UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Frâncila Weidt Neiva

Towards Pragmatic Interoperability to Support Scientific Workflows Development

Juiz de Fora

2015

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Frâncila Weidt Neiva

Towards Pragmatic Interoperability to Support Scientific Workflows Development

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientador: José Maria Nazar David
Coorientadora: Regina Maria Maciel Braga

Juiz de Fora
2015

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

Neiva, Frâncila Weidt.
Towards Pragmatic Interoperability to Support Scientific Workflows Development / Frâncila Weidt Neiva. -- 2015.
126 f. : il.

Orientador: José Maria Nazar David
Coorientadora: Regina Maria Maciel Braga
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, 2015.

1. Pragmatic Interoperability. 2. Scientific Workflows. 3. Collaborative Systems. I. David, José Maria Nazar, orient. II. Braga, Regina Maria Maciel, coorient. III. Título.

Frâncila Weidt Neiva

Towards Pragmatic Interoperability to Support Scientific Workflows Development

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Aprovada em 30 de Setembro de 2015.

BANCA EXAMINADORA

Prof. D.Sc. José Maria Nazar David - Orientador
Universidade Federal de Juiz de Fora

Prof. D.Sc. Regina Maria Maciel Braga- Coorientadora
Universidade Federal de Juiz de Fora

Prof. D.Sc. Fernanda Cláudia Alves Campos
Universidade Federal de Juiz de Fora

Prof. Ph.D. Marcos Roberto da Silva Borges
Universidade Federal do Rio de Janeiro

"For my grandmas Geralda and
Ephigênia "
In memoriam

ACKNOWLEDGMENTS

I would like to express my deepest gratitude

To God, Supreme Intelligence and First Cause of all things;

To my parents Miria and Wantuir for their constant support and love;

To Rodrigo Luis for his important advice and constant patience and kindness;

To my advisors José Maria and Regina Braga for their valuable guidance and constant encouragement;

To professor Jairo Souza, Fernanda Campos and Marco Antônio for their help and valuable suggestions;

To professor Marcos Borges for his important considerations that contributed to improve the quality of this work;

To my friends from my master degree, Marina Ivanov, Guilherme Gomes, Vitor Freitas, Jacimar Tavares, Marcos Miguel, Wellington Veiga, Tassio Ferenzini, Humberto Dalpra and Rafael Almeida for sharing the good moments and also the difficult ones in this journey.

Finally, I would like to record my sense of gratitute to one and all who, directly or indirectly, have given a helping hand in this work.

"Live as if you were to die
tomorrow. Learn as if you were
to live forever."
Mahatma Gandhi

RESUMO

Fornecer suporte a interoperabilidade apenas considerando a forma e o significado (i.e. sintaxe e semântica) na troca de dados não é suficiente para se atingir uma colaboração efetiva e significativa. Neste sentido, a interoperabilidade pragmática tem se destacado como um requisito fundamental para garantir a colaboração em sistemas distribuídos. Entretanto, preencher este requisito não é uma tarefa trivial. O objetivo deste estudo é propor e avaliar uma solução para apoiar implementação da interoperabilidade pragmática em um sistema colaborativo. Assim, a solução proposta foi implementada e avaliada em um ecossistema de software baseado na web capaz de apoiar o desenvolvimento colaborativo de workflows científicos chamado ECOS Collaborative PL-Science.

Palavras-chave: Interoperabilidade Pragmática, Workflows Científicos, Sistemas Colaborativos.

ABSTRACT

Providing interoperability support only considering the format and meaning (i.e. syntax and semantic) in data exchange is not enough to achieve effective and meaningful collaboration. Pragmatic interoperability has been identified as a key requirement to foster collaboration in a distributed environment. However, fulfilling this requirement is not a trivial task. The aim of this study is to propose and evaluate a solution to support pragmatic interoperability implementation in a collaborative system. The proposed solution was implemented and evaluated in an open source web-based software ecosystem to support collaborative development of scientific workflows.

Keywords: Pragmatic Interoperability, Scientific Workflows, Collaborative Systems.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

DSD Distributed Software Development

PI Pragmatic Interoperability

SM2PIA Support Model to Pragmatic Interoperability Achievement

PRIME PRagmatic Interoperability to MEaningful collaboration

SOA Service Oriented Architecture

CONTENTS

1 INTRODUCTION

This chapter presents the work context that motivated this research and the research questions. The goals, research methodology and how this text is structured are also presented.

Software plays an important role in society. The dissemination of software use for different purposes and our dependency on it to perform daily tasks has generated a need for solutions to increase productivity and quality and to reduce costs in development processes. In order to fulfill these needs, companies geographically distributed their activities. However, this scenario has brought new challenges because collaborative systems have complex requirements which are not easy to fulfill. This complexity is partially related to the diversity of interactions and process types that these systems support, imposing a need for flexible rules and policies. For example, suppose that a discussion forum is used to support decision making in a meeting. In this scenario, at the moment when decisions are made, roles can be modified and new activities and policies need to be redefined. Besides these challenges, the dynamism of the components used to represent the interaction context also increase the complexity of the collaborative systems. When a decision is made, modifications in the context occur during the interaction, generating additional requirements and interest objects for the participants (DAVID; MACIEL, 2009).

In order to support collaborative systems analysis, Ellis et al. (1991) established the 3C collaboration model. This model defines collaboration as the combination of three dimensions: (i) Communication - related with the message exchanges generated from user interactions; (ii) Cooperation - management of people, their activities and resources and; (iii) Coordination - activities performed in a shared workspace.

A study conducted by Steinmacher et al. (2013) revealed that tools to support the communication dimension during collaboration in a distributed software development (DSD) environment were very poorly explored. Furthermore, communication issues are usually identified as one of the main difficulties in a distributed environment, which can be considered a gap and a good opportunity for research.

The communication process is explored in semiotics, the study of meaning- making, the philosophical theory of signs and symbols in the field of linguistics. Semiotics is divided into three dimensions: syntactic, semantics and pragmatics. Syntax acts as a sign, semantics is related with which the sign refers to, and pragmatics is related to the effect of the sign on the interpreter. This effect can be noticed depending on the context where the sign is used (MORRIS, 1938). These three aspects are important in the communication process.

Using semiotics, we can also explain a key requirement that arises with the DSD scenario, interoperability across collaborative systems. In DSD, the systems are distributed and they need to communicate accordingly. In other words, they need to interoperate in order to reach specified goals. Therefore, interoperability can be considered a fundamental requirement to enable collaboration in this application domain.

In this research context, interoperability can be defined as the ability of different systems to use each other's services effectively (ASUNCION; SINDEREN, 2011). These systems can share functionalities and information at different levels. In a syntactic level, the exchanged data acts as a sign and, to achieve this interoperability level, the syntax of this sign must be previously established as a standard. Semantic interoperability is concerned with ensuring that the meaning of the data across the considered system is the same and shared in an unambiguous way. Finally, pragmatic interoperability (PI) is concerned with ensuring that the message sender and receiver share the same expectations about the effect of the exchanged messages and the context where this exchange occurs (TOLK; MUGUIRA, 2003).

Interoperability across collaborative systems is not a trivial requirement. Each system has its own particularities and needs. As a result, each system requires that different interoperability levels be achieved.

Many researchers have argued that providing interoperability support only considering the format and meaning (i.e. syntax and semantic) in data exchange is not enough to achieve complete, effective and meaningful collaboration (ASUNCION; SINDEREN, 2010), (RUOKOLAINEN, 2009). In this vein, PI support has been seen as a key requirement to meet the desired effects during message exchange and different authors have discussed this, including Bravo and Alvarado (2008) and Schoop et al. (2006).

Considering the e-science domain, scientists usually expect that services address their research needs. However, in most cases they have to specify experiments composed of different services executed in pipeline. These services act as a scientific workflow (LUDÄSCHER et al., 2006). To illustrate the pragmatic interoperability requirement in e-science, let us suppose a collaborative system designed to support experiments in biology. Scientists working in this system need to compose services which are geographically dispersed. In order to fulfill this requirement, they must know the context in which the components are inserted, the constraints on their use, the concepts that they represent, the nonfunctional requirements that they have to fulfill, among others. In this scenario, providing pragmatic interoperability support can improve collaborative activities among the scientists taking into account tacit and non-formalized context factors such as components reliability, which scientist used which components and for what purpose they were selected, among others. Another example is supposing a scenario where services must interoperate. A syntactic level covers data exchange. If service A sends a string, service B has to expect a string. But, what is this string about? At the semantic level this question can be answered. If service A sends a string that represents a DNA sequence and service B expected a RNA sequence, there is no semantic interoperability. However, even if service A sent a DNA sequence while service B expected the same, is service B the best service? Do they meet the expectations? We answer these questions based on pragmatic aspects such as user context, service context, business rules, policies, restrictions, among others.

Although pragmatic interoperability support has been mentioned as a key issue in literature, this field is still in its infancy as can be seen in (ASUNCION; SINDEREN, 2010). This study found at least 44 unique pragmatic interoperability definitions. Besides revealing the lack of consensus, the authors found that each definition is associated with a particular research domain.

1.1 RESEARCH QUESTIONS

In this context, it is important to carry out research in order to investigate PI as this is a fundamental requirement to enhance collaboration in the DSD context. As illustrated by the examples previously discussed in this section, the e-science domain has the potential to improve activities with PI. In particular activities that involve the development of scientific workflows. Thus, this work proposes and evaluates a solution to address pragmatic interoperability in order to support collaborative development of scientific workflows. To tackle this, PI was investigated considering an open source web-based software ecosystem to support collaborative development of scientific workflows (http://pgcc.github.io/plscience/). The following questions were formulated to guide the research:

- Q1:Which strategies have been proposed to implement PI in software systems (e.g.Ontology)?
- Q2:How to create an architecture to implement PI to support the collaborative development of scientific workflows?
- Q3:What is the feasibility of the resulting architecture?
- Q4: What are the benefits that this architecture provides considering the development of scientific workflows?

1.2 GOALS

The main goal of this work is to propose an architecture to support PI during the development of scientific workflows. This goal can be divided into specific goals, based on the research questions. The specific goals are:

- G1: To establish a systematic mechanism to capture the strategies used to achieve PI.
- G2: To develop an architecture to achieve PI in a collaborative system that supports the development of scientific workflows, based on the strategies captured in G1.
- G3: To evaluate the benefits of the developed architecture in order to support the development of scientific workflows.

1.3 RESEARCH METHODOLOGY

This methodology is divided into proposal definition and proposal refinement as suggested by Mafra et al. (2006). The proposal definition involved doing systematic reviews and mapping studies in order to generate knowledge about the topic of research. Based on this knowledge, the proposal was elaborated. Proposal refinement involved carrying out experimental studies in order to refine the proposal and increase comprehension. The experimental studies followed the methodology proposed by Wohlin et al. (2012) and statistical methods were used to perform quantitative analysis (WILCOX, 2003).

1.4 OUTLINE

The remainder of this work is organized as follows (see Figure 1.1): in Chapter 2 theoretical foundations and a systematic review and mapping of the literature about pragmatic interoperability computational solutions are presented. Based on this systematic review and mapping, the main strategies to implement PI were captured. Based on the strategies captured, an architecture to support PI during the development of scientific workflows is proposed in Chapter 3. Chapter 4 presents an evaluation of the proposed architecture when applied in an open source web-based software ecosystem to support collaborative development of scientific workflows. Final considerations and future works are discussed in Chapter 5.

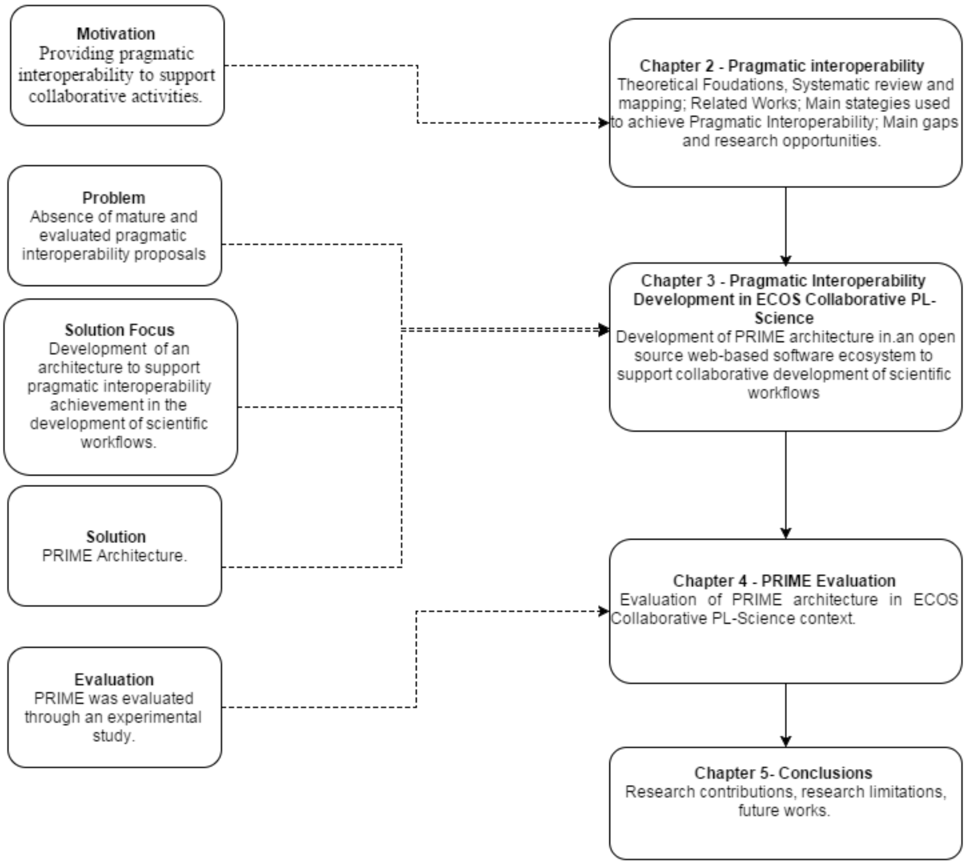

Figure 1.1: Dissertation Structure

2 PRAGMATIC INTEROPERABILITY

This chapter presents a theoretical foundation containing important concepts for the scope of this work and some results of a systematic review and mapping of the literature. The aim is to help the comprehension of the state of art about computational solutions to support pragmatic interoperability (PI).

2.1 INTRODUCTION

This work presents an architecture to achieve PI in order to support the collaborative development of scientific workflows. Before presenting some results obtained through the systematic review and mapping about PI computational solutions, it is important to examine the background to the interoperability concept and the concept of scientific workflow.

This systematic review and mapping was not restricted only to PI computational solutions related to the development of scientific workflows. It aims to help the comprehension of the state of art about pragmatic interoperability (PI) computational solutions and then, to provide support to the proposed work. These include ideas about what can be reused and/or adapted to support the development of scientific workflows.

The following two subsections provide the background on interoperability and on scientific workflow concepts, respectively.

2.2 INTEROPERABILITY

As reported in (ASUNCION; SINDEREN, 2011), there are several definitions for interoperability. IEEE defined interoperability as *"the ability of two or more systems or components to exchange information and to use the information that has been exchanged"* (GERACI et al., 1991). While, ISO defined interoperability as *"the capability to communicate, execute programs, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units"* (ISO, 1993). The Open Group defines it as *"the ability of systems to provide and receive services from other systems and to use the services so interchanged to enable them* *to operate effectively together"* (JOSEY, 2011). Researchers in the enterprise interoperability area defined interoperability as: *"the ability for two systems to understand one another and to use functionality of one another"* (CHEN et al., 2008). Finally, in the context of Service Oriented Architectures, interoperability is understood as *"the ability of the software systems to use each other's software services effectively"* (POKRAEV, 2009). In summary, interoperability allows two or more systems to interact and exchange data, based on standardized methods and to achieve some goal without having to know the uniqueness of the interacting systems (POKRAEV, 2009).

Interoperability can be categorized into levels. These levels are defined in this work scope based on the study of semiotics. They are divided into three dimensions related to the communication process: syntactic, semantics and pragmatics.

To achieve syntactic interoperability support some technologies have appeared in order to propose an standardized structure for data exchange, such as eXtensible Markup Language (XML), HLA Object Model Template (HLA OMT), Interface Description Language, Common Object Request Broker Architecture (CORBA) and Simple Object Access Protocol (SOAP) (WANG et al., 2009).

However, sometimes the interoperated data does not achieve effective interoperability, especially in collaborative systems contexts. As a result, it is necessary to transform data in information in order to address interoperability in a semantic level. The semantic level can be understood as the ability of systems to exchange information whose meaning can be inferred, interpreted and classified, without the need of human involvement (POLLOCK; HODGSON, 2004). To achieve this interoperability level, metadata associated with a certain data set are provided. To accomplish semantic support, some technologies, such as, Web Ontology Language (OWL), Unified Modeling Language (UML) and Resource Description Framework (RDF) (LI; LI, 2004) are used.

Pragmatic interoperability support goes beyond the semantic level. It aims to ensure that individuals involved in the message exchange have the same understanding about the data meaning and the same expectations about the effects caused by the exchange. Therefore, the sent messages always have an intention, in other words, they intend to cause some effect on the receiver. However, the desired effect is not always reflected in the effect caused. Pragmatic interoperability is designed to solve this sort of conflict (POKRAEV et al., 2005). In this work scope, the adopted definition of pragmatic interoperability is *the compatibility between the intended versus the actual use of a service involved in a communication process within a relevant shared context* (adapted from ((ASUNCION; SINDEREN, 2011)).

2.3 SCIENTIFIC WORKFLOWS

A Scientific Workflow is a paradigm used to describe, manage and share complex scientific analysis. These workflows are usually used to run in silico experimental studies supported by computer simulations. The scientific workflows concept is a strategy used in an attempt to minimize the complexity of scientific experiments by focusing on what needs to be done and abstracting how it should be done (ROURE et al., 2009). Despite this, scientific workflows have particularities that increase their complexity, such as (i) a great number of steps, (ii) dynamicity, as scientific workflows frequently change, and (iii) the need for parameterization for many tasks (Costa et al., 2012) (Silva et al., 2011). Furthermore, these workflows usually involve a huge number of heterogeneous datasets and different scientists working in the process (Matos et al., 2008). Datasets and scientists are frequently geographically dispersed.

Scientific workflows are usually composed of external and local services. These services are self-contained and independent and they are responsible for performing a specific task. In order to solve a specific problem, the linkage of these services comprises a scientific workflow. Regarding scientific workflow development, the need of collaboration increases, once the resources/services can be shared in order to reduce workflow development time and cost (ROURE et al., 2009). Scientists must work collaboratively and services must interoperate in order to compose the best services to fulfill scientific workflow requirements. Collaborative systems designed to support scientific workflow development can take advantage of pragmatic interoperability support services.

The following subsection presents systematic review and mapping about PI computational solutions.

2.4 PRAGMATIC INTEROPERABILITY: A SYSTEMATIC REVIEW AND MAPPING OF THE LITERATURE

Kitchenham (2004) emphasizes that an evidence-based approach to software engineering is an important research issue. Evidence-based software engineering (EBSE) highlights the need to find evidence on a specific topic using secondary studies, such as systematic literature reviews and mapping studies as a methodological framework for identifying and aggregating evidences.

The main purpose of a systematic review is to identity, evaluate and interpret studies in literature considering the research questions. Through systematic reviews, it is possible to gather evidence to identify gaps and research opportunities in the targeted area. Systematic mapping is a form of systematic literature review that provides an overview by identifying and categorizing the available research on a broad topic based on the guidelines proposed by Kitchenham (2007).

To the best of our knowledge, there is no systematic review and/or mapping which maps and analyses the existing proposed solutions in literature for providing support to pragmatic interoperability.

2.4.1 SYSTEMATIC REVIEW AND MAPPING OF THE LITERATURE

Through the systematic review, we answered (i) which strategies have been used to enhance pragmatic interoperability and (ii) how the proposed solutions address pragmatic interoperability? Through this review we obtained evidence to build our proposed solution.

This systematic review was organized based on the main activities proposed by Kitchenham (2007): planning, conducting and reporting the study. As a matter of objectivity, we do not focus on the description and detailing of the protocol and activities performed during this review, however, more details can be found in Appendix A. In this section, we present some results and conclusions that were reached to improve the understanding of this research field. This section provides an overview of the proposed solutions to address pragmatic interoperability that were included in our final set of papers.

From 1691 papers (without duplicated ones) collected from electronic databases, only 13 remained at the end of our rigorous analysis, including a quality assessment checklist. This reduction in studies can be justified by: (i) The huge number of false positive studies captured in the electronic bases caused by the use of the word "pragmatic" and its variations in our search string. The word and its variations were frequently used as a synonym of practical, operative or realistic in abstracts and titles. However, we could not remove these words from the search string because they are essential and therefore we had to deal with a huge number of false positives. (ii) The absence of solution proposals to support pragmatic interoperability. Many studies discussed the relevance of pragmatic interoperability without presenting a solution to achieve it. (iii) Finally, the low score in the quality evaluation.

The solutions proposed in these works are associated to different domains (Figure 2.1). Pragmatic interoperability solutions were frequently applied in e-business (46.1%). The set of selected papers also presented pragmatic interoperability solutions to support World Wide Web (15.4%), e-health domain (15.4%), intelligent systems (7.7%), grid computing (7.7%) and geosciences (7.7%). In spite of the fact that the range of years was not limited in this systematic mapping, the selected papers were published from 2004 onwards (Figure 2.2). This was probably due to the fact that the semantic web concept was conceived in 2001 and it was only after this year that the challenges it faced started to be discussed (BERNERS-LEE et al., 2001). The first conceptual model to deal with interoperability issues was proposed in 2003 by (TOLK; MUGUIRA, 2003), motivated by the research agenda of USA Department of Defense Net-Centric Data Strategy. Finally, the term Web 2.0 became widespread in 2004 by O'Reilly Media enterprise increasing interest in interoperability challenges.

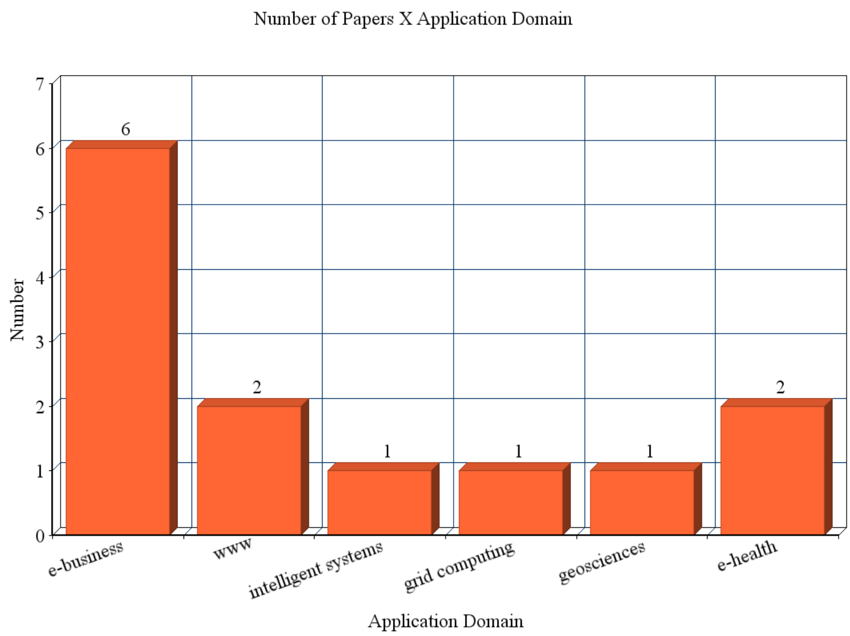

Figure 2.1: MQ4 - Application Domain

In order to answer the question "Which solutions have been used to promote pragmatic

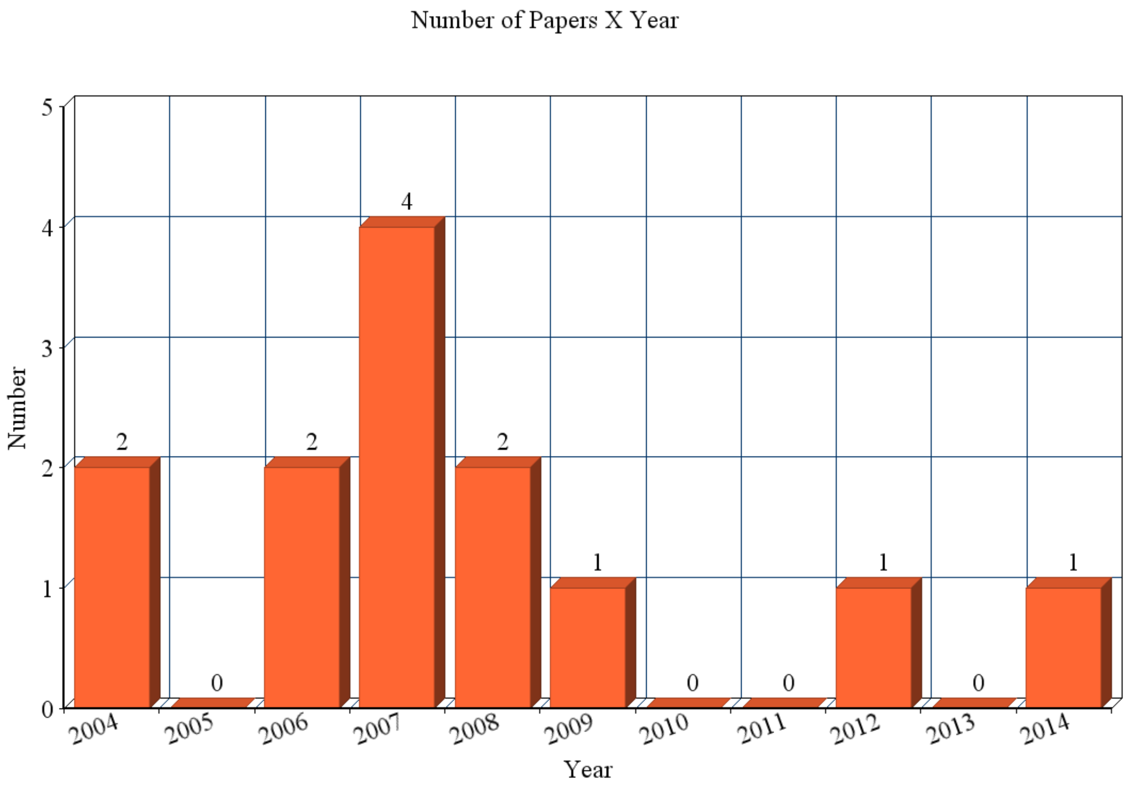

Figure 2.2: MQ1 - Number of Papers X Year

interoperability?" the aspects that were used to promote pragmatic interoperability in the proposed solutions were identified. Most of the selected papers considered service composition, discovery and/or selection (69.2%) and ontologies (69.2%) as essential aspects to achieve pragmatic interoperability support. The remainder of the solutions proposed metamodels (38.5%), software agents (30.8%), pragmatic web services (15.4%) and a pragmatic grid (7.7%).

The high number of papers considering service composition, discovery and/or selection may be explained by the nonfunctional requirements of collaborative systems. The diversity of interactions and the process types where these systems execute imposes flexible rules and policies. In order to deal with this dynamicity, the proposed solutions were usually built using Service Oriented Architecture (SOA) concepts. To reuse and compose with services from different infrastructures, interoperability is a key requirement to be fulfilled. Therefore, these non-functional requirements are important to achieve meaningful collaboration support, and to fulfill them, pragmatic interoperability support is necessary. To achieve pragmatic interoperability across geographically distributed services, techniques

Reference	Service composition, discovery or/and selection	Ontology	Software agents	Pragmatic Web Services	Pragmatic Grid	Meta model
[Ferreira et al 2012]	X	X				
[Mingxin 2009]		X				
[Liang et al. 2007]	X	X	X	X		
[Kutvonen et al. 2007]	X	X				X
[Tolk et al. 2006]	X	X	X			
[Gao et al. 2008]	X		X		X	
[Rukanova 2006]						X
[Tamani & Evripidou 2006]	X					
[Lee et al. 2007]	X	X				
[De Leenheer & Christiaens 2007]		X				
[Kutvonen 2008]	X	X				X
[Lee et al. 2004]	X	X	X	X		X

Table 2.1: RQ1 - Solutions to Promote Pragmatic Interoperability

related to discovering, selecting and composing services that meet user expectations have been developed.

The high number of papers which consider ontologies may be because ontology has been widely explored to represent concepts in the World Wide Web, especially after the beginning of the semantic web in 2001 (BERNERS-LEE et al., 2001). Thus, the use of ontology seems appropriate to represent semantic and pragmatic aspects to tackle pragmatic interoperability issues.

For a better understanding of "How the proposed solutions address pragmatic interoperability", these solutions are described in the following subsection.

2.4.2 PRAGMATIC INTEROPERABILITY PROPOSED SOLUTIONS

This subsection discusses how each proposed solution, presented in the 13 papers, address the issue of pragmatic interoperability.

(FERREIRA et al., 2012)

In (FERREIRA et al., 2012), pragmatic interoperability is used as a way to support tourism business activities. In this domain there is a huge number of heterogeneous variables and resources, such as schedules, accommodation, food, tours and others. This set of variables and resources are usually available as services in the tourism information systems and factors that constrain and impose changes to the initial plan frequently occur. Considering this, tourism services must be pragmatically interoperable in order to assure better alignment between service providers and client expectations and must be able to adapt dynamically to new scenarios without disturbing the clients.

