China.

Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., and Sinz, E. (2011). Memorandum on Design-Oriented Information Systems Research, *European Journal on Information Systems*, 20(1), 7-10.

Peffers, K., Tuunanen, T., Rothernberger, M., and Chatterjee, S. (2008). A Design Science Research Methodology for Information System Research, *Journal of Management Information Systems*, 24(3), 45-77.

Penter, K., Pervan, G., and Wreford, J. (2008). Case Studies on BPO and ITO at Large Captive Operations in India. In: *2nd Information Systems Workshop on Global Sourcing: Services, Knowledge and Innovation*.

Pries-Heje, J., Baskerville, R., and Venable, J. (2008). Strategies for Design Science Research Evaluation, In: *16th European Conference on Information Systems (ECIS 2008)*, pp. 255-266. Galway, Ireland.

Robinson N., Valeri, L., Cave, J., Starkey, T., Graux, H., Creese, S., and Hopkins, P. (2011). The Cloud: Understanding the Security and Privacy in Public Cloud Computing, *Rand Corporation*.

Saaty, T. L. (1980). *The Analytical Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw-Hill.

Saaty, T. L. (2005). The Analytical Hierarchy and Analytical Network Processes for the Measurement of Intangible Criteria and for Decision-making. In Figueira, J., Greco, S., Ehrgott, M. (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys*. New York: Springer, pp. 354-407.

Samini, P. and Patel, A. (2011). Review of Pricing Models for Grid & Cloud Computing, *IEEE Symposium on Computers & Informatics (ISCI 2011)*, pp. 634-639. Kuala Lumpur, Malaysia.

Staten, J. (2008). Is Cloud Computing Ready for the Enterprise?, *Forrester Report*.

Twentyman, J. (2012). How Far and How Fast? Comparing Three Governments' Progress in Cloud Computing, *Cisco, Government Business Council*.

von Winterfeldt, D. and Edwards, W. (1986). *Decision Analysis and Behavioral Research*, Cambridge University Press.

Walden, I. (2011). Accessing Data in the Cloud: The Long Arm of the Law Enforcement Agent, *Queen Mary University of London, School of Law*.

Weeden S. and Valiente, T. (2012). Cloud Computing: Every Silver Lining Has a Cloud, *Citi Research, Technical Report*.

Willcocks, L. and Lacity, M. (2012). *The New IT Outsourcing Landscape – From Innovation to Cloud Services*, Palgrave Macmillan.

Appendixes

Appendix A: Mapping SMI Attributes to Literature

Attribute	(Buyya, Yeo, and Venugopal, 2008)	(Nor, Alias, and Rahman, 2008)	(Armbrust et al., 2009)	(Alhamad, Dillon, and Chang, 2010)	(Kundra, 2011)	(NIST, 2011)	(Samimi and Patel, 2011)	(Conway and Curry, 2012)	(Garg, Versteeg, and Buyya, 2012)	(Low and Chen, 2012)
Governance					✓	✓		✓		
Provider contract/SLA verification	✓					✓		✓		
Auditability			✓			✓				
Data ownership					✓			✓		
Provider certifications			✓							✓
Provider personnel requirements		✓								✓
Compliance						✓				
Ownership							✓			
Sustainability									✓	
Contracting experience										
Ease of doing business										
Provider business stability										
Provider ethicality										
Provider supply chain										
Security capabilities										

Table 27. Mapping SMI accountability attributes to literature

Attribute	(Buyya, Yeo, and Venugopal, 2008)	(Nor, Alias, and Rahman, 2008)	(Armbrust et al., 2009)	(Alhamad, Dillon, and Chang, 2010)	(Kundra, 2011)	(NIST, 2011)	(Samimi and Patel, 2011)	(Conway and Curry, 2012)	(Garg, Versteeg, and Buyya, 2012)	(Low and Chen, 2012)
Scalability	✓		✓	✓	✓		✓		✓	
Portability					✓	✓				
Adaptability									✓	
Elasticity									✓	
Capacity										
Extensibility										
Flexibility										