To achieve this support, Ferreira et al. (2012) presented the Open Tourism Initiative (OTI). OTI is an architecture that handles dynamic tourism packages under a Virtual Enterprises organizational model. This architecture was designed to quickly react to external changes behaving as an adaptive information system. In order to meet clients' expectations, an architecture component, named Tourism Pragmatic Engine (TPE), enables pragmatic aspects of the communication to allow tourist involvement as a co creator of the tourism activity. The pragmatic aspects are designed to capture context, preferences, humor and other requirements related to tourists.

OTI services must be described and registered through a domain ontology. As a result, and based on pragmatic aspects, a brokering mechanism must discover, select and compose with the most appropriate services according to clients' expectations.

(TAMANI; EVRIPIDOU, 2006)

Tamani and Evripidou (2006) argued that businesses that adopt service-oriented technology tend to expand their markets and improve quality through interoperation across other businesses. However, in order to fully exploit this potential, business discovery must be effectively implemented. To do this the authors present a methodology for web services discovery that considers pragmatic aspects during the process. As a result, business services can interoperate according to the intended expectations. The proposed methodology defines that the context of web services' usage can be determined considering the context of the collaborators themselves. The contexts of the collaborators are described using a XML standard and, as a result, the context is exchanged and interpreted identically by all collaborators.

(MINGXIN, 2009)

According to Mingxin (2009) enterprise information systems (EIS) are essential to dealing with changing requirements and services in the enterprise domain. However, dynamicity is a complex requirement to be fulfilled by EIS. In this work, pragmatic interoperability was used to address this requirement. The authors proposed an architecture, named Enterprise Isomorphic Architecture (EIA) which supports syntactic, semantic and pragmatic interoperability across EIS. Relying on the core of EIA is Enterprise Isomorphic Mapping Mechanism (EIMM) which is designed to support the pragmatic interoperability level. EIMM is the architecture component that deals with requirements change and transforms these requirements from business to IT systems through mechanisms based on ontologies that keep the transformation isomorphic. This transformation must keep the coherence of concept structures to avoid transaction loss. In order to perform ontology mapping during the transformation process different aspects of EIS must be connected through a precise representation.

(LEE et al., 2007)

Lee et al. (2007) state that geospatial information sources are heterogeneous and geographically distributed. This scenario makes interoperability among different organizations a complex task. In this paper, the authors investigate the integration of real-time water quality assurance data with geographic data. They propose a context-aware geospatial data and service integration framework that includes a syntactic, a semantic and a pragmatic model based on Semantic Web technologies. The framework allows the analysis of existing dependencies, predicting causes and providing context-aware services. The pragmatic model which is necessary to enable pragmatic interoperability captures the context where a service is used and meets user expectations. Based on pragmatic knowledge, services are discovered, selected and composed. As a proof of concept, a Web-based system, named HIS-KCWater (Hydrologic Information System for Kansas City metro WATERsheds) was developed where the proposed framework was deployed. HIS-KCWater is an automated real time analysis and forecasting system for water quality in the watersheds covering the Kansas City metropolitan area.

(LIANG et al., 2007)

In this paper, a pragmatic web service framework was proposed to introduce pragmatic aspects into web service technologies. In this proposed solution, intelligent agents play a key role in managing users' intention, service context, information communication and negotiation between service consumers and service providers. Furthermore, agents must analyze users' requirements to define and rank the semantic equivalent services. Ontologies are used to represent and manipulate web resources.

In order to build the framework, a knowledge frame was designed to store pragmatic information. Pragmatic information is used to capture the intention and the context of each web service in the workflow. The pragmatic frame can be used to find and select the right web service in the web in order to solve a particular task.

(KUTVONEN, 2008),(KUTVONEN et al., 2007)

In Kutvonen (2008) and Kutvonen et al. (2007), the Web Pilarcos project is discussed. Web Pilarcos is a B2B middleware designed to reduce the cost of collaboration and to facilitate management and maintenance of networks. This project was an effort to improve participation in electronic business networks as a key for success. The web-Pilarcos architecture and middleware address the syntactic, semantic and pragmatic interoperability of autonomous business services in an inter-organizational context.

Pragmatic interoperability was used to support meaningful collaboration between enterprises. In this study, pragmatic interoperability was achieved if the participants in the message exchange had compatible business intentions, rules and organizational policies to perform digital business transactions. In order to reach these goals, a business network model and ontologies were used to define an e-Contract (collaboration contract). A B2B middleware, called Web Pilarcos, implemented the e-Contract where facilities for service discovery, selection and composition were provided.

(LEENHEER; CHRISTIAENS, 2007)

In this paper the authors argue that the aim of knowledge sharing in the World Wide Web is to collaborate and integrate communities. Contemporary knowledge engineering methods consider only the non-human system parts, investigating only technical and syntactical aspects in the concept modeling. However, the semantic (elicitation) and pragmatic (application context) must not be neglected.

In order to bridge this gap between "reality" and its modeling concepts, a framework represented through an ontology, called DOGMA, is proposed. The DOGMA ontology aims to provide extension points for engineering methods by covering semantic and pragmatic aspects. The pragmatic aspects were considered to improve the interoperability and therefore collaboration on the web.

(TOLK et al., 2006)

In order to support complex and dynamic environments, and decision challenges, interoperability in modeling and simulation (M&S) field have been considered a key issue for operational research. The authors discuss how various layered composability approaches contribute to the definition of the Levels of Conceptual Interoperability Model (LCIM). Moreover, they discuss how the results can be used to derive implications and requirements for ontologies to describe the universe of discourse. In this universe, intelligent agents serve to mediate communication between agile applications in order to compose individual systems into a meaningful system of systems. LCIM is a framework defined in a previous work which is designed to determine interoperability, including pragmatic interoperability, in the early phases of software engineering process, considering conceptual models.

Furthermore, the paper presents an example of pragmatic interoperability where an Ontology Driven Service-Oriented Architecture was designed. Based on the example, the authors conclude that ontologies are a potential contributor to support the pragmatic interoperability level, and also that this level is relevant to achieve meaningful interoperability.

(GAO et al., 2008)

In Gao et al. (2008), a pragmatic grid architecture (PGA) which extends the semantic Grid is proposed in order to include pragmatic aspects such as context elements and their purpose. As well as this, the authors also propose a pragmatic information framework, architecture of interoperation to support resource management, and a pragmatic-aware discovery service to locate a suitable resource. PGA is a middleware which encompasses pragmatic capabilities. It was extended to provide pragmatic-aware Grid services (PGS), pragmatic provisioning services (PPS), agent services, and pragmatic information service (PIS). In PGA, agents receive inputs related to the state of their environment and act on the environment change, record the usage history, analyze the historical data to learn their behavior and process negotiation between agents.

The proposed solution provides flexible and personalization-awareness resources provision by managing the pragmatic information. Query rules are automatically and dynamically generated, and real requirements are inferred.

(RUKANOVA et al., 2006)

According to (RUKANOVA et al., 2006), to achieve pragmatic interoperability it is necessary to express and compare the requirements at a pragmatic level. To do this, the authors identify the pragmatic interoperability problems that can occur if companies decide to do business transactions electronically by using standards. They propose a metamodel for describing conversations and represent interoperability aspects on a pragmatic level.

In order to describe conversations, it is necessary to define what is communicated and the intention behind the communicated information. The authors use the notions of elementary message (e-message) and action elementary message (ae-message) for this.

The use of the proposed metamodel was illustrated through an example to identify if pragmatic interoperability could be achieved using HL7 standards.

The main benefit of the metamodel is that it can be used to describe the communication requirements of a business transaction and the capabilities of the standard in the same terms and then it can be easily compared. A mismatch between communication requirements of a business transaction and the capabilities of the standard means that some pragmatic interoperability problem has occurred.

(LEE et al., 2004)

Lee et al. (2004) argued that the main goal of the semantic web is to reduce manual discovery and usage of web resources. Software agents are frequently used to automatically discover resources in order to meet user expectations. However, current studies of web services are not wide enough to provide automatic composition.

In this paper the authors propose an architecture of the knowledge processing that is necessary to compose individual services in service flows. Service description knowledge for Web service discovery and composition was divided into syntactic, semantic and pragmatic rules which enable the building of intelligent (pragmatic) web services for automating service compositions.

The models used to describe rules with service concepts allowed the system to identify the relevant rules in a certain domain and to identify and select appropriate Web services for composition. Furthermore, an ontological integration methodology that deals with heterogeneous semantics existing in web service composition was introduced. As a proof of concept, the proposed solution was applied in the cardiovascular domain which required advanced service discovery and composition across heterogeneous platforms of multiple organizations.

(LIU et al., 2014)

In the smart healthcare domain, huge numbers of heterogeneous data sources and humans and the technical factors involved make systems interoperability requirement a complex task. Many challenges such as information collision, heterogeneous data sources, policy obstacles, and procedure mismanagement during the interoperability process must be faced.

This paper presents a pragmatic interoperability analysis of systems integration in an on-going project of a radiology department at a local hospital. The project aims to achieve data sharing and interoperability across Radiology Information Systems (RIS), Electronic Patient Record (EPR), and Picture Archiving and Communication Systems (PACS). Based on the results of the analysis, a pragmatic interoperability framework was proposed. It summarizes the empirical findings and guides the integration process in the radiology context. The framework defined EAI (Enterprise Application Integration) methodology and the HL7 standards, as technologies to address interoperability barriers (http://www.hl7.org/).

2.5 ANALYSIS OF PROPOSED SOLUTIONS

Previously proposed solutions use different approaches in order to achieve pragmatic interoperability in their contexts. However, there are similarities between them.

From the analysis of the selected papers, it was possible to notice that pragmatic interoperability was treated as a key issue to meet user expectations in information systems, as in Ferreira et al. (2012) and Lee et al. (2007). The fulfillment of these expectations was constantly linked to the discovery, selection and composition of the best services. Ontologies, or other metamodels, were frequently used to represent the syntactic, semantic and pragmatic services aspects, according to Lee et al. (2004), Kutvonen (2008) and Kutvonen et al. (2007). Scenarios where pragmatic interoperability support is required, frequently involved heterogeneous, distinct and very large data sources, according to Liu et al. (2014). Pragmatic interoperability usually acts in order to improve communication between services, systems and even institutions. Adding pragmatic aspects to the communication/interoperation process could enhance collaboration activities.

In order to achieve pragmatic interoperability support, pragmatic information must be captured and stored. Pragmatic information can be enriched by context, awareness and historical information, as mentioned by Tamani and Evripidou (2006), Ferreira et al. (2012) and Gao et al. (2008).

In the e-business domain, pragmatic interoperability support often exceeds the system level and reaches the business level, considering business rules, business contracts, and policies, according to Mingxin (2009), Kutvonen (2008) and Kutvonen et al. (2007). We also noticed that the HL7 standards have been used as a base to achieve pragmatic interoperability by adding pragmatic aspects, as in Liu et al. (2014) and Rukanova et al. (2006).

In this vein, the knowledge obtained from this review found that pragmatic interoperability support is achieved when a desired effect is in accordance with the actual effect, considering a message exchange in a relevant and shared context. In other words, pragmatic interoperability support aims to meet users expectations in the collaborative processes. Pragmatic interoperability was commonly associated with a service oriented architecture (SOA). This association can be explained because SOA was proposed as a solution to enable distributed applications development which is an important factor to enhance collaborative systems design. Furthermore, a pragmatic interoperability level can only be reached if syntactic and semantic interoperability support have previously been achieved.

In order to implement pragmatic interoperability across collaborative systems, some common strategies were adopted by the selected papers. Despite the importance of pragmatic interoperability support to reach meaningful collaboration, the set of selected papers in this systematic review and mapping of the literature did not identify empirical studies to evaluate its achievement, thus threatening the validity of the proposals. Therefore, it is important to highlight that presenting a proposal which describes an experimental study seems to be a good contribution to this area.

In addition, none of the aforementioned papers reported presented a pragmatic interoperability solution to support collaborative development of scientific workflows. To improve the development of scientific workflows, there is a need for collaboration considering different systems from different institutions involving tacit knowledge.

In order to implement pragmatic interoperability in collaborative systems, some common strategies were adopted by the selected papers. Based on these strategies, a generic approach to implement pragmatic interoperability can (i) adopt oriented service architecture principles in the collaborative system; (ii) establish and share standards built using ontologies to represent syntactic, semantic and pragmatic aspects of a service, where the pragmatic aspects include context information (e.g. business rules, historical data); (iii) provide mechanisms to enable service composition, discovery and selection based on the syntactic, semantic and pragmatic service aspects stored in the ontologies. An ontology which represents these aspects in conjunction with service composition, discovery and selection mechanisms aim to address users expectations by matching the desired and the actual effect in the collaboration processes. From this generic approach, the proposed solution presented in the next chapter was developed.

Depending on the collaborative system planning and context, other strategies can be included as software agents, for example. In order to obtain more information about this systematic review and mapping of the literature, check Appendix A.

2.6 FINAL CONSIDERATIONS

This chapter has attempted to answer the research question Q1 - Which strategies have been proposed to implement PI in software systems? - by accomplishing the goal G1 - To establish a systematic mechanism to capture the strategies used to achieve PI.

In this chapter, theoretical foundations and some results of a systematic review and mapping about pragmatic interoperability computational solutions were presented. A range of publications related to the area of interest was examined in an unbiased way.

As a sub product of the systematic review and mapping performed in this chapter, a generic model to guide PI implementation in collaborative systems was produced. This model can be checked in Appendix B.

During the systematic literature review and mapping process, it was found that many studies focused on pragmatic interoperability motivation, in other words, they discussed its relevance, features and requirements. However, few studies focused on developing and evaluating a concrete proposal. Through the quality assessment checklist, it is important to highlight that all the selected papers had a low score, very close to the cutoff point (NEIVA et al., 2015a). One reason for this was that none of them presented a rigorous pragmatic interoperability evaluation. For example there were no experiments or case studies in which hypotheses, method, statistical analysis were clearly defined and described. Such findings generate evidence that, despite the fact that many studies indicate the need for pragmatic interoperability support, there are few proposals that have been implemented and evaluated through experimental studies.

Based on the evidence found in this systematic and mapping review of literature, the research was conducted in order to develop and evaluate the proposed solution related to the pragmatic interoperability support during the development of scientific workflows. The next chapter presents the proposed solution.

3 PRAGMATIC INTEROPERABILITY DEVELOPMENT IN ECOS COLLABORATIVE PL-SCIENCE

This chapter presents an architecture for PI (Pragmatic Interoperability) in a collaborative ecosystem named ECOS Collaborative PL-Science. ECOS Collaborative PL-Science infrastructure is also introduced.

3.1 ECOS-COLLABORATIVE PL-SCIENCE

This section introduces ECOS Collaborative PL-Science (FREITAS et al., 2015) environment where pragmatic interoperability has been investigated.

ECOS Collaborative PL-Science is a web-based software ecosystem designed to support researchers activities during the overall scientific workflow life cycle (PEREIRA et al., 2014). Belloum et al. (2011) describe a collaborative life cycle of a scientific workflow as an iterative process divided into four steps: (i) problem investigation, (ii) experiment prototyping, (iii) experiment execution, and (iv) results publication.

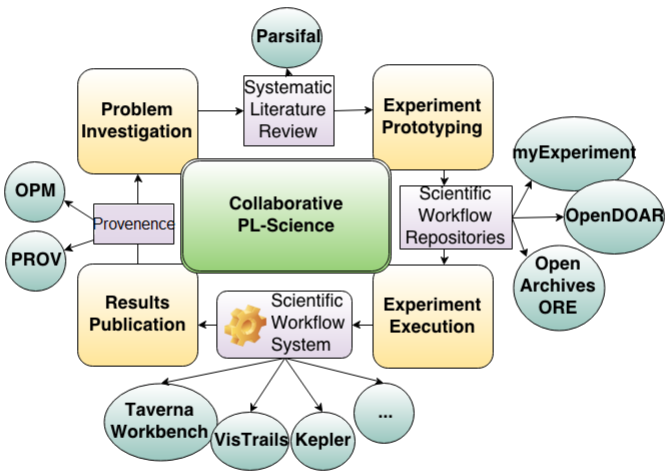

Figure 3.1: ECOS Collaborative PL-Science life cycle (adapted from (Belloum et al., 2011))

Based on this scientific experiment life cycle, an adapted life cycle considering ECOS Collaborative PL-Science context was created (Figure 3.1). During the problem investigation step, the scientist looks for similar experiments, interacts with other researchers, defines his/her goals and breaks down the experiment into smaller steps. In this step, a systematic review can be a tool to support scientists to understand, for example, the problem scope before planning and building his/her experiment. As part of ECOS Collaborative PL-Science ecosystem, a web tool named Parsifal (http://parsif.al) was developed to support systematic review process, where geographically distributed scientists can collaborate. Moreover, to support systematic review process, provenance data are used (ORNELAS et al., 2014) (GASPAR et al., 2015).

In the experiment prototyping step, scientists build an experiment prototype by designing workflows and reusing available components (i) in ECOS Collaborative PL-Science Software Product Line (SPL) or (ii) in open repositories, such as myExperiment Roure et al. (2009), or (iii) developing new components.

Through ECOS Collaborative PL-Science SPL, researchers can explore the core assets and reuse their components in order to produce new products and to contribute with new artifacts during the experiment prototype step. The systematic literature review that can be produced in first step can be linked with the experiment prototype through Parsifal API.

The output of the experiment prototype step is a workflow description file. This output comprises the input of the next life cycle step, named experiment execution. The workflow description file is generated in an intermediate language (BASTOS et al., 2014) which can be converted to different scientific workflow management systems as Taverna, Kepler, Vistrails, among others (WOLSTENCROFT et al., 2013).

In the last step, the researcher analyzes and publishes his/her results and contributions. However, publishing the results only in a scientific paper format can hinder the experiment validation, as well as its reuse and adaptation. In this step, ECOS Collaborative PL-Science enables the storage of information about the experiment process in a detailed way, including experiment steps, execution conditions, input and output data, iterations, results analysis, among other information. Moreover, the generated artifacts are used by data provenance (DALPRA et al., 2015) tasks. As a result, this guarantees experiment quality and enhances reuse. Data provenance also helps in the problem investigation step by providing strategic data to perform a new iteration of a scientific workflow life cycle.

As can be seen in Figure 3.2, ECOS Development environment is a web component where ECOS Collaborative PL-Science code is available, as open source (https://github.com/

pgcc/plscience-ecos)(http://pgcc.github.io/plscience/). As a result, the developer community can contribute through software maintenance and evolution. ECOS Collaborative PL-Science Architecture relies on a Peer-to-Peer Network where different ECOS Collaborative PL-Science instances can communicate. This Peer-to-Peer Network provides a Client Layer Peer that allows Peer Search, Lookup Peer Reference, Artifacts Search and Artifacts Download. Artifacts include services, scientific workflows, systematic literature review documents, datasets, models, among others. The ecosystem is made up of artifacts provided by different instances situated in different institutions, APIs that help the scientific workflow development in its different steps and the open source development environment. Through the APIs, users consume resources to create products and artifacts that add value to the platform. Through the open source development environment, the developer community can extend, adapt and use ECOS Collaborative PL-Science functionalities connected with other applications and applied in different domains. The ecosystem architecture connects services from different sources that must be able to interoperate at a high level in order to achieve meaningful collaboration to support the development of scientific workflows.

ECOS Collaborative PL-Science Architecture is a collaborative environment to support the development and execution of scientific workflows. As mentioned before, PI is considered a key requirement to reach meaningful collaboration. Therefore, this work aims to enhance collaboration in ECOS Collaborative PL-Science. The next section presents a solution to fulfill the interoperability requirement in ECOS Collaborative PL-Science infrastructure.

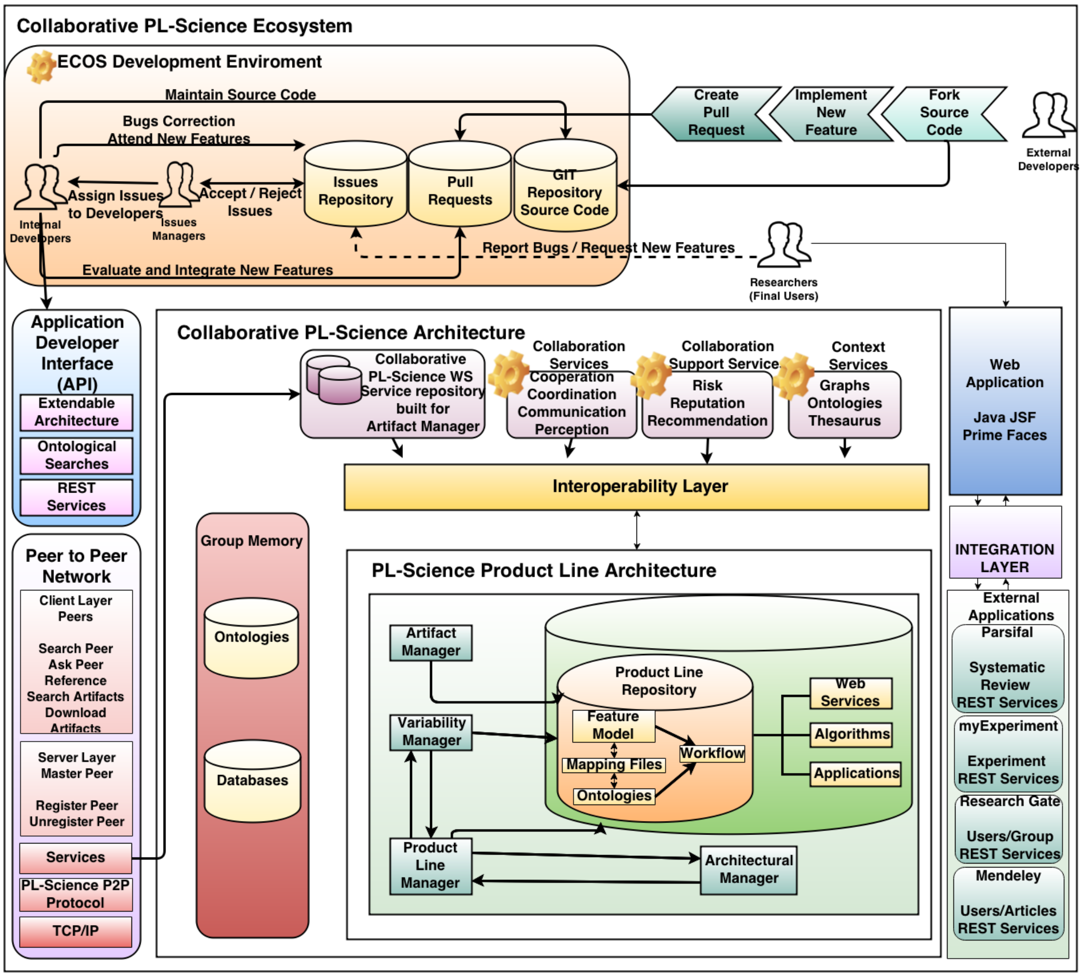

Figure 3.2: ECOS Collaborative PL-Science Architecture (adapted from (PEREIRA et al., 2014))

3.2 PRIME - PRAGMATIC INTEROPERABILITY TO MEANINGFUL COLLABORATION

In this section an approach to achieve PI in ECOS Collaborative PL-Science, named PRIME (PRagmatic Interoperability to MEaningful collaboration) (NEIVA et al., 2015b), is presented.

The strategies adopted in PRIME to achieve PI were defined based on the systematic review and mapping results discussed in Chapter 2 (NEIVA et al., 2015a).

The first step to create PRIME was to analyze the viability of achieving PI support. To accomplish this, the scientists who develop or use functionalities in ECOS Collaborative PL-Science were identified and through meetings, their needs and expectations related with the collaboration process were discussed. The ECOS Collaborative PL-Science environment, including technologies, business rules and policies, architecture issues, among others were also investigated. This investigation revealed that PI would benefit the collaboration process, and technological and organizational aspects would not hinder it achievement.

Using the knowledge obtained when performing viability analysis and the systematic review and mapping, the defined strategies were (i) service composition, discovery and/or selection, (ii) syntactic, semantic and pragmatic aspects representation by the use of ontologies, and (iii) pragmatic web service development to support item (i). The use of these strategies in PRIME is explained in this section.

PRIME was designed to improve collaboration among scientists in scientific workflow development. It focused on experiment prototyping step. This step refers to the construction of the experiment prototype and to the moment when the collaboration must be intense in order to maximize service reuse, discovery, selection and composition.

Before PRIME, experiment prototype in ECOS Collaborative PL-Science context was related to the construction of a generic workflow based on a domain ontology and features model. For example, the scientist could build a scientific workflow related with multiple sequence alignment (COSTA et al., 2013). Based on a sequence alignment ontology and sequence alignment features models, scientists could develop a generic workflow where the tasks to perform multiple sequence alignment were provided considering rules and restrictions defined by the ontology and feature model. In Figure 3.3, for example, ”A” represents possible paths to perform multiple sequence alignment and ”B” represents a workflow instance to perform multiple sequence alignment. In this way, scientists can build different products (generic workflow instances) related to multiple sequence alignment domain in ECOS Collaborative PL-Science SPL.

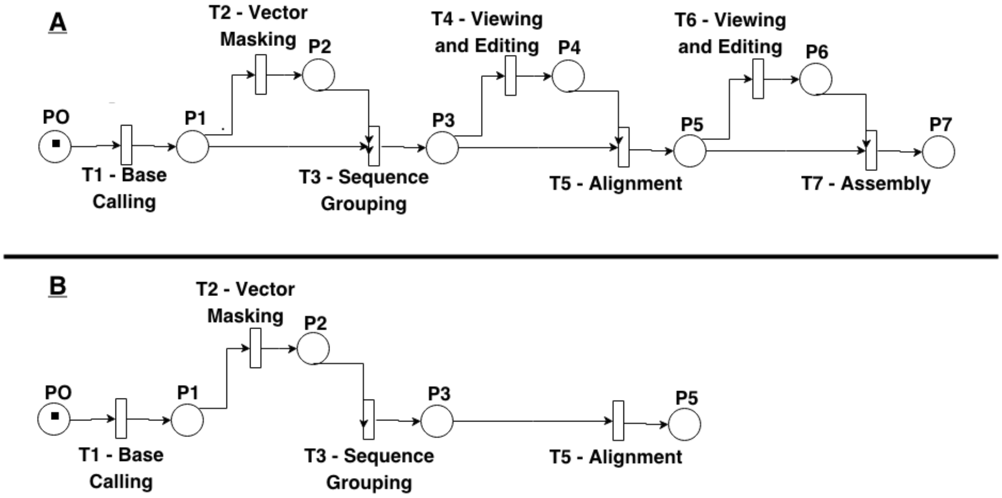

Figure 3.3: Generic Workflow for Multiple Sequence Alignment

This experiment prototype only indicated which tasks should be performed in the developed workflow and the collaboration was supported by services such as Chats. For each task, many services could be used and they must be interoperable in order to compose an executable workflow. However, it was important that services met scientists' expectations. As a result, the existing infrastructure tools (e.g. Chat) were not enough, considering that scientists have to discover services that best fit their desired tasks.

Based on the identified limitations, the PRIME solution was designed to enhance collaborative activities among scientists taking into account tacit and non-formalize context factors, as well as sharing scientists' intentions through semantic and pragmatic interoperability support.

3.2.1 REQUIREMENTS

In this section, functional and non functional requirements are presented. The elicitation process was performed through meetings with the members of ECOS Collaborative PL-Science project. In these meetings the general proposal was presented and the requirements were defined. These requirements were refined using prototyping techniques.

In this way, to address pragmatic interoperability support in ECOS Collaborative PL-Science the following functional requirements (FR) were elicited:

- FR1- Offer to scientists mechanisms to support service discovery: scientists must be able to discover services that meet their expectations.
- FR2 - Offer to scientists mechanisms to support service selection: scientists must be able to select the best service when more than one service meets their expectations. To reach this goal, services ranking based on parameters such as reputation, risk, among others, must be provided.
- FR3 - Offer to scientists mechanisms to support service composition: scientists must be able to perform services composition assuring that they are syntactically, semantically and pragmatically interoperable.
- FR4 - Offer to scientists mechanisms to support the registration of new services: scientists must be able to register new services to be discovered, selected and composed. These services can be located either in ECOS Collaborative PL-Science or in other infrastructures.
- FR5 - Offer to scientists mechanisms to generate an executable scientific workflow as the experiment prototyping output.

In addition, the following nonfunctional requirements (NFR) were elicited as high priority:

- NRF1 - Pragmatic Interoperability: this requirement must be achieved in ECOS Collaborative PL-Science experiment prototyping.

  It is important to state that there are other requirements that PRIME must fulfill regarding its integration with ECOS Collaborative PL-Science environment. The others non functional requirements considered during the development of ECOS Collaborative PL-Science were flexibility, extensibility and scalability. The flexibility requirement is important because this architecture depends on outsourced platforms and services that are constantly evolving. Furthermore, as the open source approach was adopted, the architecture has to support future integrations with other software platforms and the development of new features. ECOS Collaborative PL-Science

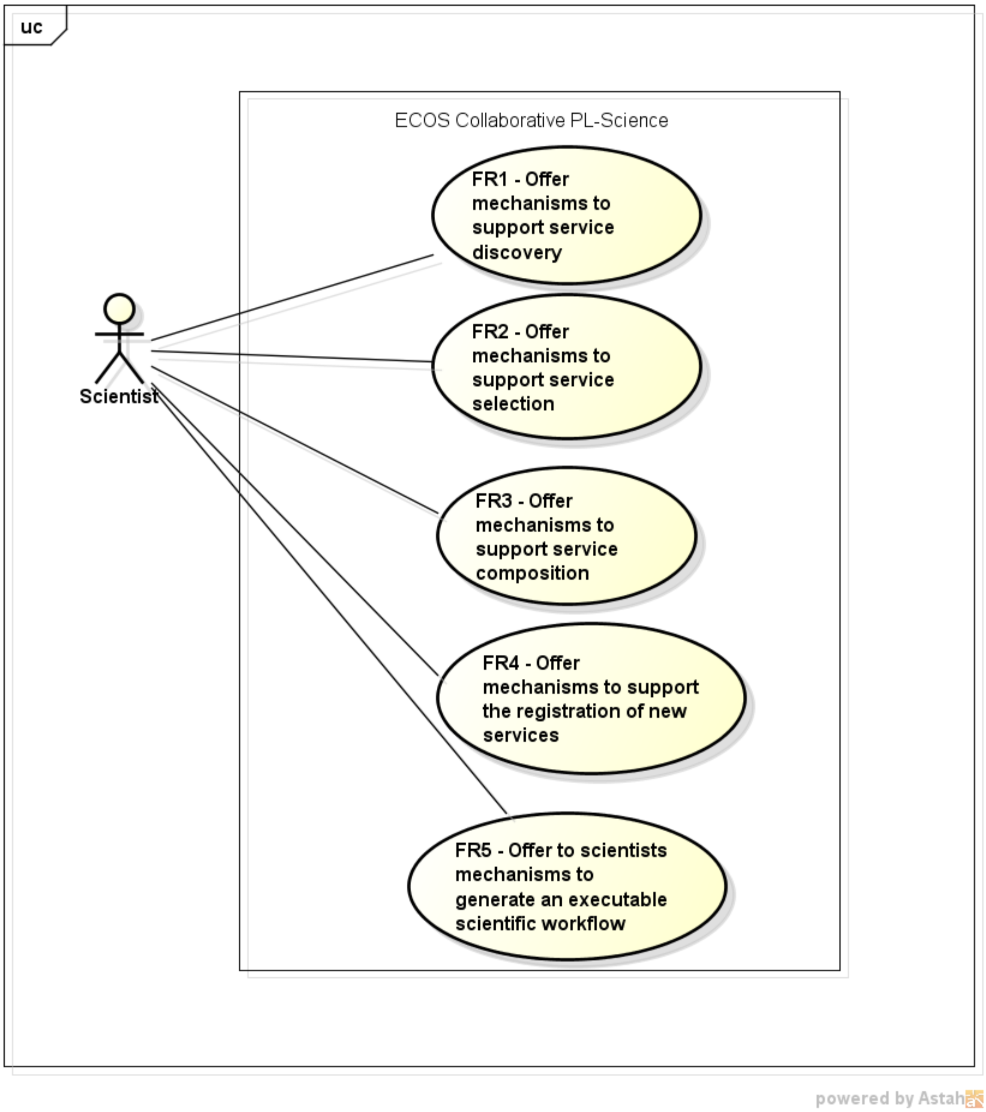

Figure 3.4: PRIME use case diagram

must allow the developer community to extend its features and connect to its architecture. Therefore, extensibility is a key requirement. Finally, the scalability requirement must be fulfilled in order to support the increase in resources.

Flexibility, extensibility and scalability have already been discussed and evaluated in (FREITAS et al., 2015). In order to focus on the goals of this work and also for evaluation purposes (Chapter 4), the pragmatic interoperability nonfunctional requirement was highlighted. In other words, this work considered pragmatic interoperability as its target nonfunctional requirement to be achieved and evaluated.

3.2.2 DESIGN

The PRIME solution brought some changes to the ECOS Collaborative PL-Science Architecture. The changes are highlighted in Figure 3.5, where new components that comprise the proposed solution and their relations are presented. As well as this, in Figure 3.5, some architecture components were omitted and/or simplified in order to focus in the architecture changes.

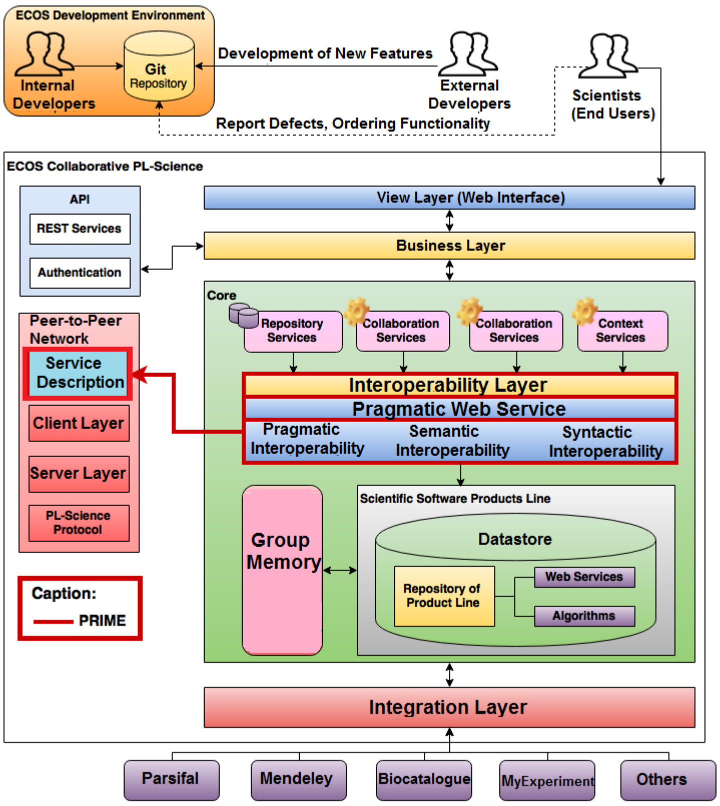

Figure 3.5: PRIME Architecture

The Pragmatic Web Service concept has been used to address pragmatic interoperability issues in different research projects such as (LEE et al., 2004). PRIME Pragmatic Web Service comprises pragmatic, semantic and syntactic interoperability modules. These modules are mainly responsible for indicating if two or more services may interoperate at a certain interoperability level or not. It is worth mentioning that all the communication events that occur in ECOS Collaborative PL-Science pass through the interoperability layer. However, different activities demand the support of different interoperability types. For example if a scientist wants to download an artifact 'A' from ECOS Collaborative PL-Science instance 'B' and he/she is situated in ECOS Collaborative PL-Science 'C', he/she can use only the syntactic interoperability services because this scenario only requires information related with data syntax/format to perform the interoperation process. However, if a scientist wants to discover a service that best fits his/her needs to compose with another service in order to produce a scientific workflow prototype, he/she must use pragmatic interoperability services.

To answer if two or more services may interoperate, the pragmatic web service accesses the Service Description file situated in P2P Network, the group memory and the artifacts in the Data Store. This Service Description file was proposed in this scope solution. The Service Description is an OWL file where registered services are described. Through this file, pragmatic, semantic and/or syntactic information about a certain service can be checked. In addition, the Service Description links the service description to group memory information (e.g. where the service is used, when, who, etc.), to domain ontologies, their mapping files and feature models in product line repository, as well as to the artifact (service) itself. All this connected information enriches pragmatic, semantic and syntactic service aspects in order to support service interoperability. As stated before, the Service Description file is distributed throughout the peer to peer (P2P) network. In this network, each super node contains a set of nodes (ECOS Collaborative PL-Science instances) that includes the services related to a same domain. For example, a super node related to multiple sequence alignment contains nodes related to this domain (CLASSE et al., 2013). This organization is important to avoid that non-related information is considered as a target domain. In order to locate services descriptions, ECOS Collaborative PL-Science instances search for the right super nodes based on this semantic distribution. The Service Description ontology is detailed in subsection 3.2.2.1.

Through the Pragmatic Web Service and Service Description file, and their relationships, we aimed to address NRF1. The architectural changes were designed to promote interoperability considering the services that comprise a scientific workflow prototype.

3.2.2.1 Service Description

The Service Description file was inspired by the work of Kutvonen et al. (2007). This ontology describes a shared standard to represent syntactic, semantic and pragmatic aspects related to a certain service. In addition, OWL-S and SAWSDL (Semantic Annotation for WSDL)were also investigated as a standard to describe services. However, both standards present limitations. OWL-S only supports WSDL. It does not support WADL

(Web Application Description Language) and JSON (Java Script Object Notation). Furthermore, extensions to connect OWL-S with domain ontologies must be built, in advance. SAWSDL does not support WADL nor JSON. Besides, SAWSDL has less semantic power than OWL-S to describe services.

As mentioned earlier, we created the Service Description file which is an ontology stored in the P2P network where all services are registered in super nodes according to their related domain and are available to be accessed by ECOS Collaborative PL-Science instances.

This ontology is able to connect with other ontologies, such as domain ontologies, enriching its semantic information without the need for extensions.

Services registered in ECOS Collaborative PL-Science are described as individuals in the Service Description ontology. This ontology concept was inspired by (ASUNCION; SINDEREN, 2010). Their study presented an analysis of a set of studies included in a systematic literature review about published definitions of pragmatic interoperability. The ontology file can be downloaded at http://goo.gl/KoqYtH. This ontology concept was first designed in a DL Specification format (see Table 3.1). For a textual description, check Appendix C.

In order to describe context domain, an ontology domain can be linked. Other ontologies can be used to describe, for example, the concept represented by semantic output or semantic input. Based on this design, we identified classes, data properties, object properties, restrictions, property chains, etc. Figure 3.6 presents an overview of this ontology through a graph representation.

Furthermore, by using property chains, it was possible to determine which services were syntactically, semantically or pragmatically equivalent. In order to build these property chains, object properties named *includesSyntactic*, *includesSemantic* and *includesPragmatic* and inverse properties, named *syntacticIsIncludedBy*, *semanticIsIncludedBy* and *pragmaticIsIncludedBy*, respectively, were defined. Based on these properties, it was possible to define relations, such as: *"Service A includesSyntactic Syntactic B"* and by inference *"Syntactic B syntacticIsIncludedBy Service A"*. The property chains stated that if *"Service A includesSyntactic Syntactic B"* and *"Syntactic B syntacticIsIncludedBy Service C"*, then *"Service A is syntacticallyEquivalent to Service C"*. Through the presented property chain, the syntactic equivalence among services can be checked. The same idea of the

Service ⊆ Thing
Service ⊆ = 1 includesSyntactic Syntactic
Service ⊆ = 1 includesSemantic Semantic
Service ⊆ = 1 includesPragmatic Pragmatic
Syntactic ⊆ Thing
Syntactic ⊆ ≥ 0 hasParameter Parameter
Syntactic ⊆ ≥ 0 hasReturn String
Syntactic ⊆ = 1 hasName String
Semantic ⊆ Thing
Semantic ⊆ ≥ 0 hasFuncionalRequirements String
Semantic ⊆ = 1 hasSemanticReception String
Semantic ⊆ = 1 hasSemanticRepresentation String
Semantic ⊆ = 1 hasSemanticReturn String
Pragmatic ⊆ Thing
Pragmatic ⊆ = 1 includesContext Context
Pragmatic ⊆ = 1 includesHardware Hardware
Pragmatic ⊆ ≥ 0 hasNonFunctionalRequirements String
Context ⊆ Thing
Context ⊆ ≥ 0 hasInvolved Scientist
Context ⊆ = 1 who Scientist
Context ⊆ ≥ 0 hasComments String
Context ⊆ = 1 hasDomain String
Context ⊆ = 1 how String
Context ⊆ = 1 where String
Context ⊆ = 1 what String
Context ⊆ = 1 when String
Context ⊆ ≥ 0 hasRestriction String
Context ⊆ = 1 hasLicense String
Context ⊆ ≥ 1 hasArtifact String
Context ⊆ = 1 hasReputation String
Hardware ⊆ Thing
Hardware ⊆ = 1 hasCPU String
Hardware ⊆ = 1 hasOperationalSystem String
Hardware ⊆ = 1 hasROM String
Hardware ⊆ = 1 hasRAM String
Scientist ⊆ Thing
Scientist ⊆ = 1 hasCompleteName String
Scientist ⊆ ≥ 1 hasEmail String

Table 3.1: Service Description Specification

presented property chain was established to define semantic and pragmatic equivalences, as can be seen in Figure 3.7.

Although the defined property chains help to answer if two or more services may interoperate, they are not enough. The defined property chains are able to identify if different services are linked to the same syntactic, semantic or pragmatic description. However, they are not able to identify similar syntactic, semantic or pragmatic descriptions. In order to reach this goal a ranking algorithm was developed to determine how similar service descriptions are. This ranking algorithm is presented in the next subsection.

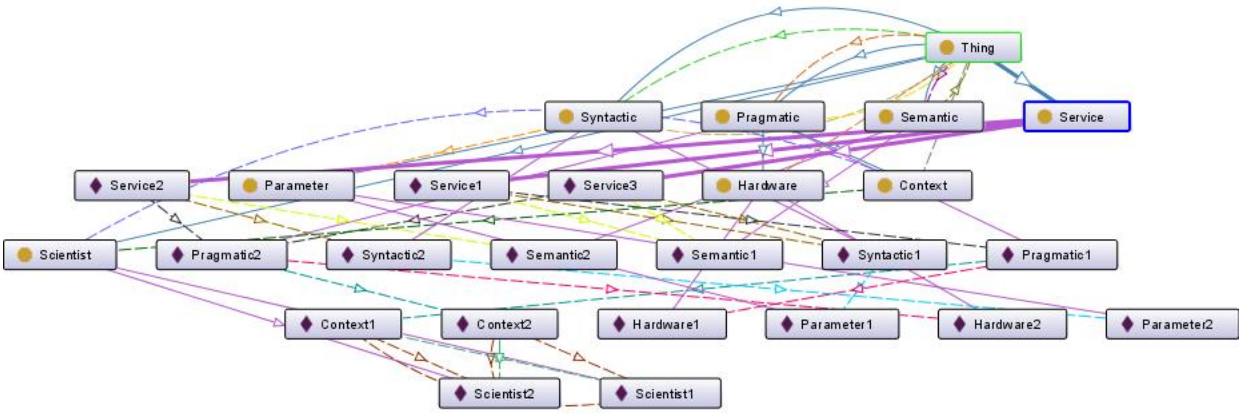

Figure 3.6: Services Description Ontology Graph Representation

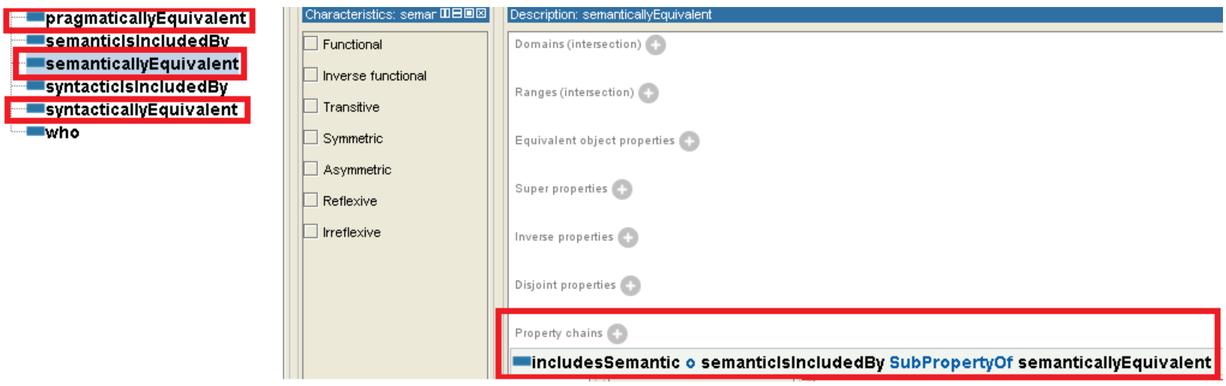

Figure 3.7: Property Chains

3.2.2.2 Ranking Algorithm

The ranking algorithm was developed in ECOS Collaborative PL-Science in order to determine how similar service descriptions are and then, to support service discovery strategy. This ranking algorithm is explained along this subsection.

The Service Description ontology data properties were divided into four categories and analyzed individually:

- Strings
- Sentences
- Texts
- Concepts

Furthermore, three categories of data properties similarity analysis were defined:

- Syntactic
- Semantic
- Pragmatic

Strings, sentences and texts were submitted to syntactic and semantic analysis, while concepts were submitted to semantic and pragmatic analysis.

The syntactic analysis was performed by applying the ”levenshtein distance”, that is an algorithm for measuring the difference between two sequences (SCHEPENS et al., 2012). This algorithm was adapted to strings, sentences and text similarity measurement by a project named HultigLib (http://www.di.ubi.pt/ jpaulo/hultiglib/). The HultigLib is a free and open source library with a set of text processing tools, written in Java language that has already been used in different research projects such as bioinformatics (CORDEIRO et al., 2007), (BURROWS et al., 2013).

It is important to state that the purpose of this dissertation is not to investigate the best algorithm to perform the similarity calculations in ECOS Collaborative PL-Science environment. The techniques were chosen because they were available and are well known but they can be replaced by others. Replacement can be easily performed. The HultgLib library provides other algorithm options, so it is only necessary to change the called method related to ”levenshtein distance” in the code to another method. Other options are to extend the HultgLib library by including new algorithms or even swap the library. The aim of this work is to propose, to implement and to evaluate an architecture to achieve PI.

The semantic analysis was performed verifying the matching of synonyms and heteronyms of different descriptions based on queries to a lexical database, named WordNet (MILLER, 1995).

The pragmatic analysis was performed based on complex networks. Complex networks are graphs that present non-trivial topological features. These features not occur in simple networks such as random graphs, but often occur in graphs modeling real systems (GANGULY et al., 2009). ECOS Collaborative PL-Science was linked to a Human Disease Network proposed in Goh et al. (2007). This network was used because we decided to conduct experiments related to this domain and also, because the network is reliable, available and was used in relevant published research, such as (BARABÁSI et al., 2011). The Service Description ontology data properties were linked to concepts in the complex network, for example, if the service domain is described as anemia, this concept will be related with the same complex network concept, in which close concepts are also considered in the service similarity analysis. The similarity between the two concepts is calculated based on the distance between them in the complex network graph.

In this way, the ranking similarity compares different services based on the standard established by Service Description ontology. This comparison can be described as a function that sums syntactic analysis (SYA) plus semantic analysis (SEA) plus pragmatic analysis (PRA) (Figure 3.8). The comparisons that obtain the lowest results are considered more similar than those that obtain the highest results. Based on this criterion the services are ranked.

Similarity = SYA + SEA + PRA

Figure 3.8: Similarity Function

3.2.2.3 Pragmatic Web Service

The engine to enable service selection, discovery and composition relies on the pragmatic web service. This strategy is crucial to allow ECOS Collaborative PL-Science users to meet their expectations in the collaboration process. As can be seen in Figure 3.9, the pragmatic web service is organized in different components. The components were inspired in the work developed by Liang et al. (2007) which was discovered in the systematic review and mapping presented in Chapter 2.

The communication component is responsible for (i) getting the user request from the user interface; (ii) translating the request into Service Description Ontology Standard; (iii) sending the request to the negotiation component for further analysis; (iv) contacting the web service component to proceed with the service invocation.

The negotiation component searches for historical information, service description ontology, complex networks, among others in the knowledge repository.

The web service component is responsible for (i) performing the similarity calculations among services; (ii) ranking the services based on the similarity calculations; and (iii)

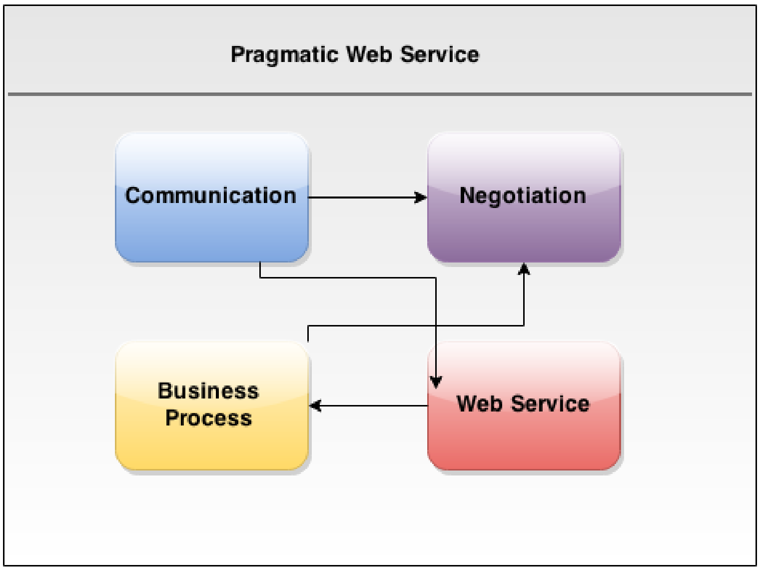

Figure 3.9: Pragmatic Web Service Components

considering information such as the reputation of services and developers.

Finally, the business process component analyzes the dependency relationship between services based on the information provided by the negotiation component in order to recommend, or not, a certain service composition.

3.2.3 IMPLEMENTATION

PRIME solution has been iteratively developed. In order to demonstrate the scientists' view considering the steps usually followed by them and the implemented functionalities, an example was created. Let's suppose a scientist wants to develop a scientific workflow prototype through ECOS Collaborative PL-Science. The scientist designs the workflow prototype and chooses the best services to perform each task defined. In Figure 3.10, the user interface that the scientists use to inform the number of scientific workflow stages and which services will be related to each stage, can be seen. In this way, FR3 was achieved.

By clicking on the search button at the end of the stages, the user interface, as illustrated in Figure 3.11, is displayed. To help in the service definition activity, the scientist uses the service discovery functionality offered by PRIME solution. This functionality allows the inclusion of desired parameters related to the service searched. These parameters cover syntactic (e.g. name and return), semantic (e.g. functional requirement) and pragmatic aspects (e.g. context). Figure 3.11 shows a part of the user interface where the search parameters can be inserted by the scientist. In this way, FR1 was achieved.

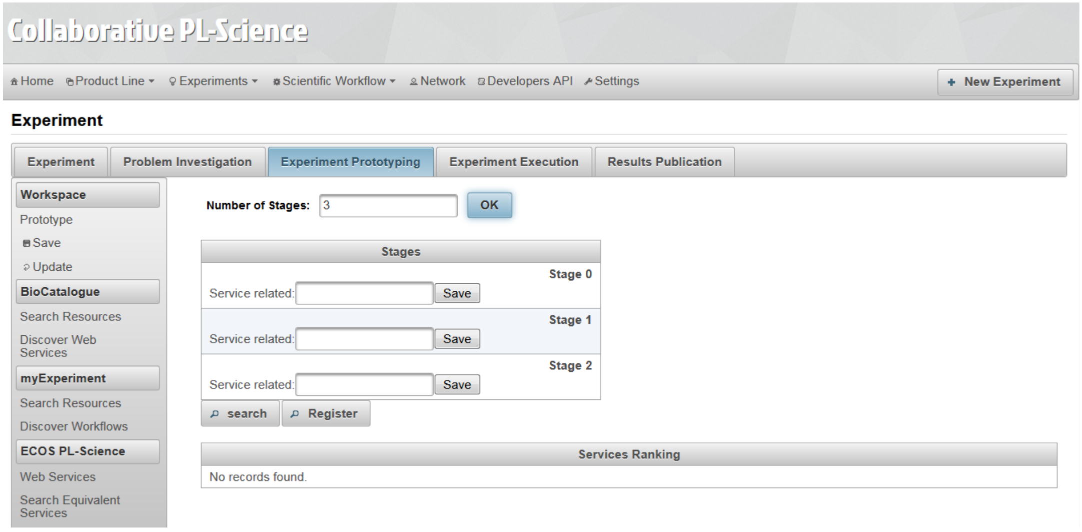

Figure 3.10: Experiment prototyping step in ECOS Collaborative PL-Science

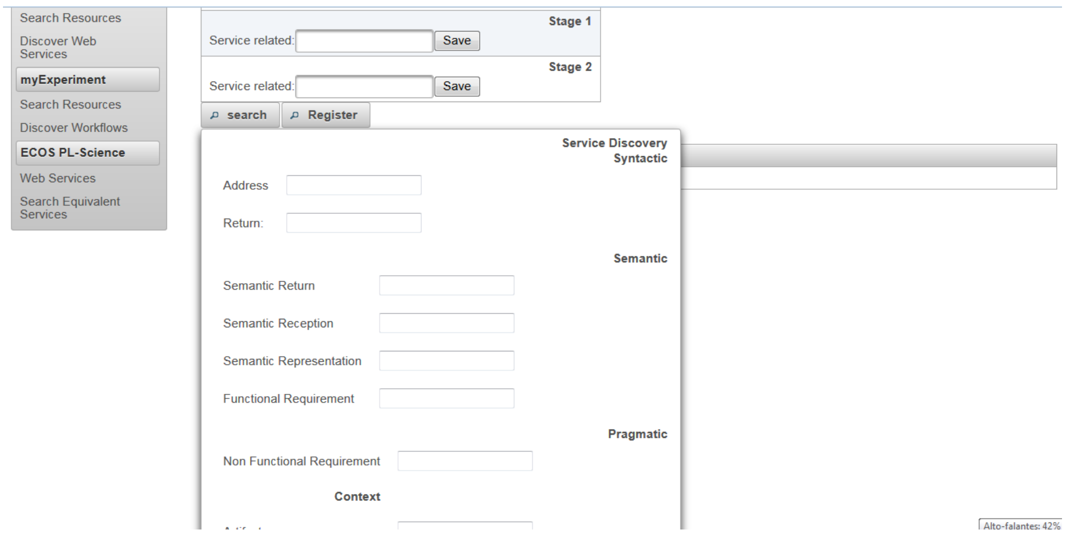

Figure 3.11: Service Discovery in ECOS Collaborative PL-Science

Sending the search information in the user interface, as illustrated in Figure 3.11, PRIME architecture processes the search as follows:

1) It finds the right node in the semantic peer-to-peer super nodes that organize ECOS Collaborative PL-Science environment.

2) It processes the data provided by the scientist based on the service description ontology standard. In other words, PRIME converts the scientist search request to the stored services standard in ECOS Collaborative PL-Science instances.

3) It analyzes the Service Description ontology file associated to the related node in the semantic peer-to-peer network.

4) It performs the similarity calculation among the recovered services and the search

parameters inserted by the scientist.

5) It orders and presents the potential interoperable services to the scientist in a ranking. The ranking is generated based on the similarity calculation. In this way, FR2 was achieved.

Based on the ranking provided through PRIME, the scientist analyzes the returned services and defines if one of them is suitable or not for the task 3.12. At this moment, if necessary the scientist can get in touch with who created or used the service to ask questions and get more information.

Figure 3.12: Services Ranking

It is important to highlight that the whole PRIME process is transparent to the scientist. The scientist is limited to provide the service desired requirements and choose the best service based on the generated ranking services. A summary of the steps followed by PRIME solution can be found in Figure 3.13.

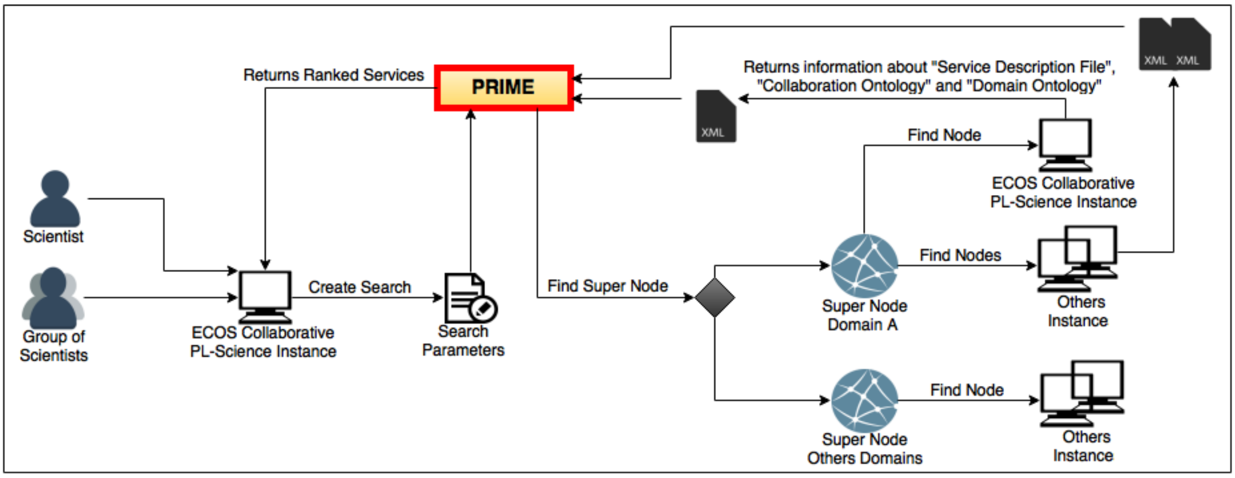

Figure 3.13: Summary of the steps followed by PRIME

The search must be performed for each task defined in the scientific workflow prototyping. If the scientist does not find any suitable service for the task, he/she can develop it and register it in ECOS Collaborative PL-Science. New service registration occurs in a similar interface to the one presented in Figure 3.14. As illustrated in this figure, the information about the new service is inserted and then, the scientist clicks on a register button in order to finish the registration process. Based on the inserted service information, the Service Description ontology is updated with a new individual (service) addressing FR4.