Table 28. Mapping SMI agility attributes to literature

Attribute	(Buyya, Yeo, and Venugopal, 2008)	(Nor, Alias, and Rahman, 2008)	(Armbrust et al., 2009)	(Alhamad, Dillon, and Chang, 2010)	(Kundra, 2011)	(NIST, 2011)	(Samimi and Patel, 2011)	(Conway and Curry, 2012)	(Garg, Versteeg, and Buyya, 2012)	(Low and Chen, 2012)
Reliability	✓	✓		✓	✓			✓	✓	✓
Availability			✓	✓	✓	✓			✓	
Maintainability						✓				✓
Recoverability				✓						
Service stability									✓	
Data geographic/political										
Resiliency/fault tolerance										
Serviceability										

Table 29. Mapping SMI assurance attributes to literature

Attribute	(Buyya, Yeo, and Venugopal, 2008)	(Nor, Alias, and Rahman, 2008)	(Armbrust et al., 2009)	(Alhamad, Dillon, and Chang, 2010)	(Kundra, 2011)	(NIST, 2011)	(Samimi and Patel, 2011)	(Conway and Curry, 2012)	(Garg, Versteeg, and Buyya, 2012)	(Low and Chen, 2012)
On-going cost				✓					✓	✓
Acquisition & transition cost							✓		✓	
Profit or cost sharing										

Table 30. Mapping SMI financial attributes to literature

Attribute	(Buyya, Yeo, and Venugopal, 2008)	(Nor, Alias, and Rahman, 2008)	(Armbrust et al., 2009)	(Alhamad, Dillon, and Chang, 2010)	(Kundra, 2011)	(NIST, 2011)	(Samimi and Patel, 2011)	(Conway and Curry, 2012)	(Garg, Versteeg, and Buyya, 2012)	(Low and Chen, 2012)
Interoperability					✓	✓		✓	✓	
Service response time		✓	✓	✓					✓	
Accuracy									✓	✓
Suitability									✓	
Functionality										

Table 31. Mapping SMI performance attributes to literature

Attribute	(Buyya, Yeo, and Venugopal, 2008)	(Nor, Alias, and Rahman, 2008)	(Armbrust et al., 2009)	(Alhamad, Dillon, and Chang, 2010)	(Kundra, 2011)	(NIST, 2011)	(Samimi and Patel, 2011)	(Conway and Curry, 2012)	(Garg, Versteeg, and Buyya, 2012)	(Low and Chen, 2012)
Access control & privilege management				✓	✓	✓		✓		
Data privacy & data loss			✓			✓		✓		✓
Data integrity					✓	✓		✓		
Data geographic/political				✓						
Physical & environmental security						✓				
Proactive threat & vulnerability management						✓				
Retention/disposition										

Table 32. Mapping SMI security and privacy attributes to literature

Attribute	(Buyya, Yeo, and Venugopal, 2008)	(Nor, Alias, and Rahman, 2008)	(Armbrust et al., 2009)	(Alhamad, Dillon, and Chang, 2010)	(Kundra, 2011)	(NIST, 2011)	(Samimi and Patel, 2011)	(Conway and Curry, 2012)	(Garg, Versteeg, and Buyya, 2012)	(Low and Chen, 2012)
Accessibility	✓						✓			
Transparency							✓		✓	
Operability							✓			
Understandability							✓			
Client personnel requirements										
Installability										
Learnability										

Table 33. Mapping SMI security and privacy attributes to literature

Appendix B: Judgments Matrices

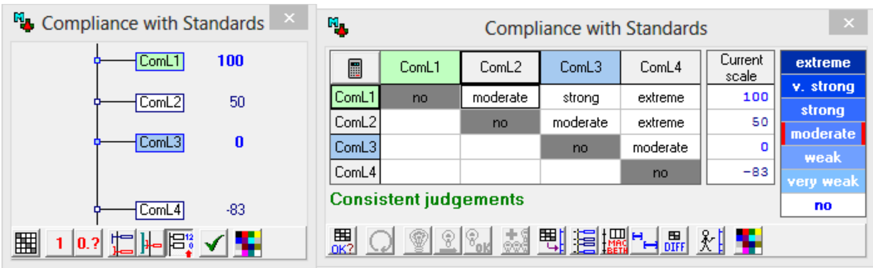

Figure 28. Judgment matrix and value function for the “Compliance with Standards” criterion