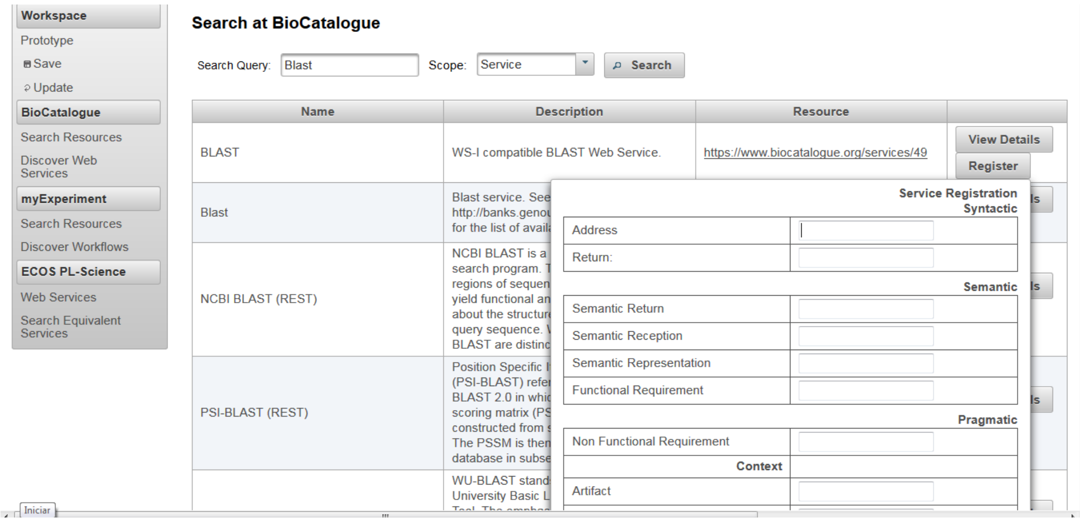

Figure 3.14: Service Registration in ECOS Collaborative PL-Science

On finishing the experiment prototype step, the scientist can export an executable workflow file, addressing FR5. The export process uses an intermediate language proposed in (BASTOS et al., 2014). This intermediate language allows that workflow prototypes can be translated to different scientific workflow management systems and be executed by them. This executable file can also be linked to Taverna (http://www.taverna.org.uk/), Kepler (https://kepler-project.org/) or Vistrails (http://www.vistrails.org/) - scientific workflow management systems - enriching the ecosystem in ECOS Collaborative PL-Science.

At the end of the implementation process, the ECOS Collaborative PL-Science members attended a meeting where the functional requirements were validated. They considered that the functional requirements were implemented properly. The interface prototypes developed during the implementation process supported the requirements validation.

3.3 FINAL CONSIDERATIONS

In this chapter an architecture to support PI was proposed and implemented in a software ecosystem named ECOS Collaborative PL-Science. The proposed solution considered the knowledge obtained through the systematic review and mapping presented in Chapter 2.

This chapter attempted to answer the research question Q2 ”How to create an architecture to implement PI to support the collaborative development of scientific workflows?” and Q3 ” What is the feasibility of the resulting architecture?” by accomplishing the goal G2 ” To develop an architecture to achieve PI in a collaborative system that supports the development of scientific workflows, based on the strategies captured.”

Despite the fact that the implementation of the proposed architecture in ECOS Collaborative PL-Science has provided evidence about its feasibility, it is necessary to conduct experimental studies in order to evaluate the defined non functional requirement. The next chapter presents an evaluation of PRIME.

4 PRIME EVALUATION

This chapter presents an evaluation of PRIME architecture that was implemented in ECOS Collaborative PL-Science. It presents the methodology used to evaluate the hypotheses, and finally, the results of these evaluations.

4.1 EXPERIMENTAL STUDY DEFINITION

In order to evaluate the functional requirements mentioned in subsection 3.2.1 a range of test cases were defined. The obtained results were positives and generated evidence that functional requirements were fulfilled. On the other hand, the defined nonfunctional requirements were evaluated through an experimental study.

The evaluation scope was defined based on GQM method (SOLINGEN et al., 2002) as following:

Analyze PRIME **for the purpose of** evaluation **with respect to** pragmatic interoperability **from the point of view** of users when developing scientific workflows in the context of ECOS Collaborative PL-Science.

From the defined scope, the research questions were derived as:

- Q1. Does PRIME architecture implies in a better PI support if compared with its non-use?

Based on this question it was investigated if PRIME increases pragmatic interoperability support.

Related with the research question, other three secondary questions were also continually under evaluation focus:

- Does PRIME architecture implies in a better efficiency to develop scientific workflows if compared with its non-use?
- Does PRIME architecture implies in a better workflow prototyping experience if compared with its non-use?

In this vein, an experiment that was planned in order to answer these questions is presented in the next section.

4.2 EXPERIMENTAL STUDY PLANNING

According to Wohlin et al. (2012), the experiment planning is a process that defines how the study will be conducted. The author divides the planning phase in seven steps: (i) context selection - can be categorized in four dimensions as, offline vs. online, students vs. professionals, real vs. toy, specific vs. generalized; (ii) hypotheses definition, which can be categorized as null or alternative. Usually, the null hypotheses are those we want to reject, while the alternative hypotheses are those we want to accept; (iii) variables selection, in which the variables are categorized as dependents and independents. The independent variables are those ones we can control and modify, while the dependents are affected by changes in the independent variables; (iv) subjects selection, that comprises the strategies used to define the population sample that will participate in the study; (v) experiment design, that defines how the tests will be organized and performed in the study; (vi) instrumentation, that aims to provide mechanisms to perform and monitor the study and (vii) validity evaluation, in order to develop strategies (whenever possible) to mitigate the validity threats.

The execution of each planning step in PRIME context is presented below:

Context Selection. It was defined that the experiment would be performed on-line. The selected subjects were defined as students and bioinformatics experts that worked in problems specially modeled to this study (toy problems).

The toy problem was modeled considering the study of inheritance of blood types in biology field. The problem consisted in calculate the probability of child inherits a certain blood type from his/her parents.

This inheritance problem can be divided into three activities: (1) the decoding of parental blood type phenotypes in genotypes, (2) the combination of parental genotypes and finally, (3) the calculation of child blood type probability (see Figure 4.1)

Regarding the aforementioned problem and its activities, the related workflow was prototyped in ECOS Collaborative PL-Science context (Figure 4.2). As can be seen in Figure 4.2, the workflow contained one stage for each defined activity. ECOS Collaborative PL-Science provided 15 registered services related to blood type domain. The blood type domain is a concept in the Human Disease network that was presented in Section 3. From these 15 services, 5 services were related to activity 1, 5 services related to activity 2, and the remained 5 services related to activity 3. The best services to perform the activities

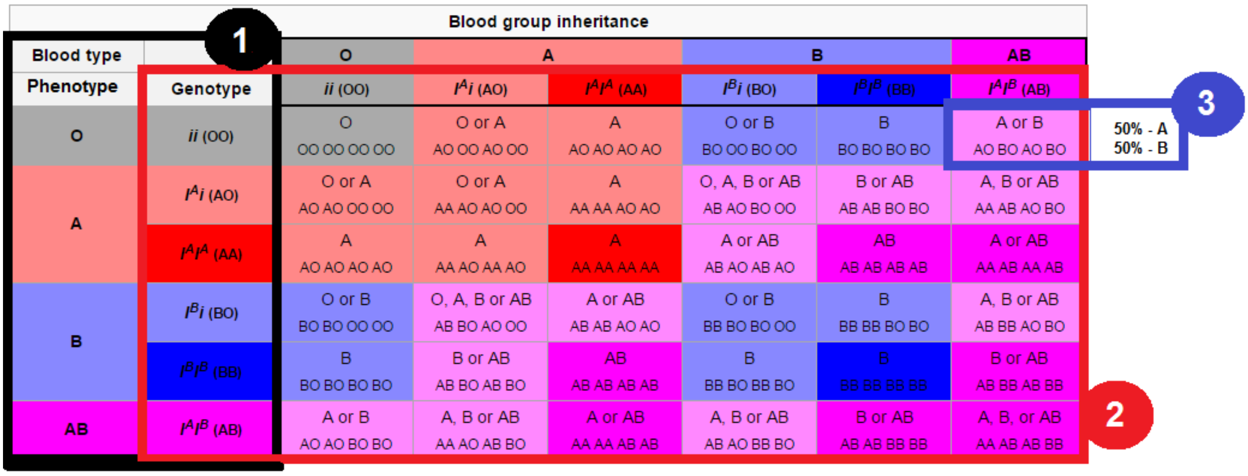

Figure 4.1: Blood group inheritance activities

should be found considering this set of 15 services available in the ecosystem.

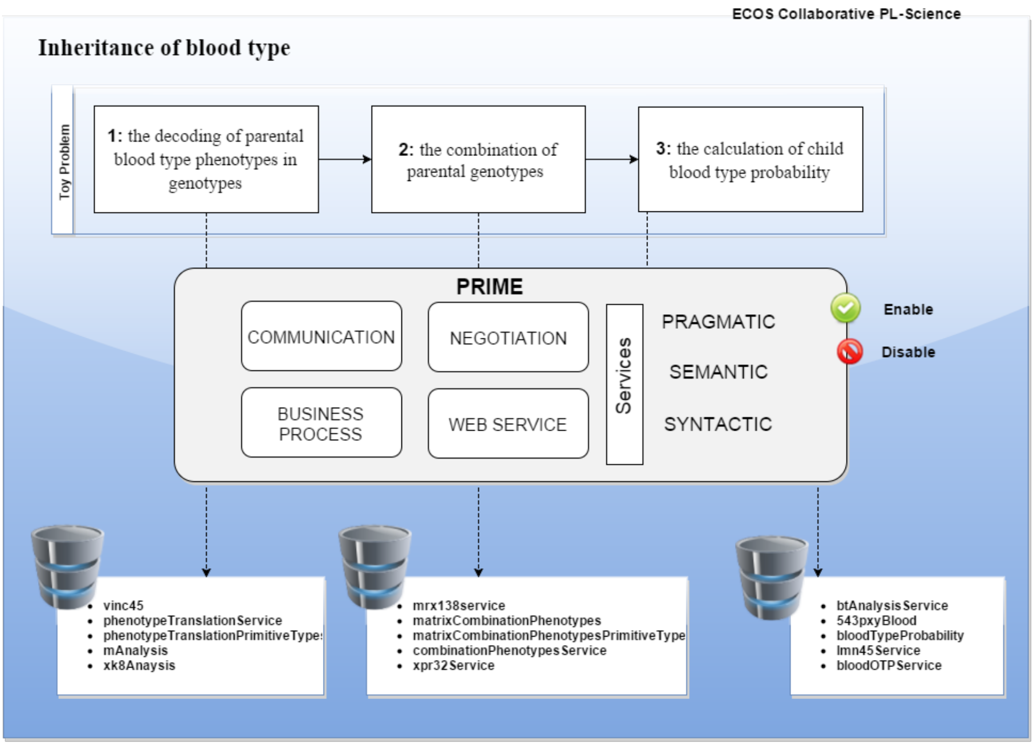

Figure 4.2: Blood type inheritance workflow

Considering the workflow detailed in Figure 4.2, we defined our hypotheses that will be presented in the following paragraphs.

Hypotheses Formulation. It was defined 3 null hypotheses and 3 alternative hypotheses as follow:

The null hypotheses were:

- (H0a) PRIME does not enhance pragmatic interoperability during scientific workflow prototyping in ECOS Collaborative PL-Science.
- (H0b) PRIME does not enhance efficiency during scientific workflow prototyping in ECOS Collaborative PL-Science.
- (H0c) PRIME does not enhance a better scientific workflow prototyping experience in ECOS Collaborative PL-Science.

The alternative hypotheses were defined as follow:

- (H1a) PRIME enhances pragmatic interoperability during scientific workflow prototyping in ECOS Collaborative PL-Science;
- (H1b) PRIME enhances efficiency during scientific workflow prototyping in ECOS Collaborative PL-Science;
- (H1c) PRIME enhances a better scientific workflow prototyping experience in ECOS Collaborative PL-Science;

Variables Formulation. This experiment was conducted using the following dependent variable: pragmatic interoperability, efficiency and user satisfaction. All dependent variables were analyzed independently between them. As independent variable, we can mention: the use of PRIME. In other words, the dependent variables were analyzed considering a task performed ad hoc and a task supported by PRIME architecture in ECOS Collaborative PL-Science. In the task performed ad-hoc, the PRIME functionalities were disabled in ECOS Collaborative PL-Science.

Pragmatic interoperability was measured as the percentage of correctness in the workflow prototyping task. Efficiency was measured as the time spent (minutes) in the workflow prototyping task. Finally, satisfaction in use was investigated using qualitative data and direct observations.

Selection of Subjects. It was selected subjects that were at least attending college.

Experiment Design. The design type of this experiment was one factor with two treatments for each hypothesis.

(FO)The factor was the percentage of correctness in the workflow prototyping task and the treatments were the use and non use of PRIME.

In this vein, two different tests were performed:

T1: The percentage of correctness in the workflow prototyping task considering the use of PRIME.

T2: The percentage of correctness in the workflow prototyping task considering the non use of PRIME.

(F1)The factor was the time spent in the workflow prototyping task and the treatments were the use and non use of PRIME.

In this vein, two different tests were performed:

T1: The time spent in the workflow prototyping task considering the use of PRIME.

T2: The time spent in the workflow prototyping task considering the non use of PRIME.

(F2)The factor was the satisfaction in the workflow prototyping task and the treatments were the use and non use of PRIME.

In order to investigate user satisfaction in the workflow prototyping task considering the use and non use of PRIME, the subjects opinions expressed in questionnaires were analyzed .

Instrumentation. In order to perform this experimental study, the following instruments were selected:

- Consent to the subject, in order to allow the publication of the collected data in this work (see Appendix D).

- Subject characterization questionnaire.

- Training material, in order to support the training session performed with the subjects before the experiment beginning.

- Document containing the explanation of the task that was proposed to the subjects (see Appendix E).

- Follow-up questionnaire, in order to collect data after the end of the subjects accomplishment of the proposed task (see Appendix F).

- ECOS Collaborative PL-Science environment was available online (http://nenc.ufjf.br/plscience/), in order to subjects accomplish the proposed task.

4.3 EXPERIMENTAL STUDY OPERATION

4.3.1 PILOT EXPERIMENTAL STUDY

4.3.1.1 Goals

The goal of this pilot experiment was to reveal improvements that could be done and eventual flaws that must be corrected before the experiment conduction. The experiment aims to evaluate PRIME with respect to pragmatic interoperability, efficiency and user satisfaction, during the workflow prototyping in ECOS Collaborative PL-Science context considering its use and non-use.

4.3.1.2 Subjects Characterization

Three subjects attended in this pilot study. These subjects were selected because they were master students in Computer Science that worked in ECOS Collaborative PL-Science project. In this way, they could contribute to discover flaws and improvements and be more comprehensive with technical problems that could appear.

4.3.1.3 Scenario

Firstly, the subjects were asked to read and sign the consent and next to answer the characterization questionnaire. Then, they attended a training session. This training presented the ECOS Collaborative PL-Science project, in order to discuss its goals and briefly its infrastructure. Once the experiment was performed in ECOS Collaborative PL-Science, it was important to provide some context information. Also, this training presented information related with the proposed tasks. In the proposed tasks the subjects should prototype a workflow to find the inheritance probability of a baby blood type regarding the parental blood types. In this way, the training discussed concepts related to blood type inheritance, the number of workflow stages needed to perform the proposed tasks and what each stage was responsible to perform. The information provided during the training can be considered equivalent to the information found during the problem investigation step of the workflow life cycle.

After accomplished the training session, the experiment study started. The tasks that were performed by the subjects was provided in tutorial form. In this way, the steps to perform the tasks were completely explained to the subjects, in order to reduce, as much as possible, interferences during tasks performance. The first task was performed ad-hoc (without PRIME use) and the second task was performed with PRIME support. The tutorials to perform the tasks can be checked in Appendix E.

Based on the information presented in the training session, in both tasks, the subjects were asked to develop a workflow prototype with 3 stages. Each stage should perform the following activities: (1) the decoding of parental blood type phenotypes in genotypes, (2) the combination of parental genotypes and (3) the calculation of child blood type probability.

In the first task the subjects discovered, selected and composed the services registered in ECOS Collaborative PL-Science using an ad-hoc method. Basically, the subjects read the description of registered services in order to choose those ones considered as best services to compose the proposed workflow prototype. There were 15 services registered in ECOS Collaborative PL-Science. For some tasks in bioinformatics domain, the number of available services can be potentially higher. However, considering the blood type inheritance modeled task and aiming to avoid an experiment too exhaustive to the subjects, the amount of 15 services was selected.

In the second task, the subjects were supported by PRIME (see Chapter 3, specially section 3.2.3). Once again, the subjects were instructed to choose those services that they considered as best services to compose the proposed workflow prototype.

4.3.1.4 Lessons Learned

The direct observations and the answers to the questions 'Which technical problems occurred during the task?' and 'Were the training session and the instruction provided enough to perform the proposed task? Justify.' in the follow-up questionnaires were used to provide lessons, in order to improve the experimental study.

Based on this source of evidences some improvements were made:

- Small adjustments in the questionnaires to improve accuracy.

- Technical problems in the application caused by Java libraries conflicts were identified and fixed.

- The following affirmative was added in the follow-up questionnaire: "PRIME supports service discovery, selection and composition considering the experiment prototyping step".

- The training was adapted to include a better explanation about workflow concept. A tour through ECOS Collaborative PL-Science web application and its functionalities was also included.

4.3.2 EXPERIMENTAL STUDY

4.3.2.1 Goals

This experiment aims to evaluate PRIME with respect to pragmatic interoperability, efficiency and satisfaction during the workflow prototyping in ECOS Collaborative PL-Science by considering its use and non-use. The improvements and flaws revealed in the pilot experiment were considered and fixed to this experiment.

4.3.2.2 Subjects Characterization

Thirty five subjects attended to this experiment. From these subjects, 31 were non-professionals mostly from Computer Science or related areas and 3 subjects were bioinformatics experts. Among these subjects, the academic degree was divided into (i) 2 doctors, (ii) 24 master students, (iii) 1 graduated and (iv) 8 undergraduate students. All of them were over 18 years. In order to select the subjects no specific knowledge was demanded. The only requirements were to be over 18 years and be at least attending college.

4.3.2.3 Scenario

The scenario of this experiment was the same of the pilot experiment. As in the pilot experiment, the subjects accomplished the first task, and then answered to the first follow-up questionnaire. After, the subjects accomplished the second task and answered the second and last follow-up questionnaire.

4.3.2.4 Lessons Learned

As in pilot study, this experiment used the direct observations and the answers to the questions 'Which technical problems occurred during the task?' and 'Were the training session and the instruction provided enough to perform the proposed task? Justify.' in the follow-up questionnaires to identify lessons learned.

Through this source of evidence, it was noticed that the adjustments performed to this experiment had worked. The technical problem occurred in the pilot study was fixed. However, some minors usability issues that can be improve in the next version of ECOS Collaborative PL-Science web application were identified. The issues were about button names and disposition in the screen. Auto-complete facilities during the insertion of search parameters were also considered as improvement to the next version. Other suggestions presented by the subjects must be discussed and analyzed with all ECOS Collaborative PL-Science project members.

4.4 ANALYSIS AND INTERPRETATION

As aforementioned, this experimental study aimed to investigate the following alternative hypotheses:

- (H1a) PRIME enhances pragmatic interoperability during scientific workflow prototyping in ECOS Collaborative PL-Science;
- (H1b) PRIME enhances efficiency during scientific workflow prototyping in ECOS Collaborative PL-Science;
- (H1c) PRIME enhances a better scientific workflow prototyping experience in ECOS Collaborative PL-Science;

In order to measure pragmatic interoperability mentioned in H1a, the percentage of correctness in the workflow prototyping task was analyzed. The efficiency mentioned in H1b was measured considering the time spent to perform the workflow prototyping task. Finally, the scientific workflow prototyping experience mentioned in H1c was analyzed based on qualitative data.

4.4.1 QUANTITATIVE ANALYSIS

This quantitative analysis aims to investigate the H1a and H1b hypotheses. The first step was to identify the outliers. To perform this activity, Grubbs test was used with 5% of significance.

For the variable 'percentual of correctness' considering the sample to group 1 (PRIME non use) and group 2 (PRIME use), the Grubbs tests were:

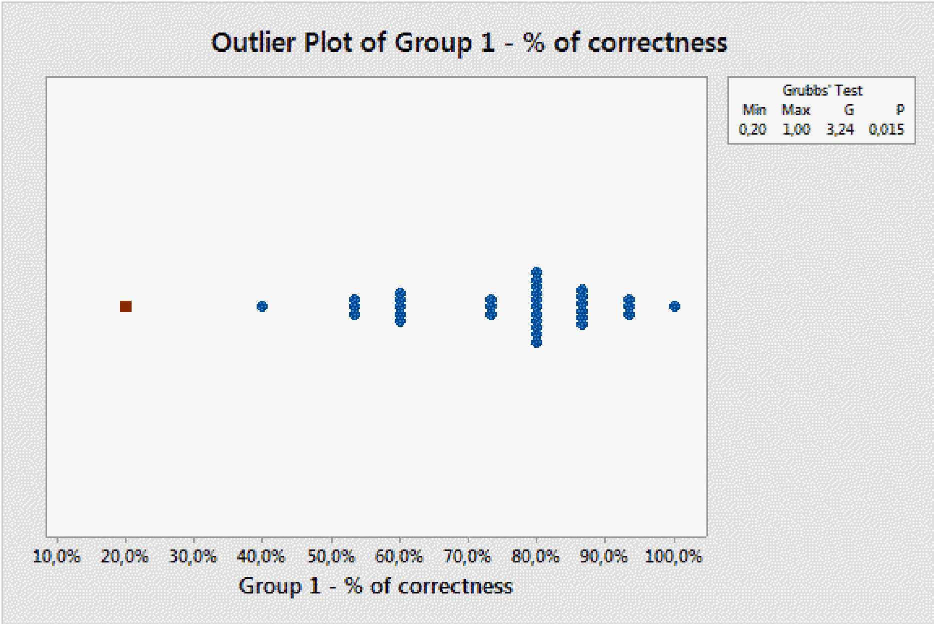

Figure 4.3: Outlier Identification - Percentage of Correctness - PRIME non-use

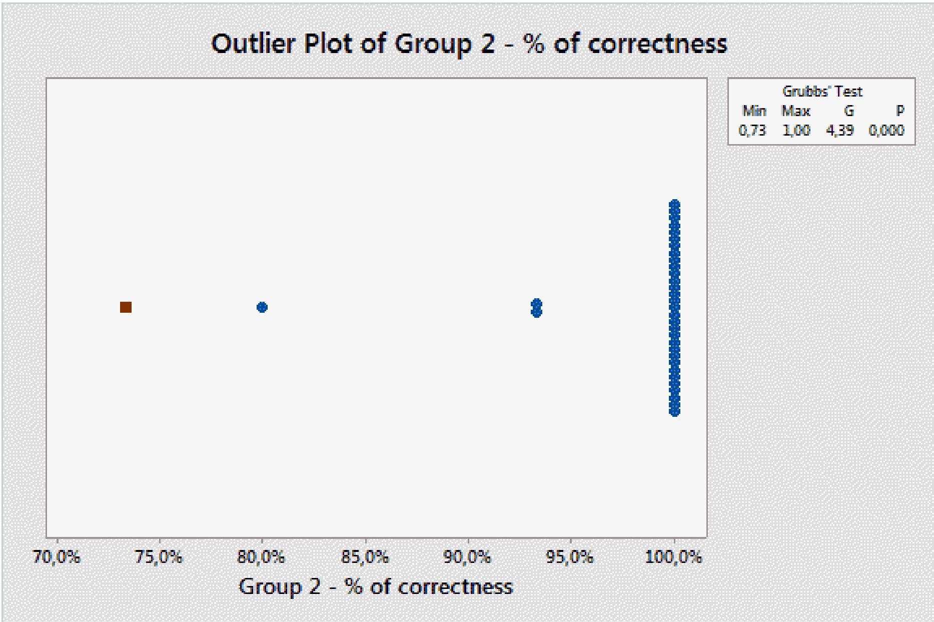

Figure 4.4: Outlier Identification - Percentage of Correctness - PRIME use

The outliers (painted in red) were analyzed, but no reasons were found to eliminate these subjects.

For the variable 'time spent', no outliers were found, for both group 1 and group 2 samples.

The next activity was to discover if the obtained samples presented a normal distribution. Once the samples were less than 50, the Shapiro-Wilk test was performed considering the following hypotheses:

- H0(Null hypothesis): The samples presented normal distribution;

- H1 (Alternative hypothesis): The samples did not presented normal distribution.

For the variable 'percentage of correctness' considering the sample to group 1, that is related to the first task where PRIME was not used, the Shapiro-Wilk test was presented in Figure 4.5:

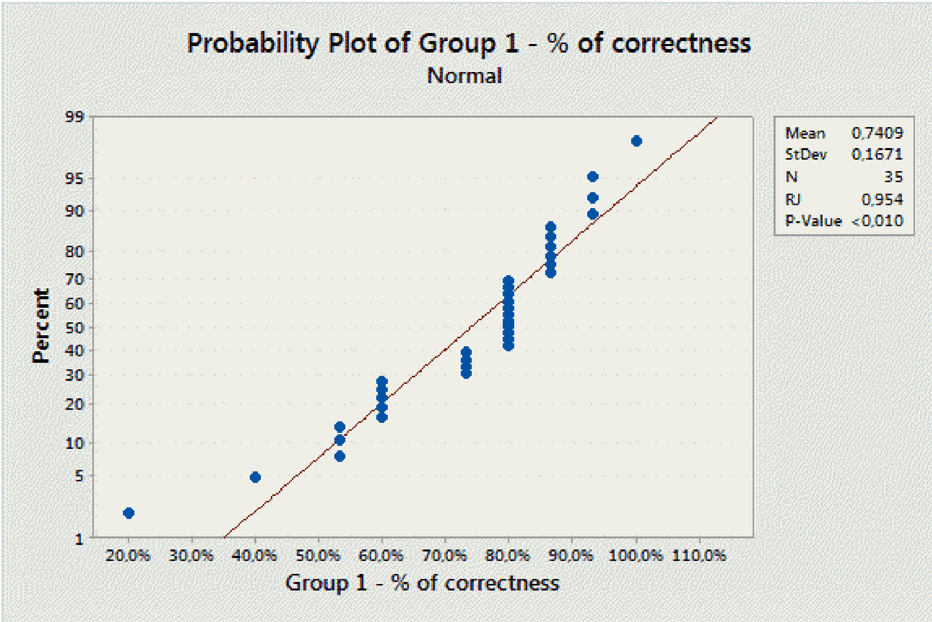

Figure 4.5: Normality test - Percentage of Correctness - PRIME non-use

For the variable 'percentage of correctness' considering the sample to group 2, that is related to the first task where PRIME was used, the Shapiro-Wilk test was presented in Figure 4.6

As can be seen in Figure 4.5 and Figure 4.6, considering a significance level of 5%, the samples were not normal because they presented p-value less than 0,05. In this way, the alternative hypothesis was accepted.

For the variable 'time spent' considering the sample to group 1, the Shapiro-Wilk test was presented in Figure 4.7.

For the variable 'time spent' considering the sample to group 2, the Shapiro-Wilk test was presented in Figure 4.8.

As can be seen in Figure 4.7 and Figure 4.8, considering a significance level of 5%, the samples were normal because they presented p-value superior than 0,05. In this way, the null hypothesis was accepted.

Once the samples related to "time spent" variable presented a normal distribution, the

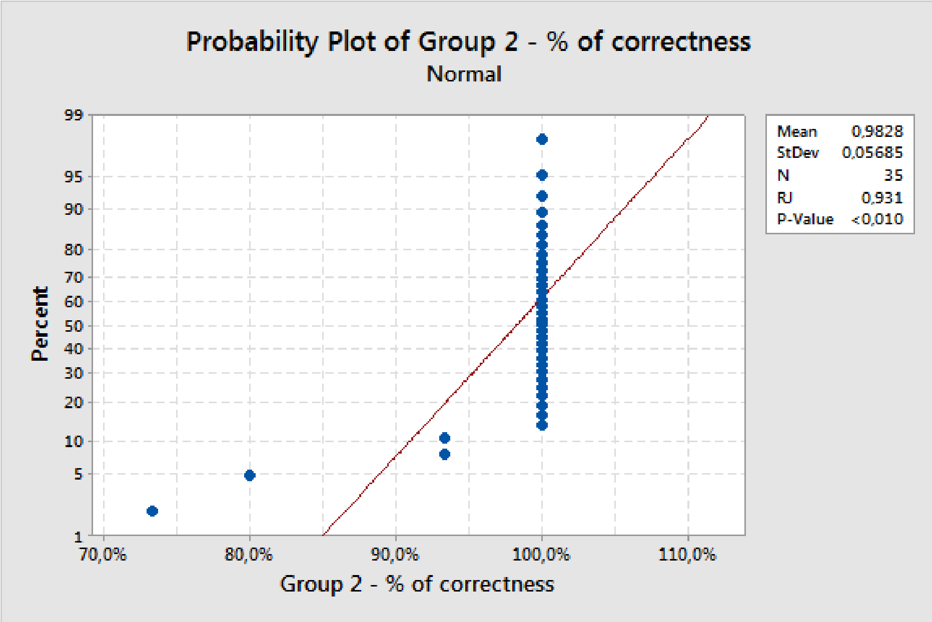

Figure 4.6: Normality - Percentage of Correctness - PRIME use

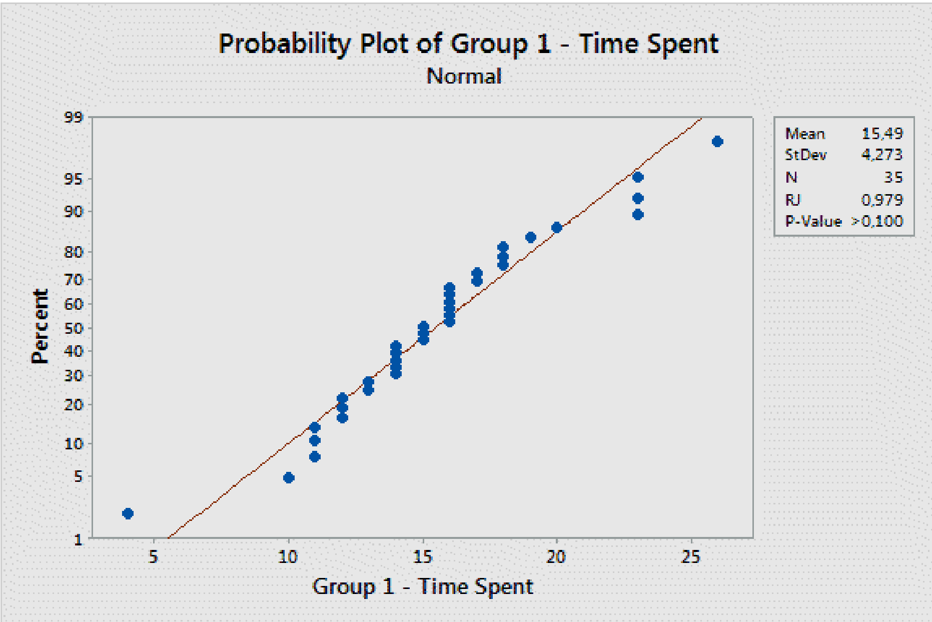

Figure 4.7: Normality test - Time Spent - PRIME non-use

homoscedasticity test was performed considering the following hyphoteses:

- H0(Null hypothesis): The data were homoscedastics;

- H1 (Alternative hypothesis): The data were not homoscedastics.

In order to perform this activity, it was used the Levene test with significance level of 5%. For the variable 'time spent' versus group (PRIME use and non-use) the Levene test was presented in Figure 4.9. In this case, the resulting p-value was superior to 0,05. In this way, the null hypothesis was accepted.

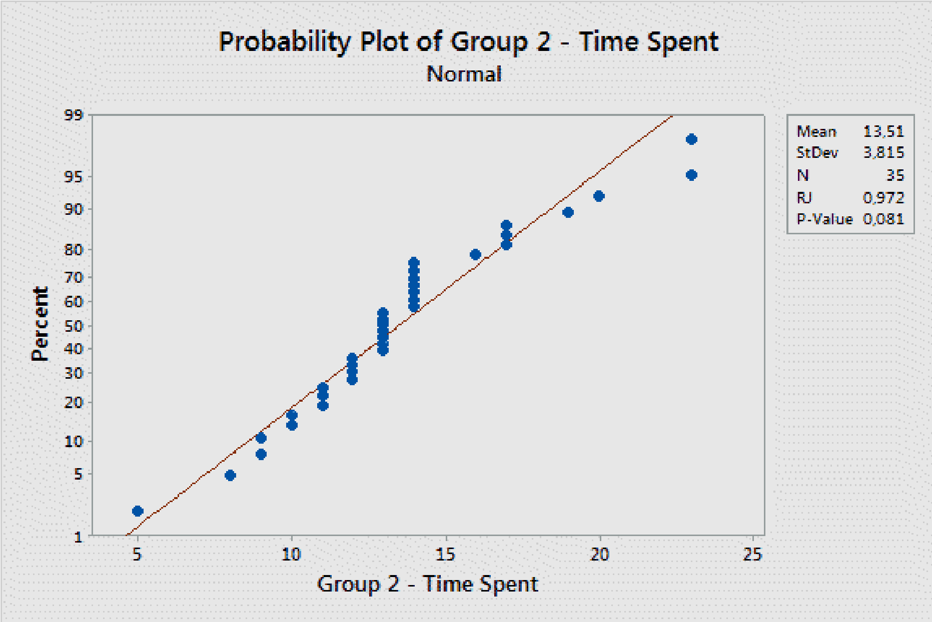

Figure 4.8: Normality test - Time Spent - PRIME use

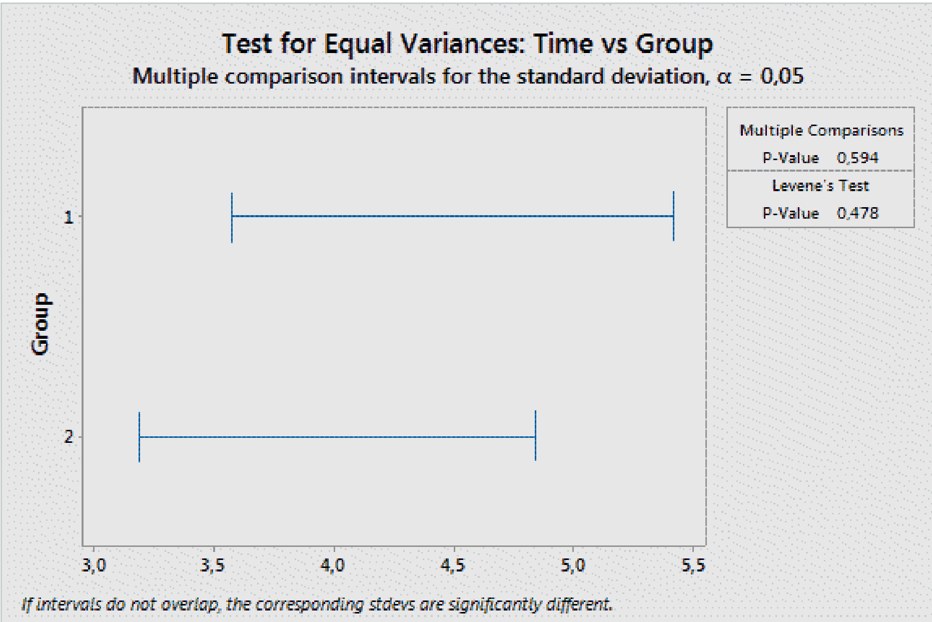

Figure 4.9: Homoscedasticity test - Percentage of Correctness

The samples related to the 'percentage of correctness' did not present a normal distribution, then a non parametric test was used. Considering that the same subjects participated in both groups and two treatments was used in the experiment,the Paired Wilcoxon test was used.

The Paired Wilcoxon test was performed considering the following hypoteses:

- H0(Null hypothesis): There is no difference between the means;

- H1 (Alternative hypothesis): There is difference between the means;

The resulting p-value was less than 0,05. Then, considering a significance level of 5%, the alternative hypothesis was accepted. The mean of the percentage of correctness without using PRIME solution (group 1) was 74% while using PRIME (group 2) was 98%. Then, group 1 mean is smaller than group 2 mean. Through this analysis, it was possible to conclude that the percentage of correctness to perform the experiment prototyping task not using PRIME (group 1) was statistically smaller than the percentage of correctness using PRIME (group 2). In this way, there are evidences that alternative hypothesis '(H1a) PRIME enhances pragmatic interoperability during scientific workflow prototyping in ECOS Collaborative PL-Science' investigated by this experimental study can be accepted.

Samples related to the 'spent time' were normal and homoscedastics, then a parametric test was used. Considering that the same subjects participated in both groups and two treatments was used in the experiment, the Paired T-Test was used. The Paired T-Test was performed considering the following hypoteses:

- H0(Null hypothesis): There is no difference between the means;

- H1 (Alternative hypothesis): There is difference between the means;

The resulting p-value was 0,027. Then, considering a significance level of 5%, the alternative hypothesis was accepted. The mean of the time spent without using PRIME was 15 minutes while using PRIME was 13 minutes. In this way, the use of PRIME made the workflow prototyping about 22% more efficient. We believe that this gain in time spent will increase according with the number of tasks in the workflow prototype. Moreover, it is possible to conclude that the time spent to perform the experiment prototyping task using PRIME (group 2) was statistically shorter than the time spent not using PRIME (group 1). This result generated evidence that the alternative hypothesis "(H1b) PRIME enhances efficiency during scientific workflow prototyping in ECOS Collaborative PL-Science" can be accepted.

4.4.2 QUALITATIVE ANALYSIS

This qualitative analysis aims to investigate the H1c hypothesis "PRIME enhances scientific workflow prototyping experienced in ECOS Collaborative PL-Science". In order to do this, the follow-up questionnaires contained questions related to the prototyping experience (Appendix F). In this experimental study, after accomplishing the first task

(PRIME non-use), the subjects answered questions related to H1c hypothesis. In the follow-up questionnaires the answers were provided in a five level Likert Scale, containing an intermediate level as suggested by Laitenberger and Dreyer (LAITENBERGER; DREYER, 1998).

Figures 4.10 and 4.11 illustrate the subjects answer to "Q1- In your opinion, the difficulty level of the proposed task performance was:" considering PRIME non-use and PRIME use respectively. In the first case (PRIME non-use) most frequent answers were 'low' and 'median' to describe the task difficulty level. With PRIME non-use, most frequent answers were 'low' and 'very low'. The obtained results evidenced that there is a tendency of PRIME to decrease the level of difficulty in workflow prototyping step.

Difficulties reported in group 1 was mainly related to the used search strategy. For example, one subject said to justify his/her answer *"The search for services is a little difficult, reading each description in order to find the desired service is tiring"*. Similar answers usually appeared. Also, subjects reported that the difficulty would increase according to the number of services registered in ECOS Collaborative PL-Science.

Low difficulty reported in group 2 was mainly related to the PRIME strategies. As stated by a subject, the second task provided *"A more automatic way to find what you are looking for."*

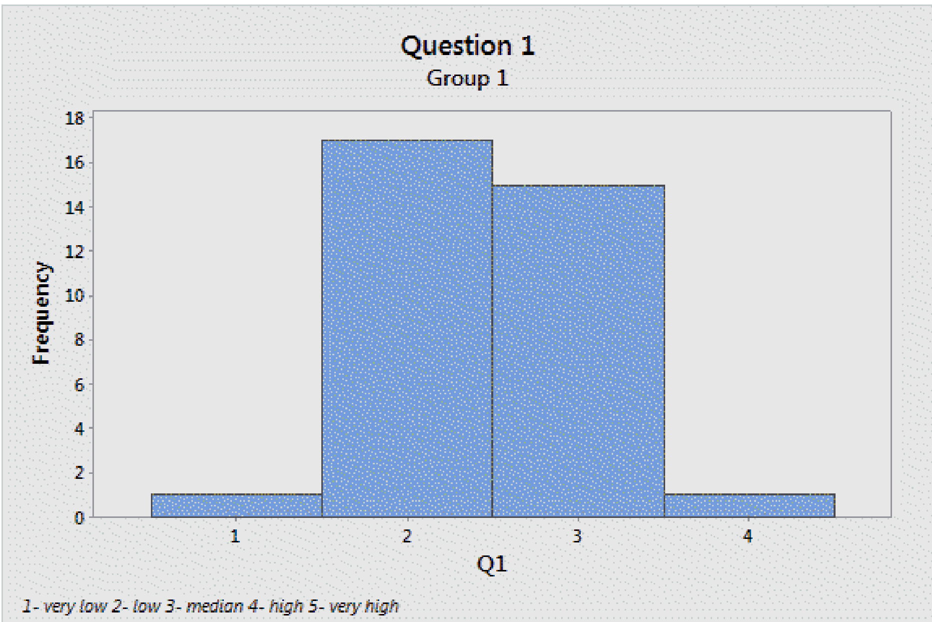

Figure 4.10: Histogram - Q1 - Group 1

The second question was about the ability of the subjects to create the scientific work-

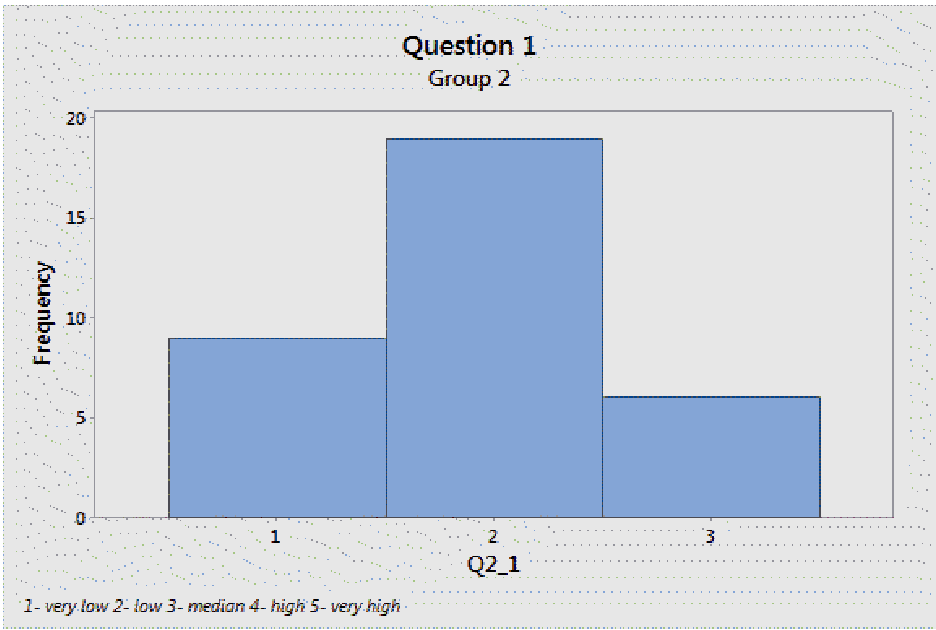

Figure 4.11: Histogram - Q1 - Group 2

flow prototype properly - "Q2- Were you able to create the scientific workflow prototype properly". As can be seen in figure 4.12, considering PRIME non-use, most frequent answers were 'agree' followed by 'neither agree nor disagree'. Considering PRIME use (Figure 4.13), most frequent answers were 'agree' and 'strongly agree'. According to this result, it is possible to realize a tendency of PRIME to increase the subjects confidence about their workflow prototype.

As reported by subjects *"...the tools gave me more confidence about my choices" and "With search tool the confidence that I created the correct workflow prototype is higher".* Similar answers were provided, which can support the understanding of Q2 results.

In figures 4.14 and 4.15 the answers were related to "Q3- You could easily discover which service to use at each workflow prototype stage." In the first case (PRIME non-use) the most frequent answer was 'disagree' to describe if they could easily discover which service to use. With PRIME, the more frequent answer was 'agree'. The results evidenced that there is a tendency of PRIME to decrease the level of difficulty to discover which service to use at each workflow prototype stage.

In taks 1, many subjects did not agree that they could easily discover which service to use because they considered that reading each service description was painful and tiring. With PRIME support and its strategies, the task was considered lighter. Some answers that illustrated theses arguments were: *"To search the descriptions of each service*

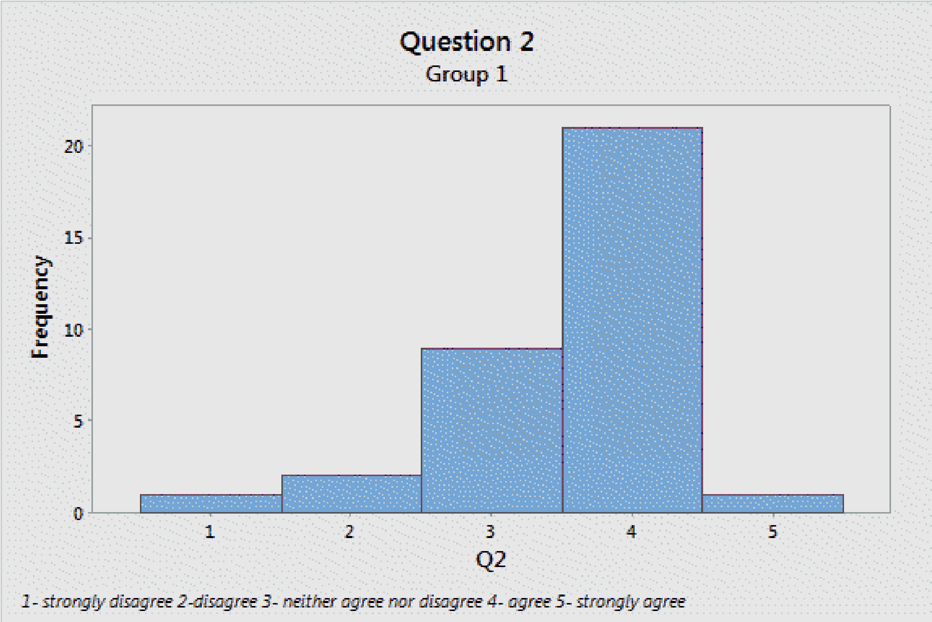

Figure 4.12: Histogram - Q2 - Group 1

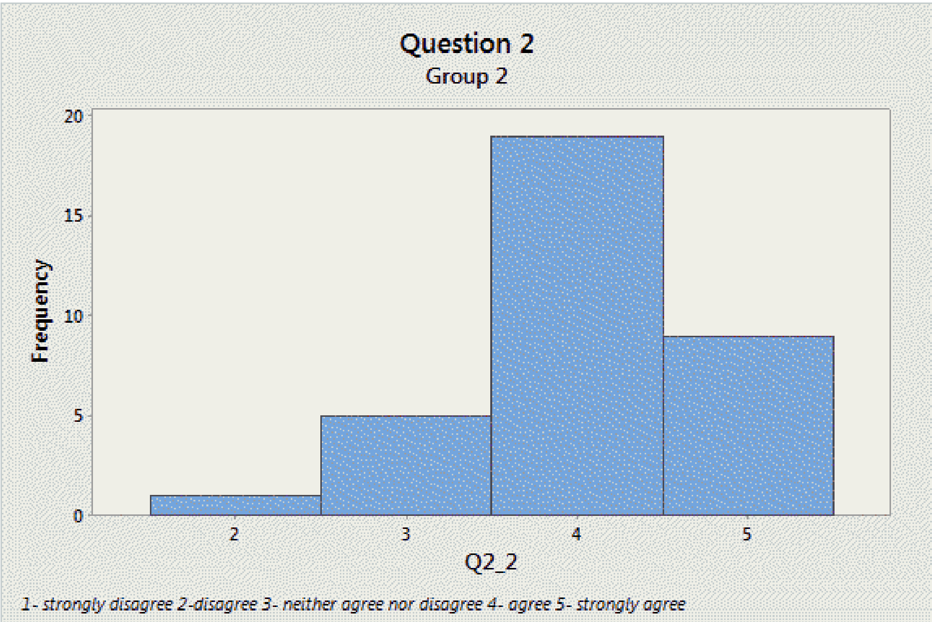

Figure 4.13: Histogram - Q2 - Group 2

is tiring" and *"With a more specific search, it is easier to choose the services because a considerable part can be eliminated."*

From direct observations, it was possible to notice that in the first task (group 1) when subjects had to read the services description, they could be divided into two categories. The first category was formed by those subjects that read very carefully each service description. The second category was formed by those subjects that were very influenced by the service name, only reading the description of the services with intuitive names in their opinions. However, this last strategy can hinder the discovery of the best services. Each research institution has its standard to define services name. This standard may not be intuitive for scientists from other institutions.

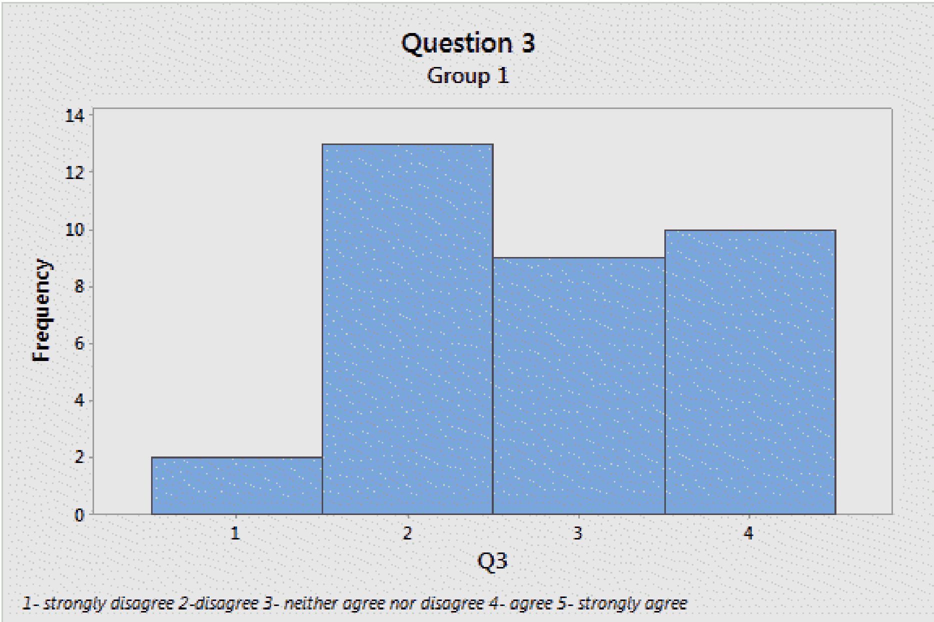

Figure 4.14: Histogram - Q3 - Group 1

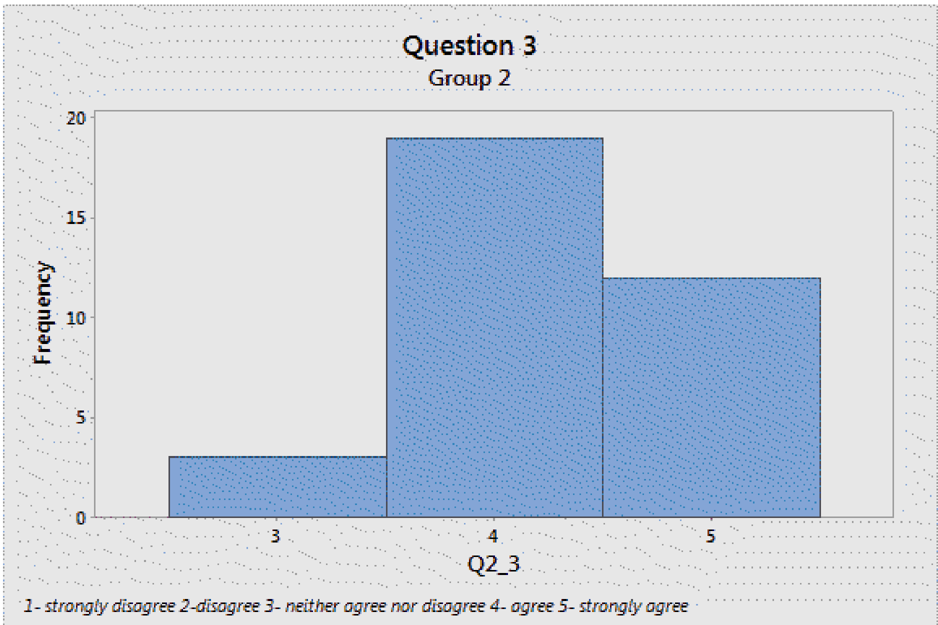

Figure 4.15: Histogram - Q3 - Group 2

The fourth question investigated was ”Q4- PRIME supports service discovery, selection and composition considering the experiment prototyping step”. This question was asked only in the second task (group 2). As can be seen in Figure 4.16, most frequent answer was 'agree'. The results evidenced that PRIME supported service discovery, selection and composition activities.

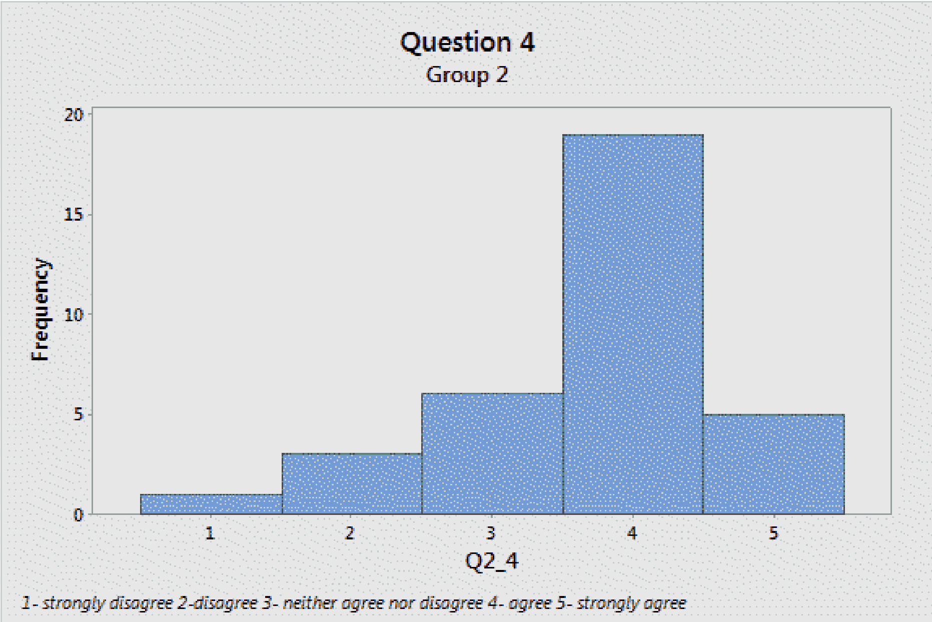

Figure 4.16: Histogram - Q4 - Group 2

The results obtained through these questions evidenced that PRIME made the experience of workflow prototyping easier and simpler. Also, the subjects revealed more confidence in their prototypes using PRIME. In this way, it was not found evidences to reject H1c hypothesis "PRIME enhances a better scientific workflow prototyping experience in ECOS Collaborative PL-Science".

In this qualitative analysis is also important to highlight the collected data only considering the bioinformatics experts (Table 4.1). Through the collected data analysis, it is possible to notice that in experts' opinion, PRIME use supported the workflow prototyping process. In their opinion, PRIME use decreased the difficult level and increased the confidence of the prototyping task. They also consider that PRIME use made the best services discovery easier.

According to these experts, the difficulty to find a web service that works fine is a big problem during workflows development. It is difficult to clearly know the purpose of the services found in the web and most times, basic information related to input parameters and services outputs are not available. This situation discourages the use of services developed by other scientists and the experts usually choose to develop their own services. In this vein, the experts stated that PRIME supports service search, discovery and selection and encourages the collaboration with other scientists and institutions. Also, they highlighted that in comparison with the systems that they normally use, PRIME made the development of scientific workflows easier.

	Questions	Ad-hoc Method	PRIME
Expert 1	Q1	Median	Very low
	Q2	Neither agree or disagree	Strongly agree
	Q3	Disagree	Agree
	Q4	-----	Strongly agree
Expert 2	Q1	Median	Low
	Q2	Agree	Strongly agree
	Q3	Agree	Strongly agree
	Q4	-----	Strongly agree
Expert 3	Q1	Low	Very low
	Q2	Agree	Agree
	Q3	Disagree	Agree
	Q4	-----	Agree

Table 4.1: Expert Analysis

Besides, considering the importance of experts evaluation, an additional evaluation with bioinformatics experts working in a real problem is planned. This evaluation is in planning with a partner institution of UFJF (Federal University of Juiz de Fora) that works with bioinformatics domain, specifically, the cattle tick problem.

4.5 THREATS TO VALIDITY

This experimental study presented some threats of validity. The threat to internal validity was related with the subject knowledge of the ECOS Collaborative PL-Science environment. This is due to the fact that an inadequate knowledge may influence the results. To reduce this threat, before the study, the subjects received a training to learn about ECOS Collaborative PL-Science purpose and functionalities. However, we understand that training is not the same of experience and then, in some way this lack of subjects experience can influence the results. The results obtained could not be generalized, in other words, the evidences are only relevant to ECOS Collaborative PL-Science context, threatening the external validity. Even in ECOS Collaborative PL-Science context, additional experiment must be done. It is necessary to consider experiments within its real-world context and professionals working in real problems (not toys).

4.6 FINAL CONSIDERATIONS

In this chapter PRIME was evaluated through experimental studies. PRIME evaluation evidenced that the non functional requirements were addressed. The evidences suggested that PRIME supported pragmatic interoperability and also increased efficiency and improve the user experience during scientific workflows prototyping in ECOS Collaborative PL-Science. Table 4.2 presents a summary of the experimental study results.

Task: Workflow Prototyping	**Ad-hoc Method**	**PRIME**
% of correctness	74%	98%
Time spent	13 minutes	15 minutes
Level of difficulty (Majority)	Low and median	Low and very low
Confidence in prototyping ability (Majority)	Agree and neither agree or disagree	Agree and strongly agree
Service discovery facility (Majority)	Disagree	Agree

Table 4.2: Experimental study summary results

Despite the evidences, additional experimental studies must be conducted, for example, experiments considering real-world context. This chapter attempted to answer the research question RQ4 " What are the benefits that this architecture provides considering the development of scientific workflows?" by accomplishing the goal G3 " To evaluate the benefits of the developed architecture in order to support the development of scientific workflows".