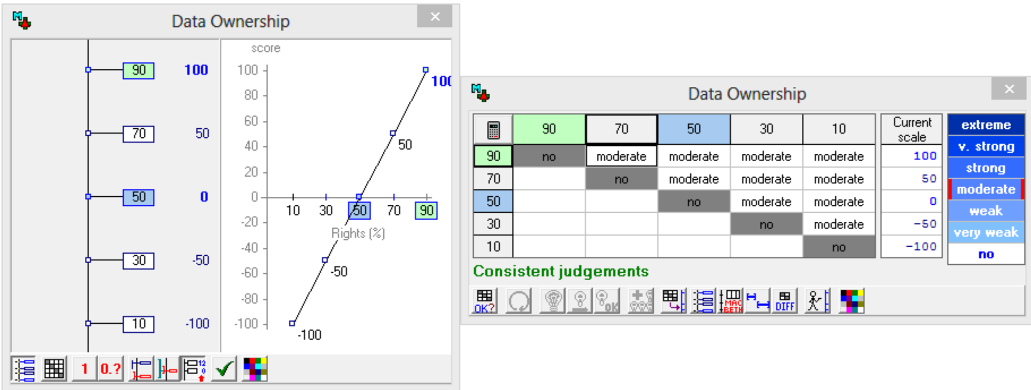

Figure 29. Judgment matrix and value function for the “Data Ownership” criterion

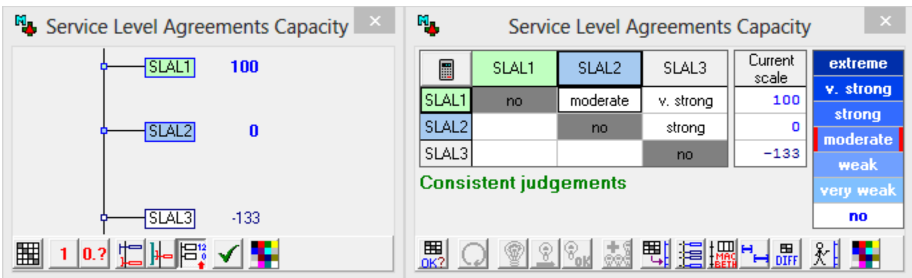

Figure 30. Judgment matrix and value function for the “Service Level Agreements Capacity” criterion

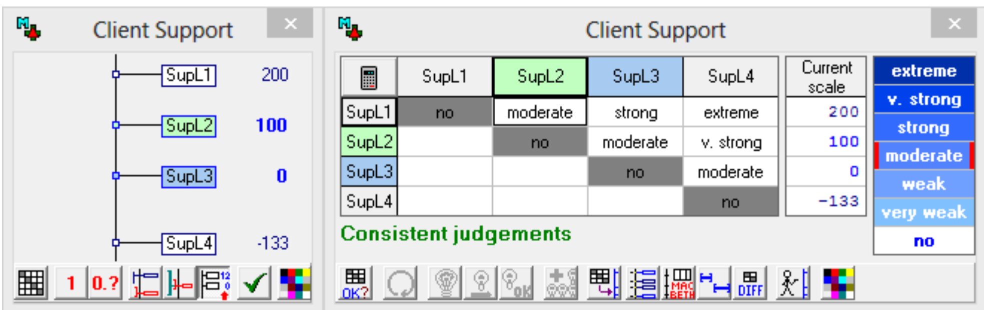

Figure 31. Judgment matrix and value function for the “Client Support” criterion

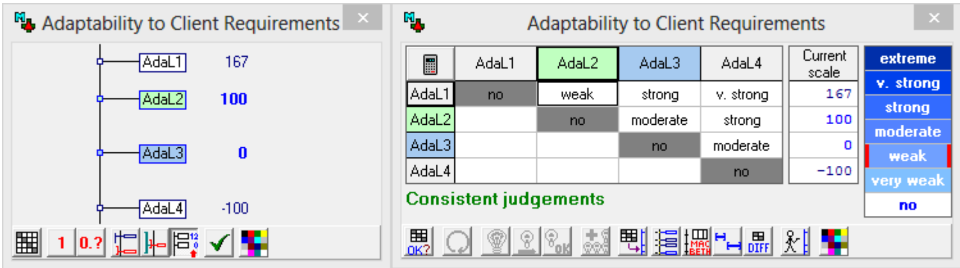

Figure 32. Judgment matrix and value function for the “Adaptability to Client Requirements” criterion

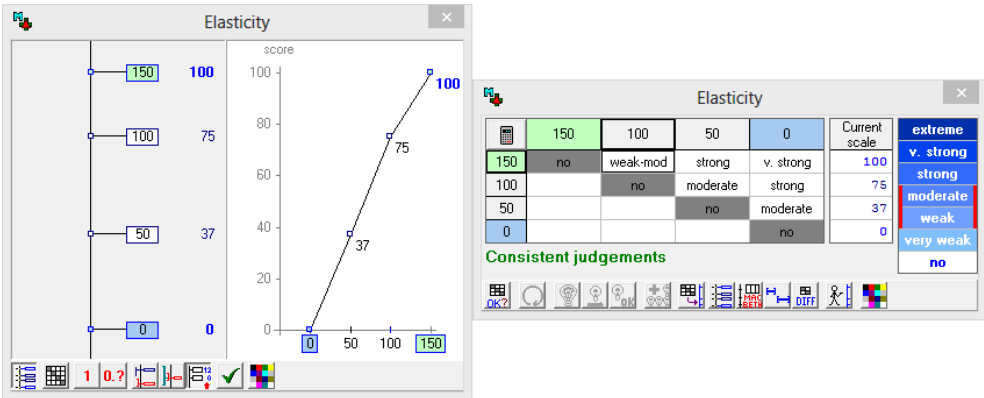

Figure 33. Judgment matrix and value function for the “Elasticity” criterion

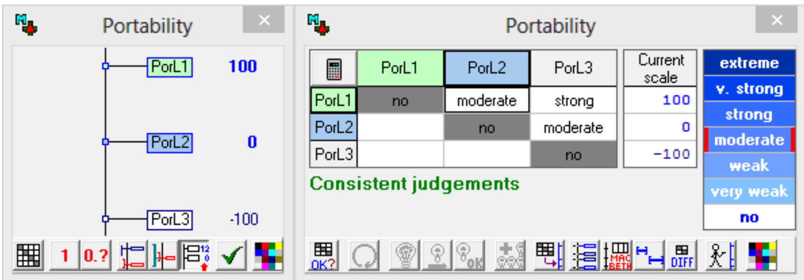

Figure 34. Judgment matrix and value function for the “Portability” criterion

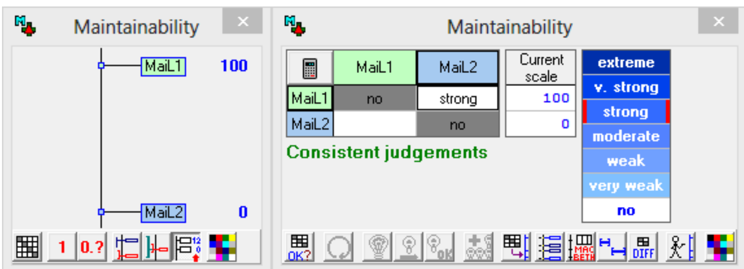

Figure 35. Judgment matrix and value function for the “Maintainability” criterion