5 CONCLUSION

Pragmatic interoperability has been considered a key requirement to support collaboration. To reach this goal, systems that support scientific experiment development must be highly interoperable.

In this work, we proposed an architecture to support pragmatic interoperability in an open source web-based software ecosystem during the collaborative development of scientific workflows. Collaboration among scientists is crucial to support complex experiments, such as those that occur in Bioinformatics application domain during the development of scientific workflows.

To support collaborative development of scientific workflows and investigate pragmatic interoperability issues we introduced ECOS Collaborative PL-Science project. It is an environment that aims to support collaborative development of scientific workflows. Through this environment, pragmatic interoperability was investigated. As a result, PRIME service was proposed. It is based on pragmatic web service component in which interoperability solution relies to support service discovery, selection and composition. In addition, an ontology named Services Description was developed, in which the candidate services to compose an experiment are registered.

PRIME evaluation contributed to investigate pragmatic interoperability design decisions. The evaluation results generated evidences that PRIME is efficient and able to support pragmatic interoperability. Moreover, the adopted strategies in PRIME can be improved and reused in other domains. E-business domain applications, for example, generally adopt distributed solutions based on SOA concepts. As this work contributes to address pragmatic interoperability among services in a workflow, PRIME architecture has potential contributions to support business workflows and processes based on services compositions.

5.0.1 RESEARCH LIMITATIONS

The following limitations were identified:

- The experimental study performed was not enough to conclude that pragmatic interoperability was completely achieved and improved the collaboration process. Despite the positive evidences, it is necessary to perform additional and different experiments. For example, considering real problems to investigate the effect of pragmatic interoperability in the collaboration process when developing and executing scientific workflows in real-world contexts. In this vein, an evaluation in a bioinformatics research institution is planned.

- PRIME evaluation was performed considering 15 services registered in ECOS Collaborative PL-Science. In real scenarios this number can be potentially higher. In this way, this research cannot guarantee a good PRIME performance considering a high number of registered services. PRIME performance evaluation is necessary.

- The user interface of workflow prototyping does not allow the development of non-linear workflows. This is a limitation, once many experiments have non-linear designs. It is important to improve this interface, for example, by including drag and drop tools to support flexible and complex workflows designs.

5.0.2 FUTURE WORKS

The following future works were identified:

-It is important to evaluate the ranking equation used in PRIME architecture. This research did not focus in defining the best equation. In this vein, it is possible that other algorithms present a better performance or accuracy than the Levenshtein distance algorithm used in this work, for example.

- PRIME usability issues need to be evaluated in ECOS Collaborative PL-Science. The experimental study conducted to evaluate PRIME in ECOS Collaborative PL-Science context revealed that some buttons positions and tools can be improved to provide a better user experience. However, these suggestions collected in the experimental study must be discussed with the members of the ECOS Collaborative PL-Science project.

- Pragmatic interoperability is an important requirement to support collaboration. However, only pragmatic interoperability does not guarantee that collaboration will be effective in a system. There is a need of other tools to support collaboration process, for example, forums, among others. In this vein, these tools must be develop in ECOS Collaborative PL-Science.

- The intermediate language used to produce executable files of the workflow prototypes was not fully integrated with ECOS Collaborative PL-Science. The integration addressed PRIME needs, however a complete integration must be performed.

- Information about the reputation of a service or a scientist needs to be investigated in depth. This data was used as manual input during service registration. However, how to derive and calculate this variable is an interesting question to be explored.

- The information about the domain, in Service Description ontology, can be enriched using, for example, linked data. In PRIME, the domain of a service was linked to a complex network and other ontologies. It would be important to investigate if the linked data can be used to support the discovery of the best services to interoperate. Linked data can be used as an additional way to capture and infer user context and expectations and then, to improve pragmatic interoperability process.

REFERENCES

ASUNCION, C. H.; SINDEREN, M. J. V. Pragmatic interoperability: A systematic review of published definitions. In: **Enterprise Architecture, Integration and Interoperability**, 2010. p. 164–175.

ASUNCION, C. H.; SINDEREN, M. van. Towards pragmatic interoperability in the new enterprisea survey of approaches. In: **Enterprise Interoperability**, 2011. p. 132–145.

BARABÁSI, A.-L.; GULBAHCE, N.; LOSCALZO, J. Network medicine: a network-based approach to human disease. **Nature Reviews Genetics**, Nature Publishing Group, v. 12, n. 1, p. 56–68, 2011.

BASTOS, B.; BRAGA, R.; GOMES, A. Uma abordagem baseada em padrões para interoperação de especificações de workflows científicos. In: **CBSOFT - WTDSOFT**, 2014.

BELLOUM, A.; INDA, M. A.; VASUNIN, D.; KORKHOV, V.; ZHAO, Z.; RAUWERDA, H.; BREIT, T. M.; BUBAK, M.; HERTZBERGER, L. O. Collaborative e-science experiments and scientific workflows. **Internet Computing, IEEE**, IEEE, v. 15, n. 4, p. 39–47, 2011.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. et al. The semantic web. **Scientific american**, New York, NY, USA:, v. 284, n. 5, p. 28–37, 2001.

BRAVO, M.; ALVARADO, M. On the pragmatic similarity between agent communication protocols: Modeling and measuring. **On the Move to Meaningful Internet Systems: OTM**, p. 128–137, 2008.

BURROWS, S.; POTTHAST, M.; STEIN, B. Paraphrase acquisition via crowdsourcing and machine learning. **ACM Transactions on Intelligent Systems and Technology (TIST)**, ACM, v. 4, n. 3, p. 43, 2013.

CHEN, D.; DOUMEINGTS, G.; VERNADAT, F. Architectures for enterprise integration and interoperability: Past, present and future. **Computers in industry**, Elsevier, v. 59, n. 7, p. 647–659, 2008.

CLASSE, T.; BRAGA, R.; CAMPOS, F.; DAVID, J. M. N. A semantic peer-to-peer network for service composition in scientific domains. **Procedia Technology**, Elsevier, v. 9, p. 215–225, 2013.

CORDEIRO, J.; DIAS, G.; CLEUZIOU, G. Biology based alignments of paraphrases for sentence compression. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing**, 2007. p. 177–184.

COSTA, C.; MURTA, L. Version control in distributed software development: A systematic mapping study. In: IEEE. **Global Software Engineering (ICGSE), 2013 IEEE 8th International Conference on**, 2013. p. 90–99.

COSTA, G. C. B.; BRAGA, R.; DAVID, J. M. N.; CAMPOS, F.; ARBEX, W. Pl-science: A scientific software product line. **Procedia Computer Science**, Elsevier, v. 18, p. 759–768, 2013.

DALPRA, H.; COSTA, G.; TASSIO, S.; ; BRAGA, R.; WERNER, C.; CAMPOS, F.; DAVID, J. M. N. Using ontology and data provenance to improve software processes. 2015.

DAVID, J. M. N.; MACIEL, R. S. P. WGWSOA - Implementing collaboration services in a middleware infrastructure. In: **Proceedings of the 2009 Simpósio Brasileiro de Sistemas Colaborativos**, 2009. (SBSC '09), p. 193–201. ISBN 978-0-7695-3918-8.

ELLIS, C. A.; GIBBS, S. J.; REIN, G. Groupware: some issues and experiences. **Communications of the ACM**, ACM, v. 34, n. 1, p. 39–58, 1991.

FERREIRA, L.; PUTNIK, G. D.; CRUZ-CUNHA, M. M.; PUTNIK, Z. Towards effective tourism dynamic packages. **Information Resources Management Journal (IRMJ)**, IGI Global, v. 25, n. 2, p. 1–21, 2012.

FREITAS, V.; DAVID, J. M. N.; BRAGA, R.; CAMPOS, F. Uma arquitetura para ecossistema de software científico. In: CBSOFT, 2015. p. 41–48.

GANGULY, N.; DEUTSCH, A.; MUKHERJEE, A. **Dynamics on and of complex networks: applications to biology, computer science, and the social sciences**, 2009.

GAO, M.; WU, Z.; LIU, K. Pragmatic grid for personalized resource provision. In: IEEE. **Service Operations and Logistics, and Informatics, 2008. IEEE/SOLI 2008. IEEE International Conference on**, 2008. v. 1, p. 1023–1028.

GASPAR, W.; BRAGA, R.; CAMPOS, F.; DAVID, J. M. N.; ORNELAS, T. Scientific provenance metadata capture and management using semantic web. **International Journal of Metadata, Semantics and Ontologies**, Inderscience Publishers (IEL), v. 10, n. 2, p. 123–138, 2015.

GERACI, A.; KATKI, F.; MCMONEGAL, L.; MEYER, B.; LANE, J.; WILSON, P.; RADATZ, J.; YEE, M.; PORTEOUS, H.; SPRINGSTEEL, F. **IEEE standard computer dictionary: Compilation of IEEE standard computer glossaries**, 1991.

GOH, K.-I.; CUSICK, M. E.; VALLE, D.; CHILDS, B.; VIDAL, M.; BARABÁSI, A.-L. The human disease network. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 104, n. 21, p. 8685–8690, 2007.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge acquisition**, Elsevier, v. 5, n. 2, p. 199–220, 1993.

ISO. {ISO/IEC} 2382-1: 1993 information technology- vocabulary-part 1:fundamental terms. 1993.

JOSEY, A. **TOGAF® Version 9.1 A Pocket Guide**, 2011.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK, Keele University**, v. 33, n. 2004, p. 1–26, 2004.

KITCHENHAM, B. **Guidelines for performing systematic literature reviews in software engineering**, 2007.

KITCHENHAM, B. Whats up with software metrics?–a preliminary mapping study. **Journal of systems and software**, Elsevier, v. 83, n. 1, p. 37–51, 2010.

KUTVONEN, L. Tools and infrastructure facilities for controlling non-functional properties in inter-enterprise in collaborations. In: IEEE. **Enterprise Distributed Object Computing Conference Workshops, 2008 12th**, 2008. p. 423–432.

KUTVONEN, L.; RUOKOLAINEN, T.; METSO, J. Interoperability middleware for federated business services in web-pilarcos. **Agent and Web Service Technologies in Virtual Enterprises**, IGI Global, p. 288, 2007.

LAITENBERGER, O.; DREYER, H. M. Evaluating the usefulness and the ease of use of a web-based inspection data collection tool. In: IEEE. **Software Metrics Symposium, 1998. Metrics 1998. Proceedings. Fifth International**, 1998. p. 122–132.

LEE, J.; LEE, Y.; SHAH, S.; GELLER, J. His-kcwater: context-aware geospatial data and service integration. In: ACM. **Proceedings of the 2007 ACM symposium on Applied computing**, 2007. p. 24–29.

LEE, Y.; PATEL, C.; CHUN, S. A.; GELLER, J. Towards intelligent web services for automating medical service composition. In: IEEE. **Web Services, 2004. Proceedings. IEEE International Conference on**, 2004. p. 384–391.

LEENHEER, P. D.; CHRISTIAENS, S. Mind the gap!: Transcending the tunnel view on ontology engineering. In: ACM. **Proceedings of the 2nd international conference on Pragmatic web**, 2007. p. 75–82.

LI, W.; LI, S. Improve the semantic interoperability of information. In: **Intelligent Systems, 2004. Proceedings. 2004 2nd International IEEE Conference**, 2004. v. 2, p. 591–59.

LIANG, L.; RONG, W.; LIU, K. Intelligent agents for pragmatic web services. In: IEEE. **Advanced Language Processing and Web Information Technology, 2007. ALPIT 2007. Sixth International Conference on**, 2007. p. 530–536.

LIU, S.; LI, W.; LIU, K. Pragmatic oriented data interoperability for smart healthcare information systems. In: IEEE. **Cluster, Cloud and Grid Computing (CCGrid), 2014 14th IEEE/ACM International Symposium on**, 2014. p. 811–818.

LUDÄSCHER, B.; ALTINTAS, I.; BERKLEY, C.; HIGGINS, D.; JAEGER, E.; JONES, M. B.; LEE, E. A.; TAO, J.; ZHAO, Y. Scientific workflow management and the kepler system. **Concurrency and Computation: Practice and Experience**, v. 18, n. 10, p. 1039–1065, 2006.

MAFRA, S. N.; BARCELOS, R. F.; TRAVASSOS, G. H. Aplicando uma metodologia baseada em evidência na definiçao de novas tecnologias de software. In: **Proceedings of the 20th Brazilian Symposium on Software Engineering (SBES 2006)**, 2006. v. 1, p. 239–254.

MILLER, G. A. Wordnet: a lexical database for english. **Communications of the ACM**, ACM, v. 38, n. 11, p. 39–41, 1995.

MINGXIN, G. Enterprise isomorphic mapping mechanism: Towards ontology interoperability in eis development. In: IEEE. **e-Business Engineering, 2009. ICEBE'09. IEEE International Conference on**, 2009. p. 340–345.

MONTERO, I.; PENA, J.; RUIZ-CORTES, A. From feature models to business processes. In: IEEE. **Services Computing, 2008. SCC'08. IEEE International Conference on**, 2008. v. 2, p. 605–608.

MORRIS, C. W. **Foundations of the Theory of Signs**, 1938.

NEIVA, F.; DAVID, J. M. N.; BRAGA, R.; CAMPOS, F. Pragmatic interoperability: A systematic review and mapping of the literature. **submitted (under review)**, 2015.

NEIVA, F.; DAVID, J. M. N.; BRAGA, R.; CAMPOS, F.; FREITAS, V. Prime: Pragmatic interoperability architecture to support collaborative development of scientific workflows. In: CBSOFT. **IX Brazilian Symposium on Components, Architectures and Reuse Software (SBCARS 2015)**, 2015. p. –.

NEIVA, F. W.; DAVID, J. M. N.; MACIEL, R. S. P.; BRAGA, R.; CAMPOS, F. Siss: Extending semantic interoperability to support collaborative system development and execution. In: IEEE. **Computer Supported Cooperative Work in Design (CSCWD), Proceedings of the 2014 IEEE 18th International Conference on**, 2014. p. 53–58.

ORNELAS, T.; BRAGA, R.; CAMPOS, F.; DAVID, J. M. N. Improvement of provenance capture and query using semantic web resoureces. In: **13th International Conference WWW/INTERNET**, 2014.

PAPAZOGLOU, M. P.; TRAVERSO, P.; DUSTDAR, S.; LEYMANN, F. Service-oriented computing: State of the art and research challenges. **Computer**, IEEE, n. 11, p. 38–45, 2007.

PEREIRA, A. F.; DAVID, J. M. N.; BRAGA, R.; CAMPOS, F. Collabpl-science: Using collaborative elements in a scientific software product line. In: IEEE. **e-Science (e-Science), 2014 IEEE 10th International Conference on**, 2014. v. 2, p. 49–54.

PETTICREW, M.; ROBERTS, H. **Systematic reviews in the social sciences: A practical guide**, 2008.

POKRAEV, S.; REICHERT, M.; STEEN, M.; WIERINGA, R. Semantic and pragmatic interoperability: A model for understanding. In: **Proceedings of the Open Interop Workshop on Enterprise Modelling and Ontologies for Interoperability (EMOI - INTEROP'05)**, 2005. (CEUR Workshop Proceedings, v. 160), p. 1–5.

POKRAEV, S. V. Model-driven semantic integration of service-oriented applications. Novay, 2009.

POLLOCK, J.; HODGSON, R. **Adaptive Information: Improving Business Through Semantic Interoperability, Grid Computing, and Enterprise Integration**, 2004. (Wiley Series in Systems Engineering and Management). ISBN 9780471714200.

ROURE, D. D.; GOBLE, C.; STEVENS, R. The design and realisation of the virtual research environment for social sharing of workflows. **Future Generation Computer Systems**, Elsevier, v. 25, n. 5, p. 561–567, 2009.

RUKANOVA, B.; SLOOTEN, K. van; STEGWEE, R. A. Towards a meta model for describing communication. In: **Enterprise Information Systems VI**, 2006. p. 203–210.

RUNESON, P.; HOST, M.; RAINER, A.; REGNELL, B. **Case study research in software engineering: Guidelines and examples**, 2012.

RUOKOLAINEN, T. Modelling framework for interoperability management in collaborative computing environments. **Licentiate Thesis, University of Helsinki**, 2009.

SCHEPENS, J.; DIJKSTRA, T.; GROOTJEN, F. Distributions of cognates in europe as based on levenshtein distance. **Bilingualism: Language and Cognition**, Cambridge Univ Press, v. 15, n. 01, p. 157–166, 2012.

SCHOOP, M.; MOOR, A. d.; DIETZ, J. L. The pragmatic web: a manifesto. **Communications of the ACM**, ACM, v. 49, n. 5, p. 75–76, 2006.

SOLINGEN, R. V.; BASILI, V.; CALDIERA, G.; ROMBACH, H. D. Goal question metric (gqm) approach. **Encyclopedia of Software Engineering**, Wiley Online Library, 2002.

SOMMERVILLE, I. **Engenharia de software**, 2010.

STEINMACHER, I.; CHAVES, A. P.; GEROSA, M. A. Awareness support in distributed software development: A systematic review and mapping of the literature. **Computer Supported Cooperative Work (CSCW)**, Springer, v. 22, n. 2-3, p. 113–158, 2013.

TAMANI, E.; EVRIPIDOU, P. A pragmatic and pervasive methodology to web service discovery. In: SPRINGER. **On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops**, 2006. p. 1285–1294.

TOLK, A.; MUGUIRA, J. A. The levels of conceptual interoperability model. In: CITESEER. **Proceedings of the 2003 Fall Simulation Interoperability Workshop**, 2003. v. 7.

TOLK, A.; TURNITSA, C. D.; DIALLO, S. Y. Ontological implications of the levels of conceptual interoperability model. In: **WMSCI2006: Proceedings of the 10th World Multi-Conference on Systemics, Cybernetics and Informatics**, 2006.

WANG, W.; TOLK, A.; WEIPING, W. The levels of conceptual interoperability model: applying systems engineering principles to mes. In: **SpringSim**, 2009.

WILCOX, R. R. **Applying contemporary statistical techniques**, 2003.

WOHLIN, C.; RUNESON, P.; HÖST, M.; OHLSSON, M.; REGNELL, B.; WESSLÉN, A. **Experimentation in Software Engineering**, 2012. (Computer Science). ISBN 9783642290435.

WOLSTENCROFT, K.; HAINES, R.; FELLOWS, D.; WILLIAMS, A.; WITHERS, D.; OWEN, S.; SOILAND-REYES, S.; DUNLOP, I.; NENADIC, A.; FISHER, P. et al. The taverna workflow suite: designing and executing workflows of web services on the desktop, web or in the cloud. **Nucleic acids research**, Oxford Univ Press, p. gkt328, 2013.

YANG, M. C. A study of prototypes, design activity, and design outcome. **Design Studies**, Elsevier, v. 26, n. 6, p. 649–669, 2005.

ZHANG, H.; BABAR, M. A. On searching relevant studies in software engineering. British Informatics Society Ltd., 2010.

Appendix A - PRAGMATIC INTEROPERABILITY: A SYSTEMATIC REVIEW AND MAPPING OF THE LITERATURE

This appendix presents a systematic review and mapping of the literature that aims to help the comprehension of the state of art about pragmatic interoperability (PI) computational solutions. This review was organized based on the main activities proposed by Kitchenham (2007): planning, conducting and reporting the study.

A.1 REVIEW PLANNING

The planning activity identifies the objectives and defines a protocol. The protocol specifies the method to be used in the systematic review and mapping in order to reduce researcher bias (STEINMACHER et al., 2013). Moreover, the systematic review and mapping must be reproducible and the protocol has an important role to fulfill in this requirement. This section summarizes the protocol.

Research Questions. The systematic mapping aims to answer the questions below:

MQ1: How many studies were published over the years?

MQ2: Who are the most active authors in the area?

MQ3: Which publication vehicles are the main targets for research production in the area?

MQ4: In which domains has pragmatic interoperability been applied? (e.g. Bioinformatics, Telemedicine, Business)

MQ5: Which type of computational support has pragmatic interoperability technique provided (e.g. framework, software architecture, etc.)?

MQ6: Which definitions of pragmatic interoperability have been used?

MQ7: Which methods were used to validate the proposed pragmatic interoperability solutions?

The systematic review aims to answer the questions below:

RQ1: Which solutions have been used to enhance pragmatic interoperability?

RQ2: How did the proposed solutions address pragmatic interoperability?

Based on the research questions, the aspects proposed by Petticrew and Roberts (2008) were taken into account to define our scope:

Population (P): Solutions that implement interoperability.

Intervention (I): Pragmatic interoperability solution.

Comparison (C): No comparison intervention.

Outcomes (O): Solution.

Context(C): Computational solutions.

To select the papers to be analyzed in order to answer the research questions some criteria were defined.

Inclusion and Exclusion Criteria. The process used to include and exclude a paper was organized in four inclusion criteria (IC) and four exclusion criteria (EC), presented below:

IC1: The papers reports the pragmatic interoperability solution (method, technique, model, tool, framework) AND

IC2: The proposed solution are applied on software OR system OR application OR service OR infrastructure AND

IC3: The papers are described in English language AND

IC4: The papers are reported in peer reviewed Workshop or Conference or Journal or Technical Reports.

The following exclusion criteria were established:

EC1: The papers do not propose a pragmatic interoperability solution AND

EC2: The proposed solution are not applied on software OR system OR application OR service OR infrastructure AND

EC3: The papers are not described in English AND

EC4: The papers are not published in a peer reviewed conference or journal.

After the definition of the research questions and the paper inclusion and exclusion criteria, the following steps were (i) to define the sources of the papers and (ii) the search string that was performed in these sources. We chose the databases according to the following requirements, based on Costa and Murta (2013):

- They are capable of using logical expressions or a similar mechanism.
- They allow full-length searches or searches only in specific fields of the works.

- They are available in the researcher's institution.
- They cover the research area of interest in this mapping: computer science.

Sources: The search was done in six electronic databases namely:

- Scopus (www.scopus.com).
- IEEExplore (ieeexplore.ieee.org).
- ScienceDirect (www.sciencedirect.com).
- Compendex (www.engineeringvillage.com).
- Web of Science (apps.webofknowledge.com).

Query String. Initially, to create the search string, the major terms from the defined research questions and PICOC and their alternate spelling and synonyms were identified.

Based on the above terms, we formed an initial string using Boolean OR/AND operators. Synonyms and alternate spellings were concatenated using Boolean OR and then these terms were concatenated using Boolean AND to form one string. We also defined a set of potential primary studies, as suggested in Zhang and Babar (2010), to verify the search string accuracy in the selected databases and if the search retrieved relevant studies. The keywords from the potential interest studies and from newly fetched ones were analyzed in order to find new relevant terms to be included as part of the search string.

The final search string was described as follows:

(pragmatic OR pragmatics OR pragmatism) AND (interoperability OR interoperate OR interoperable OR interoperation OR similarity OR integrate OR integration) AND (solution OR method OR technique OR model OR tool OR framework OR architecture OR infrastructure OR approach) AND (computational OR system OR application OR software)

Tolk et al. (2006) and Kutvonen (2008) were known potential interest studies that were used to control if the search string was finding relevant studies.

The mapping was carried out and the details for this step are presented in the following section.

A.2 REVIEW AND MAPPING CONDUCTION

The first step in this phase was to execute the search string considering the selected sources. Afterwards, the study selection process was performed. The process comprised the following steps:

1. The results from all databases were merged in JabRef (http://jabref.sourceforge.net/) and duplicates were removed.

2. The papers were analyzed based on their titles and abstracts, considering the inclusion/exclusion criteria. Those papers considered clearly irrelevant were excluded. The introduction, background and conclusion sections of the papers considered doubtful were read. Finally, papers considered included were analyzed in the next step.

3. Included papers were fully read and analyzed considering the research questions. The selected papers were submitted to a quality assessment checklist and information was extracted and put on a form. A snowballing process was also performed in the set of papers, which was approved in the quality assessment, by reviewing their references in order to find other potential primary studies. The papers not obtained initially were requested to the authors by e-mail.

This process was performed in two rounds. The first round was performed in December, 2013 and this process is illustrated in Figure A.1.

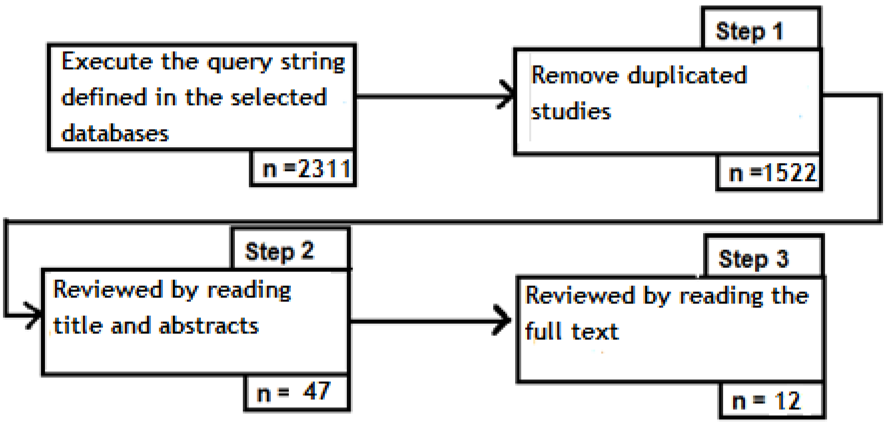

Figure A.1: Paper Selection Process - Round 1

The second round was performed in September, 2014 and this process is illustrated in Figure A.2

For analysis purposes, both rounds were integrated in one single process (Figure A.3).

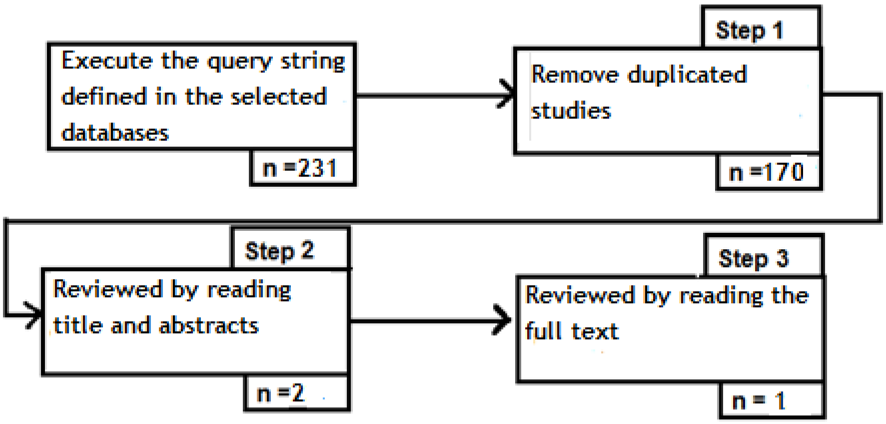

Figure A.2: Paper Selection Process - Round 2

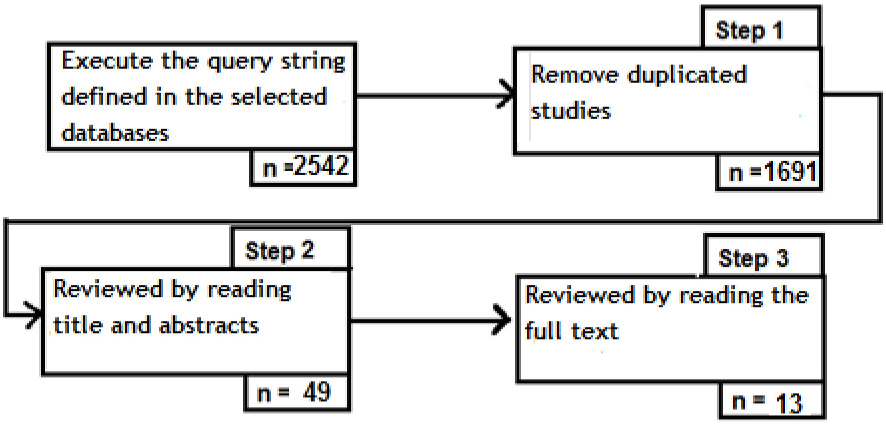

Figure A.3: Paper Selection Process - Round Both Rounds

Based on both rounds, the following results were obtained. In Step1, from 2542 papers returned by the query, 1691 (66.52%) remained after removing duplicates. Step 2 involved title and abstract analysis. From 1691 papers, 1542 (91.2%) were excluded and 149 (8.8%) were included or considered doubtful. The doubtful papers were analyzed based on the introduction, background and conclusion sections. At the end of this step 49 papers remained, which represent a reduction (97.1%) compared with the initial set of 1691 papers. Throughout this process, it was possible to notice that a huge number of false positive papers were captured. This occurred because of the word "pragmatic", and its variations, in our search string. The word and its variations were frequently used as a synonym of practical, operative or realistic in abstracts and titles. However, it was not possible to remove these words from the string because they are essential and therefore it was necessary to deal with a huge number of false positives. Finally, in Step 3, the papers were read and their quality was evaluated using a quality assessment checklist (Table A.1). Each one of the 13 questions could score 1 point if the answer was ”Yes”, 0.5 point if the answer was ”Partial” or 0 point if the answer was ”No”. Using this system, each paper could obtain a score from 0 to 13 points. The first quartile (13/4= 3.25) was used as the cutoff point for a paper to be included. If a paper scored less than 3.25, it was excluded it from our final list to avoid low quality works. In this step we also performed data extraction.

Question	Score
1. Are the research aims clearly specified?	Y/N/Partial
2. Was the study designed to achieve these aims?	Y/N/Partial
3. Are the used techniques clearly described and their selection justified?	Y/N/Partial
4. Are the variables considered by the study suitably measured?	Y/N/Partial
5. Are the data collection methods adequately described?	Y/N/Partial
6. Is the data collected adequately described?	Y/N/Partial
7. Is the purpose of the data analysis clear?	Y/N/Partial
8. Are statistical techniques used to analyze data adequately described and their use justified?	Y/N/Partial
9. Are negative results (if any) presented?	Y/N/Partial
10. Do the researchers discuss any problems with the validity/reliability of their results?	Y/N/Partial
11. Are all research questions answered adequately?	Y/N/Partial
12. How clear are the links between data, interpretation and conclusions?	Y/N/Partial
13. Are the findings based on multiple projects?	Y/N/Partial

Table A.1: Quality Assessment Checklist Adapted From Muhammad et al. (2014) and Kitchenham (2007)

A database schema (Figure A.4) supported the data extraction task. In order to build the schema, PgAdminIII tool for PostgreSQL was used. As illustrated in Figure 2, the data were extracted from papers in order to answer the defined research questions. General information about the papers' authors (name, email and affiliation) was extracted and linked to the published papers (publication). Thus, it was avoided inserting repeated information about authors and it was possible to easily identify the authors that published more than one paper in the set of papers. Information about published papers was extracted such as title, publication year, publication type (conference, journal, workshop, etc.), publication channel (e.g. conference or journal name), status (excluded or included), status report (e.g. A justification of a paper exclusion), etc. Specific information about pragmatic interoperability content was also extracted, such as the type of pragmatic interoperability technique used (e.g. ontology, thesaurus, pragmatics web service, intelligent agents, etc.), the name of pragmatic interoperability technique investigated (e.g. FIPA ACL), the domain where pragmatic interoperability technique was applied (e.g bioinformatics, telemedicine, business, etc.), the type of computational support the pragmatic interoperability technique had provided (e.g. framework, software architecture, etc.) and the main findings reported by the papers. When the published paper presented some experimental study, the data was also extracted, such as type of study (e.g. case study, experiment, survey etc.), study participants (e.g. students, practitioners, both), data collection method (e.g. interview, questionnaire) and others. Information about quality assessment process was also stored in the database schema.

At the end of Step 3, from 49 papers, 13 (26.5%) were included. Briefly, from 1691 papers at the end of Step 1, only 0.8% remained at the end of Step 3. This dramatic reduction in studies can be justified by: (i) The huge number of false positives studies captured in the electronic bases caused by the use of the word ”pragmatic” and its variations in our search string. (ii) Absence of solution proposals to support pragmatic interoperability. Many studies discussed the relevance of pragmatic interoperability without introducing a solution to its achievement. (iii) Finally, the low score in the quality assessment checklist.

A.3 SYSTEMATIC MAPPING REPORT

The 13 selected papers were analyzed and the information to answer the mapping questions (MQ) was extracted.

Figure A.5 represents the answer to MQ1 (How many studies were published over the years?) graphically. Despite the fact that the range of years was not limited in this systematic mapping, the selected papers were published from 2004 onwards. This may have happened because the semantic web concept was conceived in 2001 and it was only

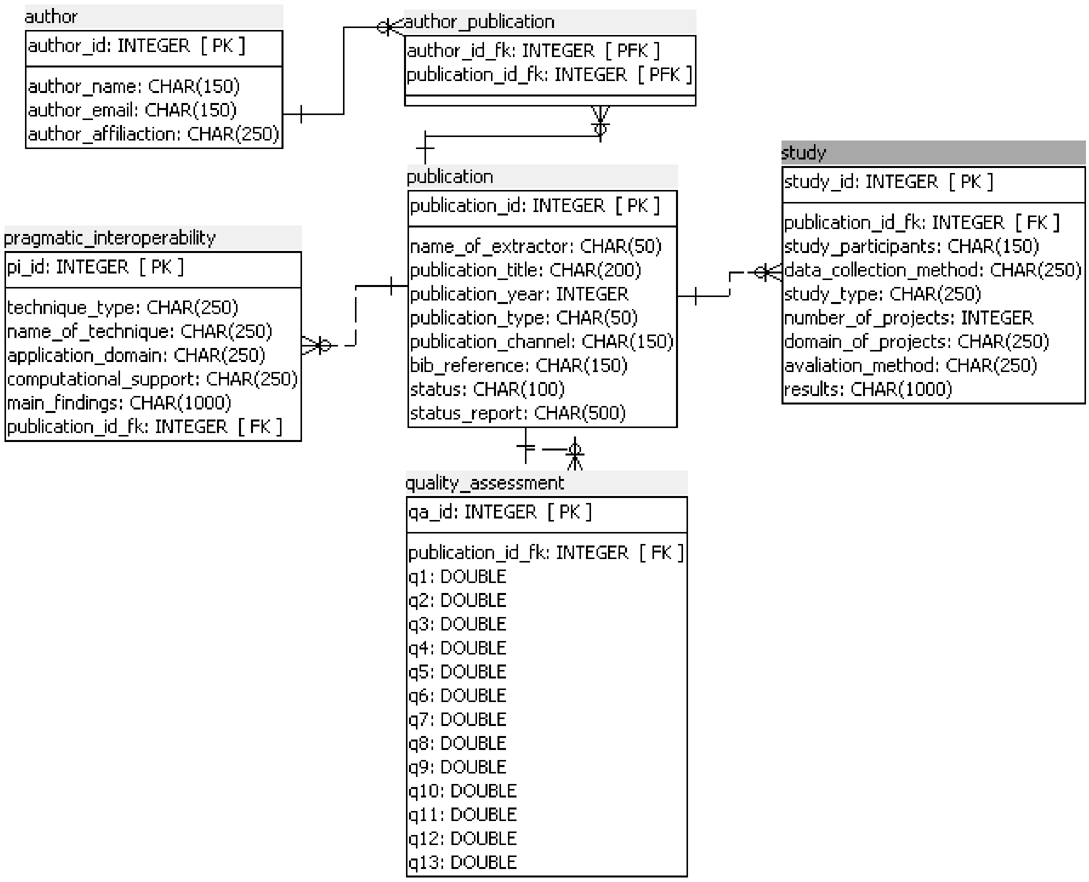

Figure A.4: Database Schema to Support Data Extraction

after this that the challenges it faced started to be discussed Berners-Lee et al. (2001). The first conceptual model to deal with interoperability issues was proposed in 2003 by Tolk and Muguira (2003), motivated by the research agenda of USA Department of Defense Net-Centric Data Strategy. Finally, the Web 2.0 term became widespread in 2004 by O'Reilly Media enterprise increasing interest in interoperability challenges.

To answer MQ2, the authors from the selected papers were identified. Only the researchers Kecheng Liu, James Geller, Yugyung Lee and Lea Kutnoven appeared more than once in our results. The authors' name and the total number of related publications are illustrated in Table A.2.

Regarding MQ3 (Which publication vehicles are the main targets for research production in the area?), Figure A.3 shows where the selected papers were published. Most papers, 8 from a set of 13, were published at Conferences (61.5%) and the other 5 (38.5%) papers were published in Journals (15.4%) and at Workshops (23.1%). In Computer Science, the number of Conferences is significantly higher compared with the number of Journals, which explains why most papers were published through this channel.

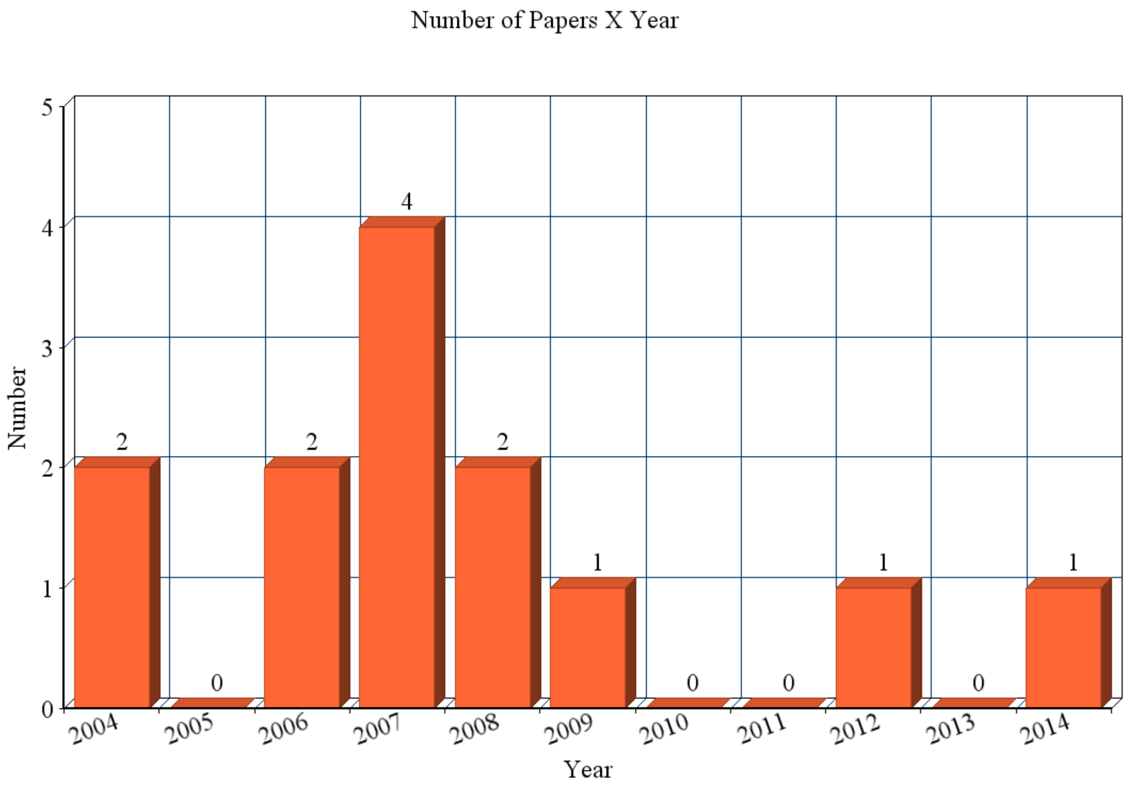

Figure A.5: MQ1 - Number of Papers X Year

Name	Total
Kecheng Liu	3
James Geller, Yugyung Lee, Lea Kutvonen	2
Zhongfu Wu, Boriana Rukanova, Lin Liang, Pieter De Leenheer, Goran D. Putnik, Gan Mingxin, Min Gao, Wenge Rong, Zlata Putnik, Robert A. Stegwee, Andreas Tolk, Soon Ae Chun, Jejung Lee, Luís Ferreira, Electra Tamani, Saikou Y. Diallo, Sanket Shah, Janne Metso, Chintan Patel, Kecheng Liu, Kees van Slooten, Paraskevas Evripidou, Toni Ruokolainen, Stijn Christiaens, Maria Manuela Cruz-Cunha, Charles D. Turnitsa, Weizi Li, Shixiong Liu	1

Table A.2: Authors' Names and Number of Publications

The channel of the published papers is also important information to understand the main targets in the area of pragmatic interoperability, according with Table A.3. The h-index value was also included for a better understanding about the productivity and impact of the published works in a certain publication channel.

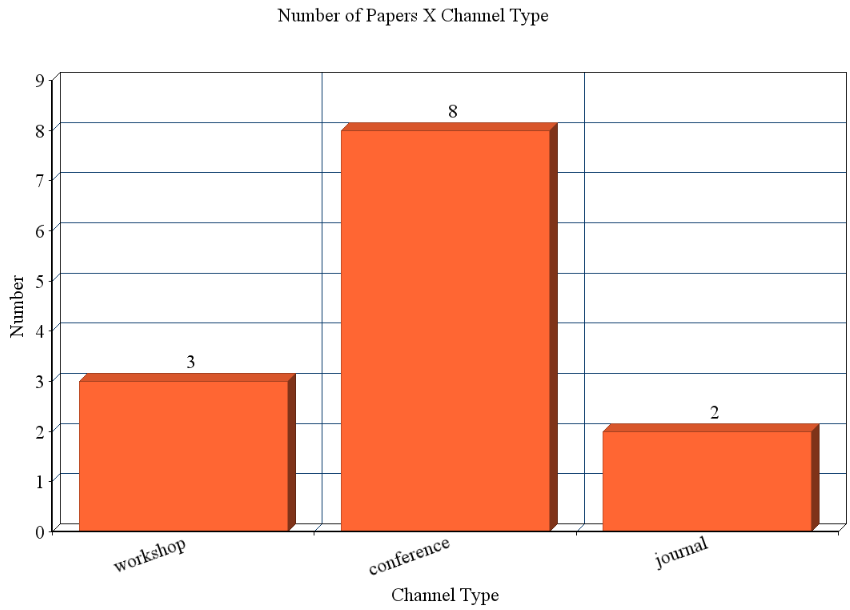

Figure A.6: MQ3 - Publication Channel

The selected papers were also classified according to the domain, where the pragmatic interoperability solution was applied, in order to answer MQ4 (In which domain has pragmatic interoperability been applied?). As can be seen in Figure A.7, pragmatic interoperability solutions were applied in e-business (46.1%). The set of selected papers also presented pragmatic interoperability solutions to support World Wide Web (15.4%), e-health domain (15.4%), intelligent systems (7.7%), grid computing (7.7%) and geosciences (7.7%).

Regarding MQ5 (Which type of computational support has the pragmatic interoperability technique provided?), Figure A.8 presents the type of computational support proposed by the selected papers. Pragmatic interoperability solutions mostly involved different approaches, such as: architecture model, other models, framework and/or methodology. Most of selected papers (53.8%) proposed an architecture to support pragmatic interoperability issues, followed by 38.5% that proposed a framework, 38.5% other models and 15.4% that proposed a method to address pragmatic interoperability research questions.

The classification of the papers in order to answer MQ5 was performed based on the explicit mention in the paper of the type of computer support. For example, if the authors mentioned that a framework was used, the computational support was classified as a "framework". However, sometimes it was possible to notice that the concepts of framework

Reference	Channel Name	H-Index
[Lee et al. 2007]	ACM Symposium on Applied Computing (SAC)	61
[Liang et al. 2007]	International Conference on Advanced Language Processing and Web Information Technology (ALPIT)	7
[Kutvonen 2008]	IEEE Enterprise Distributed Object Computing Conference Workshops (EDOCW)	13
[Mingxin 2009]	International Conference on e-Business Engineering (ICEBE)	20
[Rukanova 2006]	International Conference on Enterprise Information Systems (ICEIS)	24
[De Leenheer & Christiaens 2007][Tamani & Evripidou 2006]	International Conference on the Pragmatic Web (ICPW)	-
[Lee et al. 2004]	IEEE International Conference on Web Services (ICWS)	58
[Kutvonen et al. 2007]	International Journal of Enterprise Information Systems (IJEIS)	11
[Ferreira et al 2012]	Information Resources Management Journal (IRMJ)	22
[Gao et al. 2008]	IEEE International Conference on Service Operations and Logistics and Informatics (SOLI)	-
[Tolk et al. 2006]	World Multi-conference on Systemics, Cybernetics and Informatics (WMSCI)	5
[Liu et al. 2014]	IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid)	60

Table A.3: MQ3 - Publication Channel

and model were mixed and merged. The architecture concept was usually used to represent systems components and their relationships (GAO et al., 2008), (KUTVONEN, 2008), (LEE et al., 2004), (TOLK et al., 2006), (KUTVONEN et al., 2007), (FERREIRA et al., 2012) and (MINGXIN, 2009). The framework concept was used to represent different issues. In (LIU et al., 2014) the term "framework" was used to describe a table that summarizes empirical findings (best practices, recommendations, among others) in order to guide the integration process in the context of radiology. In (LEE et al., 2007), a combination of models and semantic web technologies was termed as a framework. Liang et al. (2007) called a framework as a set of ideas and structures to enable the operation

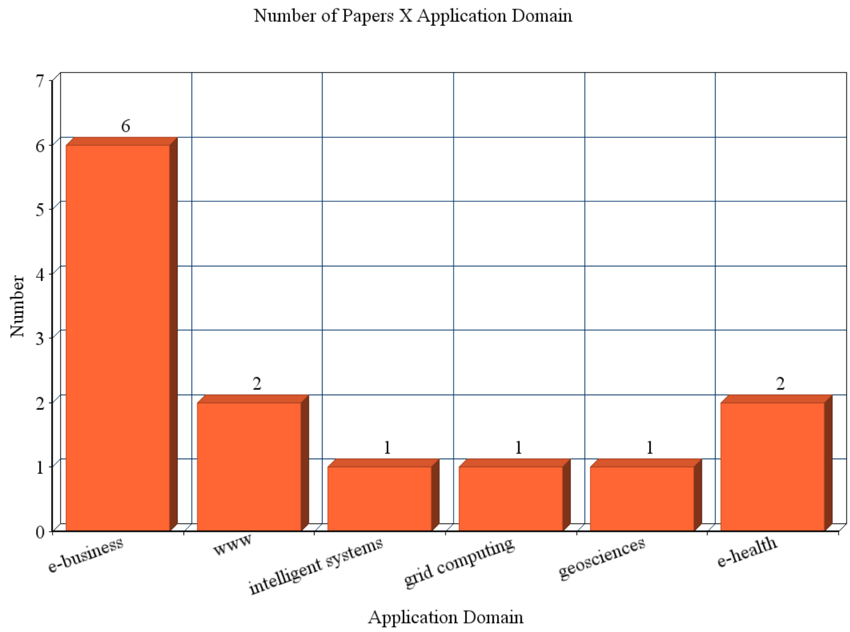

Figure A.7: MQ4 - Application Domain

of a pragmatic web service. In (LEENHEER; CHRISTIAENS, 2007) the framework was considered as an ontology that represents aspects of semiotics engineering aspects. Finally, Gao et al. (2008) proposed a structure to store pragmatic information and called it a framework. Sometimes model and framework were used as synonyms. Model concept was used to represent a structure or a description that supports how to describe something in (LEE et al., 2007), (KUTVONEN et al., 2007), (KUTVONEN, 2008), (RUKANOVA et al., 2006) and (LEE et al., 2004). In (LEENHEER; CHRISTIAENS, 2007) and (GAO et al., 2008) the same representation was called as a framework instead of a model. The concept of methodology was used to represent a sequence of procedures in (TAMANI; EVRIPIDOU, 2006) and (LEE et al., 2004)

Regarding MQ6 (Which definition of pragmatic interoperability has been used?), Table A.4 presents the definitions used in our set of selected papers. Most papers (76.9%) did not explicitly state the pragmatic interoperability definition. As a result, there were differences about what requirements need to be fulfilled to achieve pragmatic interoperability.

In addition, Tolk et al. (2006) followed the definition proposed by Tolk and Muguira (2003) and the 2 remaining papers (14.3%) presented distinct but very close definitions of pragmatic interoperability.

In order to answer MQ7, the methods used to validate the proposed pragmatic interoperability solutions were identified. Table A.5 presents the reference of the papers

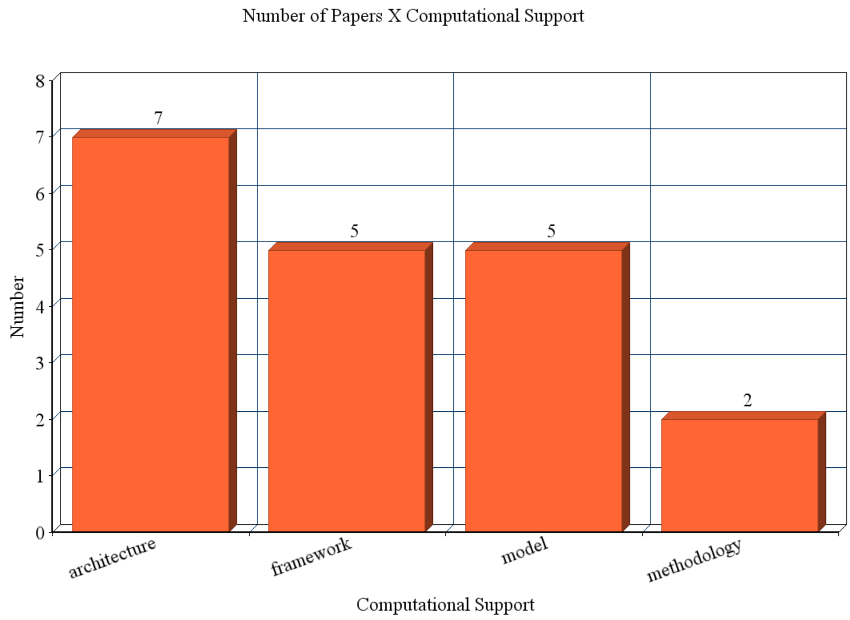

Figure A.8: MQ5 - Type of Computational Support

Reference	**Definition**
[Liu et al., 2014]	The ability to aggregate and optimize various business processes to achieve intended purposes of different information systems
[Rukanova 2006]	The ability to express and compare the requirements of the business transaction and the capability of the standard to cover these requirements
[Tolk et al, 2006]	The definition proposed by the Levels of Conceptual Interoperability Model (LCIM) [Tolk and Muguira, 2003]: The use of the data – or the context of its application – is understood by the participating systems; the context in which the information is exchanged is unambiguously defined.
[Tamani and Evripidou] [Mingxin, 2009] [Lee et al., 2007] [Liang et al., 2007][Kutvonen et al., 2007][De Leenheer and Christiaens, 2007][Gao et al., 2008][Kutvonen 2008] [Ferreira et al., 2012][Lee at al., 2004]	Not explicitly stated

Table A.4: Pragmatic Interoperability Definitions

with the corresponding identified validation method. The classification was based on the explicit mention of the validation method that was presented. For example, if the paper authors mentioned that a case study was presented, the validation method was classified as a case study. From the 13 papers, 5 papers (38.5%) did not use any validation method at all to evaluate the proposed pragmatic interoperability solution, which threatens the validity of the proposals. The most commonly used validation method in the set of select studies was illustration (30.8%), wherein the illustration includes examples and scenarios as well. However, the evidence generated from this method does not have the same rigor as empirical studies have, which threatens its reliability. Among the remaining papers, 2 of them identified the validation method as a case study and the other 2 papers as proof of concepts.

According to Runeson et al. (2012): *case study in software engineering is an empirical enquiry that draws on multiple sources of evidence to investigate one instance (or a small number of instances) of a contemporary software engineering phenomenon within its real-life context, especially when the boundary between phenomenon and context cannot be clearly specified.*

Regarding the definition above, it was possible to notice that although (LEENHEER; CHRISTIAENS, 2007) and (GAO et al., 2008) classified their validation method as a case study, the applied method does not fit in with the definition of a case study in software engineering context. In both papers, the evaluation was performed through a defined scenario where the use of the pragmatic interoperability solution was illustrated.

Proof of concepts is usually applied in software engineering at the initial stages of a project to demonstrate whether an idea has the potential to be used and is feasible (MONTERO et al., 2008), (YANG, 2005). In (LEE et al., 2007) and (LEE et al., 2004), the validation method was defined as a proof of concepts by the authors. In both cases, the authors developed a system where the pragmatic interoperability solution could be applied in order to demonstrate its feasibility. In (LEE et al., 2007) a web based system was developed and in (LEE et al., 2004) the system was under development.

Once the pragmatic interoperability field was mapped it was necessary to do a deeper analysis in order to identity, evaluate and interpret the included studies to answer the systematic review research questions. The next section presents the systematic review report.

Reference	Validation Method
[Mingxin, 2009] [Liang et al., 2007] [Kutvonen at al., 2007] [Kutvonen 2008] [Ferreira et al., 2012]	No validation
[Liu et al, 2014][Tamani and Evripidou] [Tolk et al, 2006] [Rukanova 2006]	Illustration
[De Leenheer and Christiaens, 2007] [Gao et al., 2008]	Case Study
[Lee et al., 2007] [Lee et al., 2004]	Proof of concepts

Table A.5: Validation Methods

A.4 SYSTEMATIC REVIEW REPORT

In order to answer RQ1 (Which solutions have been used to promote pragmatic interoperability?) the aspects that were used to promote pragmatic interoperability in the proposed solutions were identified. A summary of the identified aspects and their use by the selected papers are presented in Table 6. Generally, these aspects were used more than once by different papers.

From the results summarized in Table A.6, it is possible to see that most of the selected papers considered service composition, discovery and/or selection (69.2%) and ontologies (69.2%) as essential aspects to achieve pragmatic interoperability support. The remainder of the solutions proposed meta models (38.5%), software agents (30.8%), pragmatic web services (15.4%) and a pragmatic grid (7.7%) respectively.

The high number of papers considering service composition, discovery and/or selection may be explained by the nonfunctional requirements of collaborative systems. The diversity of interactions and the process types where these systems occur imposes flexible rules and policies. In order to deal with this dynamicity the proposed solutions were usually built using Service Oriented Architecture (SOA) concepts. In order to reuse and compose with services from different infrastructures, interoperability is a key requirement to be fulfilled. Therefore these nonfunctional requirements are important to enable collaboration and to achieve meaningful collaboration, and to do this pragmatic interoperability is necessary. To achieve pragmatic interoperability across geographically distributed services, techniques related to discovering, selecting and composing services that meet user expectations have been developed.

The high number of papers considering ontologies may have been found because of ontology - explicit formal specifications of the terms in the domain and relations among them (GRUBER, 1993) - has been widely explored to represent concepts in World Wide Web, especially after the conception of the semantic web in 2001 (BERNERS-LEE et al., 2001). Thus, the use of ontologies seems appropriate to represent semantic and pragmatic aspects to tackle pragmatic interoperability issues.

In order to answer RQ2 (How solutions addressed pragmatic interoperability?), the following subsection discusses the pragmatic interoperability solutions.

A.4.1 PRAGMATIC INTEROPERABILITY PROPOSED SOLUTIONS

This subsection discusses how each proposed solution, presented in the 13 papers, addressed the issue of pragmatic interoperability.

(FERREIRA et al., 2012)

In (FERREIRA et al., 2012), pragmatic interoperability was used as a way to support tourism business activities. In this domain there is a huge number of heterogeneous variables and resources, such as schedules, accommodation, food, tours and others. This set of variables and resources are usually available as services in the tourism information systems and frequently factors that constrain and impose changes in the initial plan appear. Considering this set, tourism services must be pragmatically interoperable in order to assure better alignment between service providers and client expectations and be able to dynamically adapt to new scenarios without disturbing the clients.

To achieve this, this paper presented the Open Tourism Initiative (OTI). OTI is an architecture that handles dynamic tourism packages under a Virtual Enterprises organizational model. This architecture was designed to quickly react to external changes behaving as an adaptive information system. In order to meet clients' expectations, an architecture component, named Tourism Pragmatic Engine (TPE), enables pragmatic aspects of the communication to allow tourist involvement as a co creator of the tourism

Reference	Service composition, discovery or/and selection	Ontology	Software agents	Pragmatic Web Services	Pragmatic Grid	Meta model
[Ferreira et al 2012]	X	X				
[Mingxin 2009]		X				
[Liang et al. 2007]	X	X	X	X		
[Kutvonen et al. 2007]	X	X				X
[Tolk et al. 2006]	X	X	X			
[Gao et al. 2008]	X		X		X	
[Rukanova 2006]						X
[Tamani & Evripidou 2006]	X					
[Lee et al. 2007]	X	X				
[De Leenheer & Christiaens 2007]		X				
[Kutvonen 2008]	X	X				X
[Lee et al. 2004]	X	X	X	X		X

Table A.6: RQ1 - Solutions to Promote Pragmatic Interoperability

activity. The pragmatic aspects are designed to capture context, preferences, humor and other requirements related to tourists.

OTI services must be described and registered through a domain ontology and as a result, based on pragmatic aspects, a brokering mechanism must discover, select and compose with the most appropriate services according to clients' expectations.

(TAMANI; EVRIPIDOU, 2006)

Tamani and Evripidou (2006) argued that businesses that adopt service-oriented technology tend to expand their markets and improve quality through interoperation with other businesses. However, in order to fully exploit this potential scenario, business discovery must be effectively implemented. To reach this goal, the authors proposed a methodology for web services discovery that considers pragmatic aspects during the process. In this way, business services can interoperate according to the intended expectations. The proposed methodology states that the context of web services' usage can be determined considering the context of the collaborators themselves. The contexts of the collaborators are described using a XML standard and, as a result, the context is exchanged and interpreted identically by all collaborators.

(MINGXIN, 2009)

According to Mingxin (2009) enterprise information systems (EIS) are essential to deal with changing requirements and services in the enterprise domain. However, dynamicity is a complex requirement to be fulfilled by EIS. In this work, pragmatic interoperability was used in order to address this requirement. The authors proposed an architecture, named Enterprise Isomorphic Architecture (EIA), that supports syntactic, semantic and pragmatic interoperability across EIS. Relying on the core of EIA is Enterprise Isomorphic Mapping Mechanism (EIMM) which is designed to support the pragmatic interoperability level. EIMM is the architecture component that deals with requirements change and transforms these requirements from business to IT systems through mechanisms based on ontologies that keeps the transformation isomorphic. This transformation must keep the coherence of concept structures to avoid transaction loss. In order to perform ontology mapping during the transformation process, the different aspects of EIS must be connected through a precise representation.