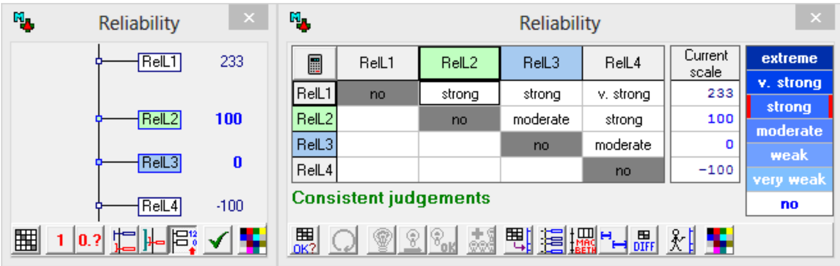

Figure 36. Judgment matrix and value function for the “Reliability” criterion

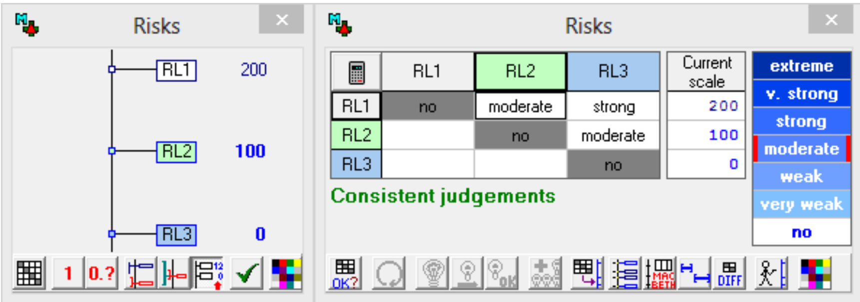

Figure 37. Judgment matrix and value function for the “Risks” criterion

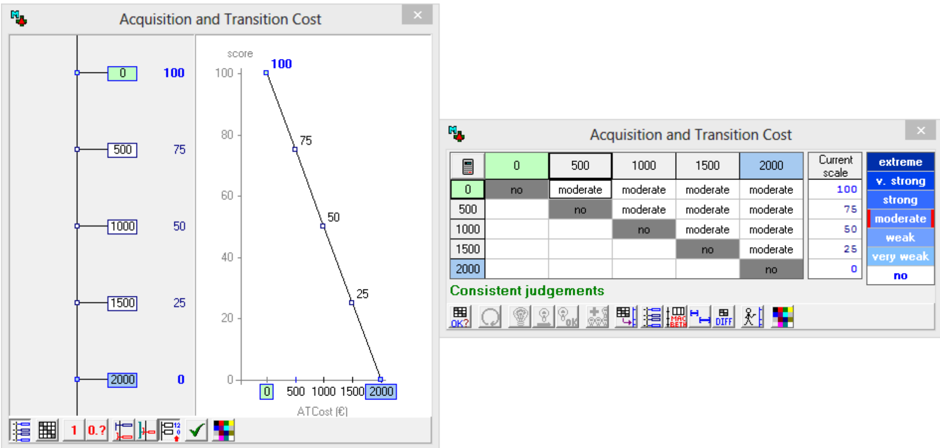

Figure 38. Judgment matrix and value function for the “Acquisition and Transition Cost” criterion

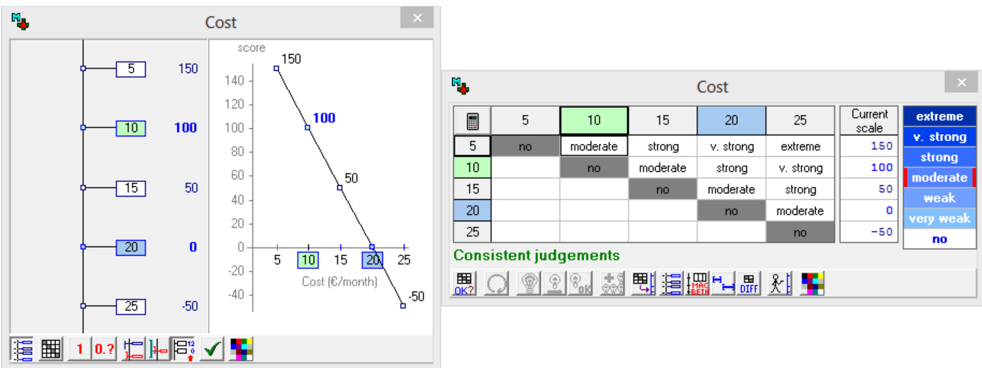

Figure 39. Judgment matrix and value function for the “Cost” criterion

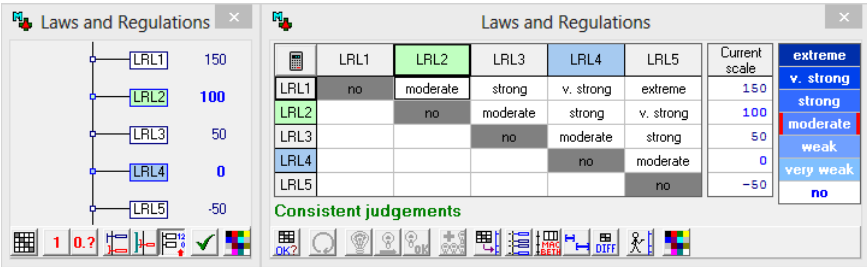

Figure 40. Judgment matrix and value function for the “Laws and Regulations” criterion

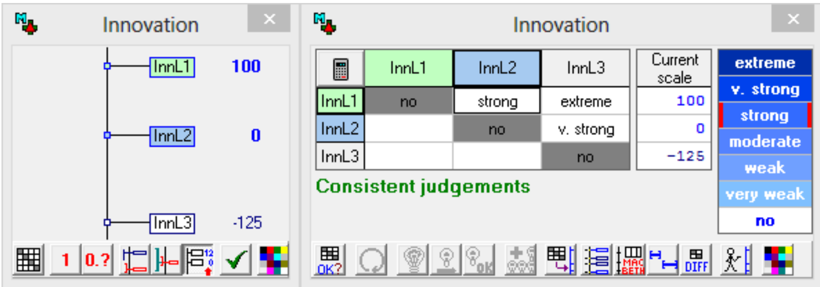

Figure 41. Judgment matrix and value function for the “Innovation” criterion

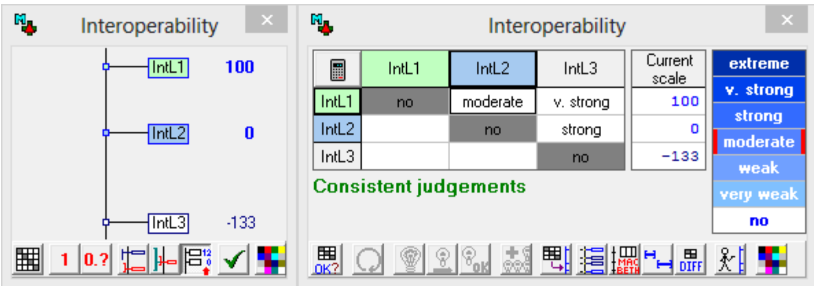

Figure 42. Judgment matrix and value function for the “Interoperability” criterion

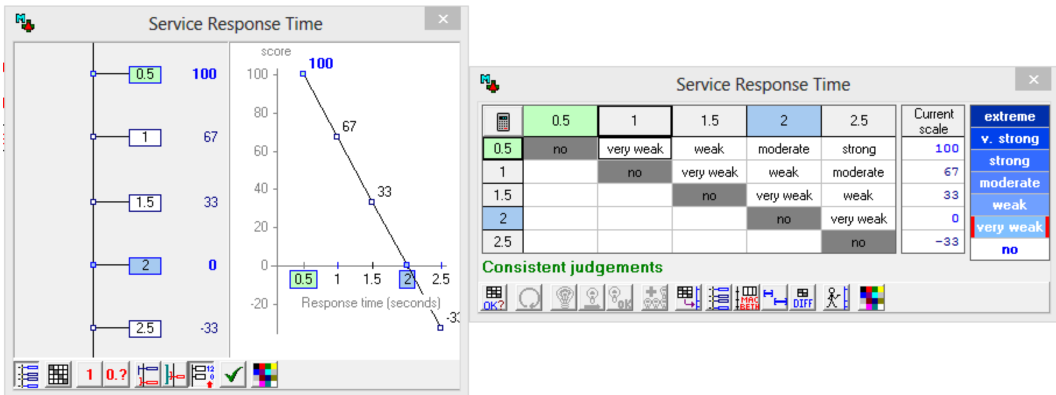

Figure 43. Judgment matrix and value function for the “Service Response Time” criterion

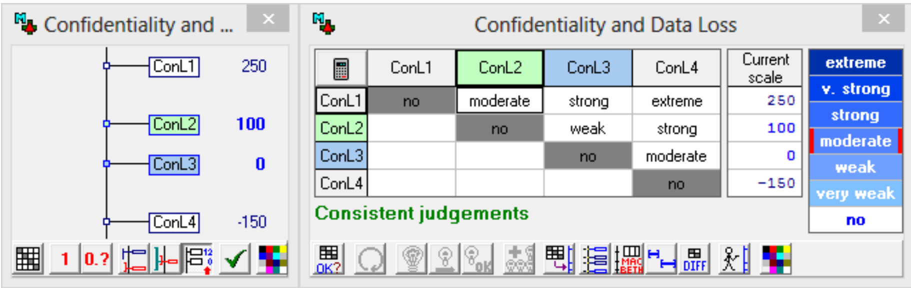

Figure 44. Judgment matrix and value function for the “Confidentiality and Data Loss” criterion