(LEE et al., 2007)

Lee et al. (2007) state that geospatial information sources are heterogeneous and geographically distributed. This scenario makes interoperability among different organizations a complex task. In this paper, the authors aim to investigate the integration of real-time water quality assurance data with geographic data. In this vein, they proposed a context-aware geospatial data and service integration framework that includes a syntactic, a semantic and a pragmatic model based on Semantic Web technologies. The framework allows the analysis of existing dependencies, predicting causes and providing context-aware services. The pragmatic model that is necessary to enable pragmatic interoperability aims to capture the context where a service is used and meets user expectations. Based on pragmatic knowledge, services are discovered, selected and composed. As a proof of concept, a Web-based system, named HIS-KCWater (Hydrologic Information System for Kansas City metro WATERsheds) was developed where the proposed framework as applied. HIS-KCWater is an automated real time analysis and forecasting system for water quality in the watersheds covering the Kansas City metropolitan area.

(LIANG et al., 2007)

In this paper, a pragmatic web service framework was proposed in order to introduce pragmatic aspects in web service technologies. In this proposed solution, intelligent agents play a vital role in managing users' intention, service context, information communication and negotiation between service consumers and service providers. Furthermore, agents must analyze users' requirements to define and rank the semantic equivalent services. Ontologies are used to represent and manipulate web resources.

In order to build the pragmatic web service framework, a pragmatic frame, a knowledge frame, was designed to store pragmatic information. Pragmatic information is used aiming to capture the intention and the context of each web service in the workflow. The pragmatic frame can be used to find and select the right web service in the web in order to solve a particular task.

(KUTVONEN, 2008),(KUTVONEN et al., 2007)

In Kutvonen (2008) and Kutvonen et al. (2007), the Web Pilarcos project are discussed. Web Pilarcos is a B2B middleware designed to reduce the cost of collaboration establishment and to facilitate management and maintenance of networks. This project was a joint effort with VTT, Elisa, SysOpen, Tellabs and TEKES in order to improve participation in electronic business networks as a key for success. The web-Pilarcos architecture and middleware addresses the syntactic, semantic and pragmatic interoperability of autonomous business services in an inter-organizational context.

Pragmatic interoperability was used to support meaningful collaboration between enterprises. In this study, pragmatic interoperability was achieved if the participants in the message exchange had compatible business intentions, rules and organizational policies to perform digital business transactions. In order to reach these goals, a business network model and ontologies were used to define an e-Contract (collaboration contract). A B2B middleware, called Web Pilarcos, implemented the eContract where facilities for service discovery, selection and composition were provided.

(LEENHEER; CHRISTIAENS, 2007)

In this paper the authors argued that the aim of knowledge sharing in the World Wide Web is to collaborate and integrate within and between communities. Contemporary knowledge engineering methods consider only the non-human system parts, investigating only technical and syntactical aspects in the concept modelling. However, the semantic (elicitation) and pragmatic (application context) must not be neglected.

In order to bridge this gap between ”reality” and its modelling concepts, a framework represented through an ontology, named DOGMA was proposed. The DOGMA ontology aims to provide extension points to engineering methods by covering semantic and pragmatic aspects. The pragmatic aspects were considered to improve the interoperability and therefore collaboration on the web.

(TOLK et al., 2006)

In order to support complex and dynamic environments, and decision challenges, interoperability in modeling and simulation (M&S) field have been considered a key issue for operational research. The authors discuss how various layered composability approaches contributed to the definition of the Levels of Conceptual Interoperability Model (LCIM). Moreover, they discuss how the results can be used to derive implications and requirements for ontologies to describe the universe of discourse. In this universe, intelligent agents serve to mediate the communication between agile applications in order to compose the individual systems into a meaningful system of systems. LCIM is a framework defined in previous work which is designed to determine interoperability, including pragmatic interoperability, in the early phases of software engineering process, considering conceptual models.

Furthermore, the paper presents an example of pragmatic interoperability where an Ontology Driven Service-Oriented Architecture was designed. Based on the example, the authors concluded that ontologies are a potential contributor to support the pragmatic interoperability level, and also that this level is relevant to achieve meaningful interoperability.

(GAO et al., 2008)

In Gao et al. (2008), a pragmatic grid architecture (PGA) that extends the semantic Grid is proposed in order to include pragmatic aspects such as context elements and their purpose. As well as this, the authors propose a pragmatic information framework, architecture of interoperation for resource management, and a pragmatic-aware discovery service to locate a suitable resource. PGA is a middleware which encompasses pragmatic capabilities. It was extended to provide pragmatic-aware Grid services (PGS), pragmatic provisioning services (PPS), agent services, and pragmatic information service (PIS). In PGA, agents receive inputs related to the state of their environment and act on the environment change, record the usage history, analyze the historical data to learn their behavior and process negotiation between agents.

The proposed solution benefits the flexible and personalization-awareness resources provision by managing the pragmatic information. Query rules are automatically and dynamically generated, and real requirements are inferred.

(RUKANOVA et al., 2006)

According to (RUKANOVA et al., 2006), to achieve pragmatic interoperability it is necessary to express and compare the requirements on a pragmatic level. To do this, the authors identify the pragmatic interoperability problems that can occur if companies decide to do business transactions electronically by using standards. They propose a metamodel for describing conversations and, as a result to represent interoperability aspects on a pragmatic level.

In order to describe conversations, it was necessary to define what is communicated and the intention behind the communicated information. In this way, the authors used the notions of elementary message (e-message) and action elementary message (ae-message).

The proposed metamodel use was illustrated through an example to identify if pragmatic interoperability could be achieved using HL7 standard .

The main benefit of the metamodel is that it can be used to describe the communication requirements of a business transaction and the capabilities of the standard in the same terms and then it can be easily compared. A mismatch between communication requirements of a business transaction and the capabilities of the standard means that some pragmatic interoperability problem has occurred.

(LEE et al., 2004)

Lee et al. (2004) argued that the main goal of the semantic web is to reduce manual discovery and usage of web resources. Software agents are frequently used to automatically discover resources in order to meet user expectations. However, current studies of web services are not big enough to provide automatic composition.

In this paper the authors propose an architecture of the knowledge processing that is necessary for composing individual services in service flows. Service description knowledge for Web service discovery and composition was divided into syntactic, semantic and pragmatic rules which enables the building of intelligent (pragmatic) web services for automating service compositions.

The models used to describe rules with service concepts allowed the system to identify the relevant rules in a certain domain and to identify and select appropriate Web services for composition. Furthermore, an ontological integration methodology that deals with heterogeneous semantics existing in web service composition was introduced. As a proof of concept, the proposed solution was applied in the cardiovascular domain which required advanced service discovery and composition across heterogeneous platforms of multiple organizations.

(LIU et al., 2014)

In the smart healthcare domain huge numbers of heterogeneous data sources and humans and the technical factors involved make systems interoperability requirement a complex task. Many challenges such as information collision, heterogeneous data sources, policy obstacles, and procedure mismanagement during the interoperability process must be faced.

This paper presented a pragmatic interoperability analysis of systems integration in an on-going project of a radiology department at a local hospital. The project aims to achieve data sharing and interoperability among Radiology Information Systems (RIS), Electronic Patient Record (EPR), and Picture Archiving and Communication Systems (PACS). Based on the results of the analysis, a pragmatic interoperability framework that summarizes the empirical findings and guides the integration process in the radiology context was proposed.

As for the technologies to be used to address interoperability barriers, the framework recommended EAI (Enterprise Application Integration) methodology and the HL7 standards (http://www.hl7.org/).

A.5 ANALYSIS OF PROPOSED SOLUTIONS

These 13 proposed solutions presented different approaches in order to achieve pragmatic interoperability in their contexts. However, similarities that were reinforced in Figure 7 can be noticed. In order to provide a different perspective on the systematic review and mapping results, Figure 7 presents a bubble chart where different combined results can be verified.

Combining MQ1 and MQ5 results, it is possible to notice that papers that proposed frameworks in 2007, followed by models in 2007 and architectures in 2008, were more frequent. In general, most papers proposed architecture, framework and/or model between 2006 and 2008. Combining MQ1 and RQ1, the papers are mainly proposals that use service composition, discovery and/or selection and ontologies to achieve pragmatic interoperability in 2007.

Architecture proposals considering the e-business domain had the highest frequency when MQ4 and MQ5 were combined, followed by other models considering the same domain. Also in the e-business domain, only service composition, discovery and/or selection and ontologies were used to achieve pragmatic interoperability, while in other domains additional techniques were included, for example software agents.

From the analysis of the selected papers, it is possible to notice that pragmatic interoperability was treated as a key issue to meet user expectations in information systems, as Ferreira et al. (2012) and Lee et al. (2007). The satisfaction of these expectations was constantly linked to the discovery, selection and composition of the best services. Ontologies, or other metamodels were frequently associated to representing the syntactic, semantic and pragmatic services aspects, as Lee et al. (2004), Kutvonen (2008) and Kutvonen et al. (2007). Scenarios where pragmatic interoperability is mainly required, frequently involved heterogeneous, distinct and very large data sources, as Liu et al. (2014). Pragmatic interoperability usually acts in order to improve communication between services, systems and even institutions. Communication is a crucial element for collaboration. Adding pragmatic aspects to the communication/interoperation process could enhance collaboration activities.

In order to achieve pragmatic interoperability, pragmatic information must be captured and stored. Ontologies and other metamodels were frequently used for storage, as Rukanova et al. (2006), Kutvonen (2008) and Kutvonen et al. (2007). Pragmatic information was enriched by context, awareness and historical information, as Tamani and Evripidou (2006), Ferreira et al. (2012) and Gao et al. (2008).

In the e-business domain, pragmatic interoperability often exceeds the system level and

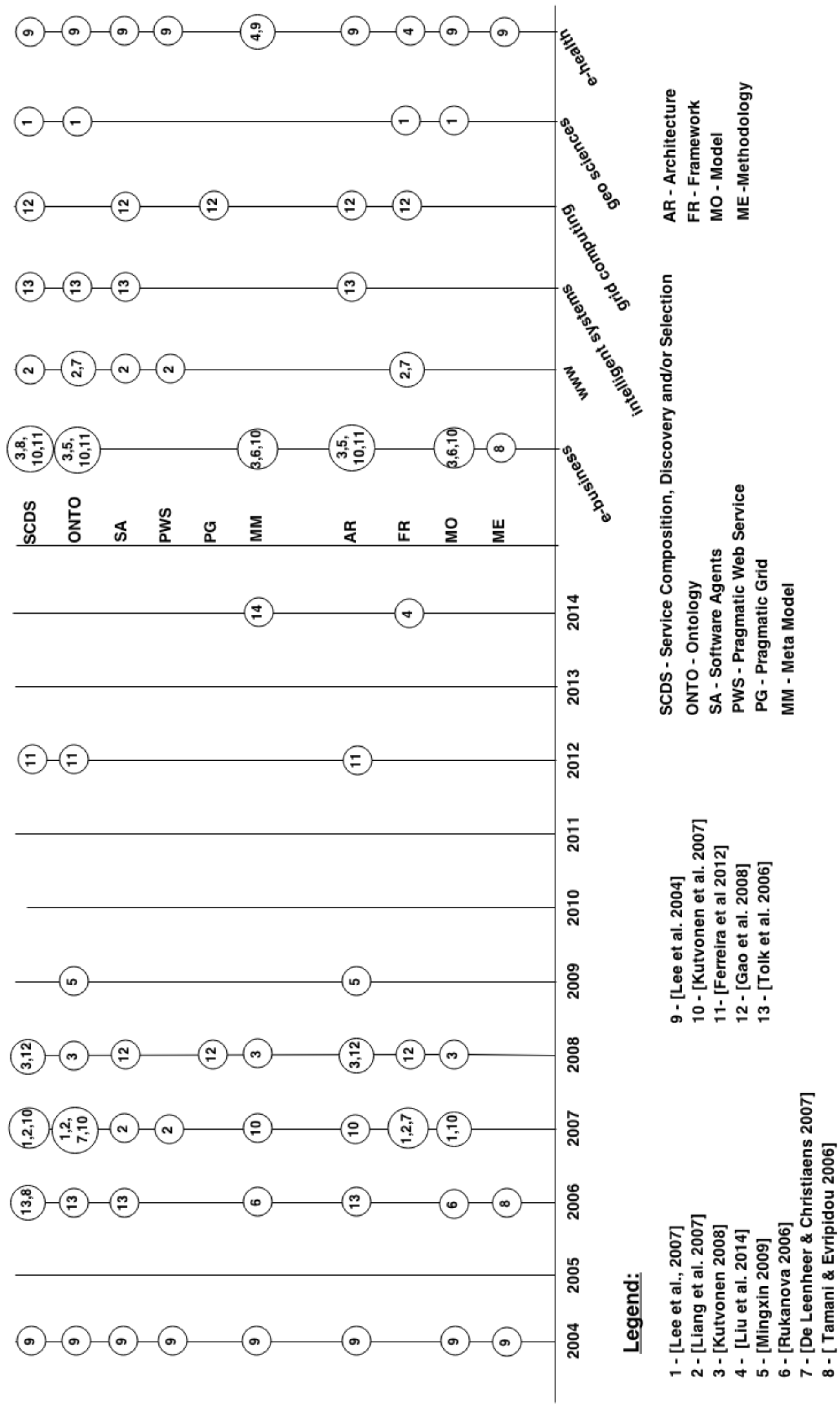

Figure A.9: Summary of Results

starts to act at a business level, considering business rules, business contracts, policies, etc, as Mingxin (2009), Kutvonen (2008) and Kutvonen et al. (2007). We also noticed that the HL7 standard has been used as a base to achieve pragmatic interoperability by adding pragmatic aspects, as Liu et al. (2014) and Rukanova et al. (2006).

In this vein, the knowledge obtained through this study evidenced that pragmatic interoperability is achieved when a desired effect is in accordance with the actual effect considering a message exchange in a relevant and shared context. In other words, pragmatic interoperability aims to meet users expectations in the collaborative processes. Pragmatic interoperability was commonly associated with a service oriented architecture (SOA) (NEIVA et al., 2014). This association can be explained because (SOA) was proposed as solution to enable distributed applications development that is an important factor to enhance collaborative systems design (PAPAZOGLOU et al., 2007). Also, pragmatic interoperability only can be reach if syntactic and semantic interoperability were established.

In order to implement pragmatic interoperability in collaborative systems, some common strategies were adopted by the selected papers (as answered by RQ1). Despite of the importance of pragmatic interoperability in meaningful collaboration, the set of selected papers in this sistematic review and mapping of the literature did not present empirical studies to evaluate its achievement, thus threatening the validity of the proposals. Therefore, it is important to highlight that presenting a proposal which describes an experimental study seems to be a good contribution to this area. Furthermore, a clear and well define requirements to achieve pragmatic interoperability in collaborative systems could be proposed.

A.6 THREATS TO VALIDITY

This systematic literature review and mapping aimed to identify, categorize and analyze pragmatic interoperability solutions in the computational domain. However, as any method, there are threats to its validity and limitations. The results of this study may have been influenced by certain uncontrollable limitations. First of all, there might be bias regarding the number of researchers selecting the papers. In spite of reviewing the overall process, and aiming to mitigate this threat to validity, more than one researcher was able to reproduce this process to reduce the possibility of bias.

Additionally, errors can be inserted in the protocol definition and the search string may not contain all the relevant keywords, causing loss of some valuable studies. To mitigate this, other researchers reviewed the planning presented. The search string was evaluated using papers to control the results. The papers appeared in the results generating evidence about the search string correctness. Furthermore, it was not considered all the relevant electronic databases such as the ACM Digital Library (http://dl.acm.org/) and Springer Link (http://link.springer.com/) excluded by the criteria already explained in section 1, so it is possible that relevant studies were not indexed by our selection. However, it was considered that the selected electronic databases were enough to obtain a big picture of the pragmatic interoperability research area. (KITCHENHAM, 2010) for example, conducted a systematic mapping only using the SCOPUS database. The author argued that SCOPUS indexes IEEE, ACM and Elsevier publications, which means that it indexes many of the leading publications. SCOPUS was in our set of selected databases.

Appendix B - MODEL TO SUPPORT PI ACHIEVEMENT (SM2PIA)

This appendix presents a model to support Pragmatic Interoperability (PI) implementation .

B.1 INTRODUCTION

Despite of the need to achieve PI support in order to enhance collaboration, few researches have been made to investigate how PI can be implemented in software projects. In this vein, based on knowledge acquired from the systematic literature review performed in this work and on study performed by Asuncion and Sinderen (2010), a model to support PI achievement was developed. From the aforementioned researches it was possible to establish requirements to achieve PI. These requirements were organized in a model to PI achievement which was described below.

B.2 MODEL TO SUPPORT PI ACHIEVEMENT (SM2PIA)

This model is divided into two processes. Each process has a purpose and expected results.

Process: Establishment of Pragmatic Interoperability in a System Level (SL)

Purpose: The interaction between systems / services must be in accordance with the intended purpose by the involved parties.

Expected Results:

SL1: The message exchanged in the interaction between systems / services contains the use intention for the receiver;

SL2: The message intention is clearly, and explicitly, defined;

SL3: The message is sent automatically, via a communication channel established between the parties;

SL4: A standard to represent the message intention is unambiguously established;

SL5: The standard to represent the message intention is understood and shared by the involved parties in the interaction;

SL6: Message effect is compatible with its intention;

SL7: The context in which the exchanged message is inserted is clearly and explicitly defined;

SL8: A standard to represent the context in which the exchanged message is inserted is unambiguously established;

SL9: The standard to represent the context in which the exchanged message is inserted is understood and shared by the parties involved in the interaction;

Process: Establishment of Pragmatic Interoperability in Business Level (BL)

Purpose: The collaboration between organizations and / or processes and / or human actors must meet the expectations of the involved parties.

Expected Results:

BL1: Organizational intentions such as business rules and organizational policies are clearly and explicitly defined;

BL2: A standard to represent organizational intent is unambiguously established;

BL3: The standard to represent organizational intentions is understood and shared by the involved parties in collaboration;

BL4: A continuous channel of communication between the involved parties is established;

BL5: Responsibilities and agreements are established among the involved parties;

BL6: Responsibilities and agreements are performed, negotiated and monitored by the involved parties;

BL7: Information about services use (who, how, where, when) are stored and shared;

BL8: The collaborators context (social, cultural, workspace, and so on) is clearly and explicitly defined;

BL9: A standard to represent the collaborators context is unambiguously established;

BL10: The standard to represent the collaborators context is understood and shared by the involved parties in the collaboration;

BL11: Mechanisms to support the establishment of trust among the involved parties are established;

BL12: Organizational intentions are enriched with historical information about the use of shared resources and about the collaborators context in order to better meet the expectations of the involved parties;

SM2PIA provides a guide to implement PI in software projects by compiling requirements to achieve PI. Despite the fact that it establishes ”what” must be done, the model does not define ”how” to be done. ”How” to accomplish the suggested process and their expected results depends of particularities of each system and domain. In this way, an extension of software development process that considers SM2PIA is also discussed in the next section.

B.3 EXTENSION OF SOFTWARE DEVELOPMENT PROCESS TO SUPPORT SM2PIA

As stated by Tolk and Muguira (2003), interoperability must be considered in early stages of software development process. In this way, in order to consider SM2PIA to achieve PI, the main stages of software development processes were extended. Regardless of the software development model adopted (e.g. waterfall, iterative, spiral) the main stages of most software development process are: Requirements, Design, Implementation, Verification and Validation, Maintenance (SOMMERVILLE, 2010).

To illustrate the proposed extension, the waterfall model was taken as example. However, we can extrapolate the proposal of extension in order to be considered in other models (Fig B.1).

As can be seen in Fig B.1, the main differences to consider SM2PIA in software process development occur in Requirement and Design stages, where new artifacts must be generated. A more detailed explanation about these differences is presented in Table B.1.

In this way, the extended process proposed in this section suggests to add the generation of new artifacts in Requirements and Design stages to support SM2PIA in software development process. The artifacts generated in the extended process help to achieve the expected results defined in the SM2PIA.

After these initial stages, the software development process proceeds to implementation stage. In this stage, the Requirements Document, PI Requirements Document, Design Document and PI Design Document are used to enable the code writing. The Verification and Validation and Maintenance stages also considerer these documents to perform their

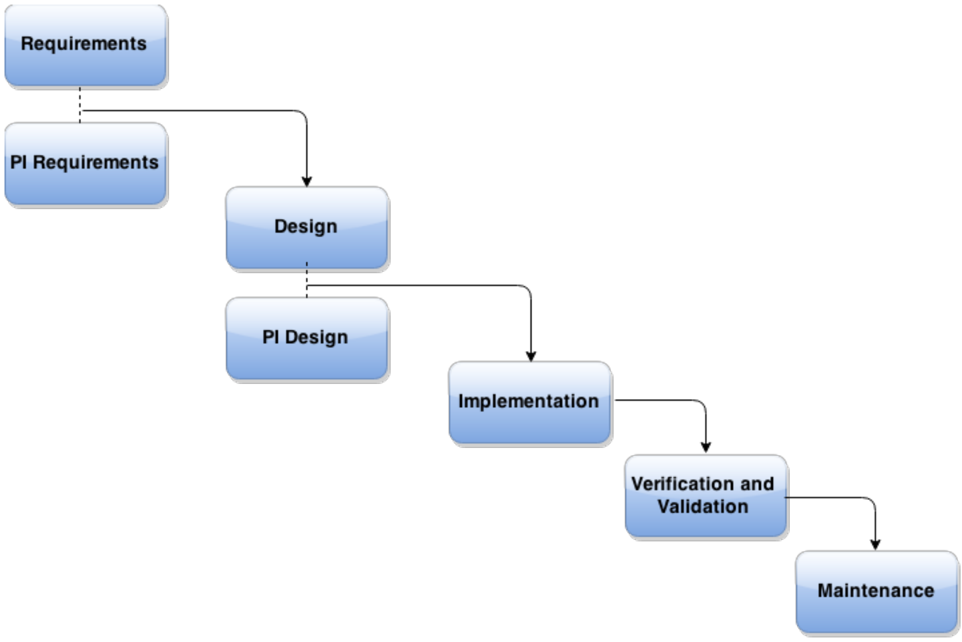

Figure B.1: Extended Waterfall Process Model

activities.

B.4 FINAL CONSIDERATIONS

In this appendix it was presented a support model to guide PI achievement in software projects. In conjunction with the model, it was presented an extended software development process where new activities were proposed in order to consider the support model recommendations.

However, in order to verify and validate the model support and the extended software development process feasibility, it is necessary to implement them in a real software project context.

Stage	Input Artifacts	Actions	Output Artifacts
1.PI Requirements			
1.1 Viability Analysis	None	-Initial survey of the involved parties at system and business levels in the interaction processes. -Impediments or risks of PI implementation should be investigated. (On a system level this can be done by checking the technologies involved, data sources involved, etc. On a business level this can be done by checking organizational policies, financial costs involved, etc.)	PI Viability Report
1.2 Elicitation and Analysis	PI Viability Report	- Interoperability interfaces definition of the systems/services involved; - Definition of systems and organization parts affected by the interaction/collaboration. - Apply elicitation techniques to functional and non-functional requirements discovery related to pragmatic interoperability. - Analysis of the requirements obtained from modeling the sequence of actions of pragmatic interoperability process (use cases and prototypes can be used to question established requirements)	PI Models
1.3 Specification	PI Models	-Documentation of functional and non-functional requirements related to PI obtained in 1.2 in a clear, precise and unambiguous way.	PI Requirements Document
1.4 Validation	PI Requirements Document	The Requirements Document is validated and agreed between the parties as a kind of contract for the establishment of PI.	PI Requirements Document Validated
2.PI Design	PI Requirements Document Validated	-Unambiguous standard to represent the following information is defined: Message intention; Context; Organizational Intentions; Collaborators context. - Definition of interoperability architecture. -Definition of collaboration model. - Definition of storage and sharing models. (How to share and store Message intention; Context; Organizational Intentions; Collaborators context; Information about services use; Responsibilities and agreements)	PI Design Document

Table B.1: Extended PI Process

Appendix C - SERVICE DESCRIPTION ONTOLOGY

This appendix presents the Service Description Ontology in a textual format as follows:

- Service has syntactic, semantic and pragmatic aspects.

- Syntactic has name, parameters and return.

- Parameter has type and value.

- Semantic has semantic output, semantic input, semantic representation and functional requirements.

- Pragmatic includes hardware and context information and has non functional requirements

- Context is described by what, how, when and who (Scientists), domain, license, reputation and restrictions

- Hardware has an operational system, RAM, CPU and ROM.

- Scientist has a name, email and institution.

Appendix D - EXPERIMENTAL STUDY - CONSENT

TERMS OF CONSENT

Research:

”Proposed solution to support pragmatic interoperability in a software ecosystem infrastructure ”

Dear Madam/ Sir,

Providing interoperability support only considering the format and meaning (i.e. syntax and semantic) in data exchange is not enough to achieve effective and meaningful collaboration. Pragmatic interoperability has been identified as a key requirement to foster collaboration in a distributed environment. However, fulfilling this requirement is not a trivial task. The aim of this study is to analyze a solution proposed to support pragmatic interoperability from the point of view of users when developing scientific workflows prototypes in the context of ECOS Collaborative PL-Science.

1) Procedure

The ECOS Collaborative PL-Science environment will be used for the development of a scientific workflow prototype. To participate in this study we asked you a special collaboration to: (1) allow the study of the data resulting from this evaluation, (2) attend an interview and / or a questionnaire. When data is collected, your name will be removed and will not be used at any time during the results analysis or presentation.

2) Treatment of Possible Risks

All the arrangements will be made to ensure your privacy and anonymity. The data collected during the study are intended strictly to research activities, not being used in any form of professional or personal assessment.

3) Benefits and Costs

It is expected that as a result of this study, you can increase your knowledge about collaborative systems. This study will also contribute to the development of collaborative systems and pragmatic interoperability research areas.

You will have no expense or burden on your participation in this study and will not receive any kind of refund or bonus.

4) Research Confidentiality

All information collected in this study is confidential and your name and your organization will not be identified in any way, except in case of express authorization for this purpose.

5) Participation

Your participation in this study is very important and voluntary. You have the right to leave this study at any time without penalty. If you decide to withdraw from this study, please notify a responsible researcher.

The responsible researchers for this study will provide any explanation, just contact the following emails:

Researcher: Frâncila Weidt Neiva fran.weidt@gmail.com

Advisor: José Maria David Nazar - jmndavid@gmail.com

Co-Advisor: Regina Braga regina@acessa.com

6) Declaration of Consent

I read or someone read to me the information in this document before I sign this consent form. I declare that all the technical language used in describing this research study was explained adequately and I received answers to all my questions. I also confirm that I received a copy of these Terms of Consent. I understand that I am free to withdraw from the study at any time without penalty. I declare to be older than 18 years old and I give my consent to participate in this study.

Place e Date:

Organization:

Name:

Signature:

Name: Frâncila Weidt Neiva

Signature:

Appendix E - EXPERIMENTAL STUDY - TASKS

This appendix presents the tasks proposed in the experimental study.

The presented tasks to the subjects were:

Task 1

Group 1 - Ad-Hoc Method

NOTE 1: Please note the time when you start this task and also the time when you finish it. This information will be requested in the follow-up questionnaire that will be answered after this task accomplishment.

Goal: To develop the prototype of a scientific workflow on blood heritage domain.

Instructions:

1) Go to the address: http://www.nenc.ufjf.br/plscience/

2) In ECOS Collaborative PL-Science environment click in "+ New Experiment"

3) In title field type: "Inheritance of blood types" + your name. In description field type: "Participation in PRIME evaluation" + date of participation + Group. Then, click in 'Save' button.

4) Select Experiments -> My Experiments

5) Click in 'edit' button in the experiment that you created.

6) Select the tab "Experiment Prototyping".

7) Now it's time to discover services:

7.1) Select the best services to perform the following activities:

Step 0: the decoding of parental blood type phenotypes in genotypes.

Step 1: the combination of parental genotypes.

Step 2: the calculation of child blood type probability.

8) Answer the Group 1 - Follow-up questionnaire.

Task 2

Group 2 - PRIME

In this task, PRIME architecture was activated. In this way, subjects followed the same instructions that was presented in task 1, but in this case, they accessed PRIME functionalities.

By finishing this task, subjects answered the Group 2 - Follow-up questionnaire.

After the subjects have accomplished the tasks and have answered the questionnaires, they were asked to expose suggestions and claims in order to improve the experiment.

Appendix F - FOLLOW-UP QUESTIONNAIRE

This appendix presents the follow-up questionnaire.

In the experimental study, after to accomplish the first task (PRIME non-use), the subjects answered the following questions:

Q1- In you opinion, the difficulty level of the proposed task performance was:

[ ] very low

[ ] low

[ ] median

[ ] high

[ ] very high

Justify:

Q2- Were you able to create the scientific workflow prototype properly.

[ ] strongly disagree

[ ] disagree

[ ] neither agree nor disagree

[ ] agree

[ ] strongly agree

Justify:

Q3- You could easily discover which service to use at each workflow prototype stage.

[ ] strongly disagree

[ ] disagree

[ ] neither agree nor disagree

[ ] agree

[ ] strongly agree

Justify:

The same three questions were made after the subjects have accomplished the second task (PRIME use). However, a fourth question was added:

Q4- PRIME supports service discovery, selection and composition considering the experiment prototyping step.

[ ] strongly disagree

[ ] disagree

[ ] neither agree nor disagree

[ ] agree

[ ] strongly agree

Justify: